



UNIVERSITÉ LIBRE DE BRUXELLES
FACULTÉ DE LETTRES, TRADUCTION ET COMMUNICATION

ULB

La production automatisée d'informations en appui aux pratiques journalistiques

Analyse des représentations, des conditions d'association
et de la structuration des usages en Belgique francophone

Laurence DIERICKX

Thèse présentée en vue de l'obtention du grade
académique de Docteur en Langues, Lettres et
Traductologie / Information et Communication
sous la direction de Monsieur le Professeur
David DOMINGO et de Monsieur le
Professeur Seth VAN HOOLAND

Année académique 2019-2020

Remerciements

Cette thèse n'aurait pas vu le jour sans la confiance que m'ont témoignée mes deux co-directeurs, David Domingo et Seth van Hooland. Je les remercie de m'avoir permis de poursuivre cette belle aventure et pour leur appui dans son cheminement. Je remercie également les membres de mon comité d'accompagnement, Isabelle Boydens, Florence Le Cam et Sébastien de Valeriola, pour leurs conseils avisés qui me furent précieux.

Merci à la rédaction d'Alter Échos pour avoir accepté de jouer les cobayes en ouvrant ses portes à "Bxl'air bot", et un merci particulier à Sandrine Warsztacki et Céline Gautier. Un autre merci particulier à Nicolas Becquet qui, en m'invitant dans le projet "Quotebot", a permis d'enrichir le propos de cette thèse. Merci aussi aux journalistes du service "Investir" de L'Echo pour avoir accepté de me consacrer un peu de leur temps dans des journées que je sais intenses.

Merci à tous les professionnels qui ont accepté mes sollicitations pour éclairer cette recherche de leurs témoignages, et à Olivier Trédan pour m'avoir conseillée dans mes doutes et questionnements. Merci à encore à Roland Huvinen, le meilleur traducteur suédois-français que je connaisse, Steve Hill pour ses (nombreuses) relectures en langue anglaise et à Mehmet Koksal, pour son aide ponctuelle; ainsi qu'à Marie-Soleil Frère, pour m'avoir encouragée à me lancer dans cette expérience, et à Gonzague Hervy, Sandrine Saint Hilaire et Pascal Durant pour m'avoir soutenue tout au long de ces presque cinq années.

Résumé

Le phénomène de la production automatisée d'informations, caractérisé par un processus de transformation de données structurées en textes en langage naturel ou en toute autre forme de représentation visuelle, interroge la manière dont il se développe dans le monde du journalisme. S'il peut être envisagé comme un produit fini qui sera délivré aux audiences sans médiation journalistique, cette thèse s'intéresse à son autre versant : celui d'un appui aux pratiques journalistiques. Souvent désigné par la métaphore du "robot journaliste", ce phénomène nourrit des représentations duales qui reflètent ses enjeux : tantôt abordé en termes de menace sur l'emploi, voire sur l'identité professionnelle des journalistes ; tantôt perçu comme une opportunité, dans une vision enchantée d'un journalisme en réinvention. Ceci pose la question du rapport du journaliste à l'objet technique, dès lors que celui-ci est conçu comme un outil qui lui fera gagner du temps dans ses routines, ou qui se posera en tant que source d'information s'ajoutant à un arsenal existant. Les objectifs de cette thèse sont de comprendre les mécanismes de structuration des usages professionnels dans le contexte de la production automatisée d'informations, d'identifier les sources de résistance qui ne permettent pas de réaliser ces usages et, partant, de contribuer à une meilleure compréhension d'une relation homme-machine placée sous tension. Considérant que l'on ne peut faire usage d'une technologie sans se la représenter, de quelle manière les représentations sociales et les pratiques culturelles des journalistes vont-elles participer à la structuration des usages (et des non-usages) d'artefacts de la production automatisée d'informations lorsque ceux-ci sont envisagés en tant qu'outils du journalisme ? Cette question de recherche a été envisagée selon deux hypothèses : celle de l'impact de la métaphore du "robot" quant à des résistances contre l'intégration de l'objet dans les routines professionnelles, et celle de l'adéquation des productions matérielles de l'artefact avec les savoir-faire et exigences journalistiques. L'analyse mise en œuvre s'appuie sur les cadres conceptuels et théoriques proposés par le modèle SCOT (*Social Construction Of Technology*) et l'école française de la sociologie des usages. Elle s'est déployée dans le cadre d'une étude empirique menée dans deux rédactions en Belgique francophone, lesquelles se distinguent tant par leur culture professionnelle que par leur structure organisationnelle. Dans ces deux expériences, il s'agissait de suivre le processus d'une construction sociotechnique qui a activement impliqué les journalistes dans la conception des artefacts. Bien que cette première forme d'usage ne garantisse pas des usages finaux, elle pose les premiers jalons d'usages d'adoption. Les résistances observées ne sont pas nécessairement liées à la nature de l'objet technique. Elles sont également à trouver dans un manque d'intérêt envers le domaine d'application, une forme de technophobie, un manque d'appétence pour une approche par données dans le journalisme, une mise en cause du management et de l'organisation, ou encore dans une représentation duale induite par la métaphore du robot. Le mécanisme de structuration des usages va donc relever d'un processus complexe dépendant à la fois de facteurs endogènes, liés au contexte organisationnel et aux routines journalistiques, et de facteurs exogènes, liés aux cadres plus larges de l'imaginaire technologique et de la relation homme-machine.

Mots clés : automatisation, pratiques professionnelles, journalisme algorithmique, journalisme computationnel, diffusion de l'innovation, imaginaires technologiques, TAL

Abstract

The phenomenon of news automation, characterized by the process of transformation of structured data into texts in natural language or any other form of visual representation, asks the question about how it is integrated within newsrooms. If it can be considered as a finished product that will be delivered to audiences without journalistic mediation, this thesis is interested in its aspect related to the support of journalistic practices. Often associated with the metaphor of the "robot journalist," news automation encourages dual representations that reflect its challenges : either a threat against employment and professional identity or an opportunity through an enchanted vision of the reinvention of journalism. This thesis explores how journalists tackle this new form of relationship as an artefact, as a tool, that is supposed to free them from repetitive and time-consuming tasks or to provide them an additional source of information. The aims of this research project are, therefore, to understand how it will shape new professional uses, to identify the causes of resistance that could lead to non-uses, and to contribute to a better understanding of a man-machine relationship developed under tension. Considering that one cannot make use of a given technology without representing it, how will the social representations and cultural practices of journalists contribute to shaping the uses (and the non-uses) of news automation software when it is considered as a journalistic tool? This research question was examined according to two hypotheses : that of the impact of the robot metaphor, and that of the adequacy of the material productions of the artefact with journalistic know-how and requirements. The analysis is based on conceptual and theoretical frameworks proposed by the SCOT model (Social Construction of Technology) and by the French school of the sociology of uses. It is approached through an empirical study conducted within two newsrooms in French-speaking Belgium, which are radically different both in their professional culture and in their organizational structure. In these two experiences, we have followed the process of a socio-technical construction which involved journalists in the design process. Although this first form of use does not guarantee the end-use of the two automation artefacts, it does lay the groundwork for adoption. The reasons to explain the observed resistances are not necessarily related to the nature of the object. They are also to be found in a lack of interest, a form of technophobia, a lack of appetite for a data-driven approach, a questioning of the management, or the dual representation induced by the robot metaphor. The mechanism for structuring the use will, therefore, be developed through a complex process shaped by both endogenous factors linked to the organizational context and journalistic routines, and exogenous factors related to the broader frameworks of the technological imaginaries and the man-machine relationship.

Keywords : news automation, professional practices, algorithmic journalism, computational journalism, diffusion of innovation, technological imaginaries, NLG

Table des matières

Introduction générale	7
0.1 Problématique et question de recherche	9
0.2 Cadre théorique et conceptuel	12
0.2.1 Construction sociale de la technologie (SCOT)	13
0.2.2 Sociologie des usages	15
0.3 Plan de la thèse	22
0.3.1 Première partie : représentations sociales et pratiques culturelles	23
0.3.2 Deuxième partie : structuration des usages	26
0.4 Implication personnelle	28
0.5 Limites de la recherche	30
I Représentations sociales et pratiques culturelles	31
1 Imaginaires technologiques	32
1.1 Etat de l'art sociohistorique	34
1.1.1 Identification des résistances	41
1.1.2 Reconfiguration des processus sociaux et éditoriaux	44
1.2 Particularités du contexte belge	47
1.2.1 Journalisme de données en Belgique francophone	48
1.2.2 Journalisme de données en Flandre	56
1.3 Enjeux et défis professionnels	57
1.3.1 Gérer les angoisses professionnelles	58
1.3.2 Trouver l'équilibre : les promesses du partenariat homme-machine	64
1.3.3 Comprendre le fonctionnement de la machine	70
1.4 Imaginaires du "robot journalisme"	77
1.4.1 Méthode	78
1.4.1.1 Approche internationale	79
1.4.1.2 Focus sur les titres	81
1.4.1.3 Diversité des médias et des auteurs	82
1.4.1.4 Modèle d'analyse du discours	83
1.4.2 Résultats	87
1.4.3 Mise en perspective	94

2	Entre deux mondes : les dimensions culturelles de l'automatisation	96
2.1	Dimension cognitive	99
2.1.1	Les données, une cause matérielle	100
2.1.2	Gérer la qualité des données	108
2.1.3	Modèle conceptuel d'évaluation de la qualité des données	114
2.1.4	Traitement des données : l'illusion de l'objectivité	119
2.2	Dimension normative	122
2.2.1	Dans le monde du journalisme	124
2.2.2	Dans le monde de la technique	131
2.2.3	Définir un cadre commun	138
2.3	Dimension évaluative	145
	Conclusions de la première partie	148
II	Structuration des usages	154
	Introduction aux études de cas	155
3	L'artefact "Bxl'air bot" dans les usages de la rédaction d'Alter Échos	159
3.1	Contexte de l'expérience	161
3.1.1	Chaîne sociotechnique et cadre technologique	163
3.1.1.1	Le magazine Alter Échos	164
3.1.1.2	Les membres de la rédaction	168
3.1.2	Conception de l'artefact technologique	171
3.1.2.1	Accessibilité des données	172
3.1.2.2	Évaluation de la qualité des données	174
3.1.2.3	Architecture du système d'information	177
3.1.2.4	Prototypage de l'interface	180
3.1.2.5	Maintenance du système	182
3.2	Construction sociale de l'artefact technologique	183
3.2.1	Premières résistances et nouvelles formes d'usages	189
3.2.2	Clôture et évaluation de l'expérience	196
3.3	Analyse des observations	202
4	L'artefact "Quotebot" dans les usages de la rédaction de L'Echo	208
4.1	Contexte de l'expérience	210
4.1.1	Chaîne sociotechnique et cadre technologique	213
4.1.1.1	Le journal L'Echo	214
4.1.1.2	Les membres de la rédaction de L'Echo associés au projet	218
4.1.1.3	Syllabs	220
4.1.1.4	VWD Group	228
4.1.2	Modèle d'analyse et d'évaluation du dispositif sociotechnique	229
4.1.2.1	Fondements théoriques	229
4.1.2.2	Méthodes d'évaluation	231

4.1.2.3	Médiation sociotechnique et norme ISO 9000	233
4.2	Construction sociale de l'artefact technologique	235
4.2.1	Dans le monde du journalisme	237
4.2.2	Dans le monde de la technique	240
4.2.3	Processus de correction	244
4.2.4	Clôture et évaluation de l'expérience	247
4.2.4.1	Analyse de la médiation sociotechnique	248
4.2.4.2	Évaluations automatiques des textes générés	253
4.2.4.3	Évaluations basées sur des jugement humain	258
4.3	Analyse des observations	263
 Conclusions générales		271
 Retour sur un processus de recherche et des situations d'immersion		281
 Bibliographie		289
 Sources		318
 Annexes		322
Annexe 1.	Offre d'emploi pour un poste de datajournaliste au Guardian	322
Annexe 2.	Code pour l'analyse de sentiment du discours médiatique des journalistes .	324
Annexe 3.	The Critical Engineering Manifesto	326
Annexe 4.	Modélisation du processus de Voitto, développé par le radiodiffuseur Yle (Finlande)	327
Annexe 5.	Structure pour l'entraînement du "robot" de l'agence de presse norvégienne NTB	328
Annexe 6.	Présentation du processus de génération automatique de textes météorolo- giques relatifs à des destinations de vacances	330
Annexe 7.	Différenciation des solutions de génération automatique en langage naturel et directions du marché	331
Annexe 8.	Formulaires d'évaluation diffusés via Google Forms dans le cadre de l'expé- rience "Bxl'airbot"	332
Annexe 9.	Guide d'entretien pour l'évaluation finale de l'expérience "Bxl'air bot" . . .	334
Annexe 10.	Données en temps réel diffusées par CELINE (web service)	335
Annexe 11.	Première version de l'interface de l'application "Bxl'air bot"	337
Annexe 12.	Génération automatique de textes en langue naturelle	338
Annexe 13.	Échantillon d'anomalies détectées dans le cadre du monitoring des données	340
Annexe 14.	"Bxl'air bot" : tableau de bord (modules d'analyses statistiques et module "Moyennes par station de mesure")	342
Annexe 15.	"Bxl'air bot" : modèle système d'alerte automatisé	343

Annexe 16. Article publié dans le dossier "Pollution, l'air de rien"	344
Annexe 17. Évolution de l'interface du site web de <i>L'Echo</i> , de 2002 à 2019	350
Annexe 18. Modes de calcul des indicateurs automatiques	351
Annexe 19. Formulaire d'évaluation diffusés via Google Forms dans le cadre de l'expérience "Quotebot"	352
Annexe 20. Échantillon de textes rédigés par les journalistes du service "Investir" de <i>L'Echo</i>	355
Annexe 21. Interface de démonstration de Syllabs	358
Annexe 22. Échantillons de textes sources et de texte cibles	359
Annexe 23. Liste des templates de "Quotebot" et indications sur la récupération des données	365

Table des figures

1.1 Typologie d'une approche par données dans le journalisme (Coddington 2014)	40
1.2 Processus théorique "en pipeline" de génération automatique de textes, d'après Danlos (1991) et Reiter & Dale (2000)	68
1.3 Corpus de recherche	80
1.4 Répartition du corpus par langue et par pays	80
1.5 Répartition des échantillons du corpus par type de média en ligne.	83
1.6 Analyse du discours : termes employés	87
1.7 Évolution de la recherche des locutions "Intelligence artificielle" et "Artificial intelligence" dans Google Trends	88
1.8 Analyse du discours : connotation des titres	89
1.9 Influence de l'usage de la métaphore du robot par type de publication	90
1.10 Analyse du discours : enjeux socioprofessionnels (niveau pragmatique)	91
1.11 Analyse des similitudes : arbre maximum (corpus francophone)	92
1.12 Analyse des similitudes : arbre maximum (corpus anglophone)	93
2.1 L'application "Crime L.A." du <i>L.A. Times</i>	103
2.2 L'application "Quakebot" du <i>L.A. Times</i>	104
2.3 L'application "Fenstaub Monitor" du <i>Berliner Morgenpost</i>	104
2.4 La plateforme en ligne "Données du Monde"	105
2.5 "JSA Figures", la rubrique automatisée de <i>On The Wight</i>	106
2.6 Premières productions de "RADAR" publiées dans la presse britannique	106
2.7 Premières productions de "Tobi" publiées sur les sites du groupe Tamedia	107
2.8 Modèle réalisé pour l'évaluation du portail <i>open data</i> de Rennes Métropole	115

2.9	Modèle conceptuel du cycle de vie des données	118
2.10	Composant logique et composant de contrôle d'un algorithme (Kowalski 1979)	137
2.11	Points de convergence et points à négocier entre les mondes de la technique et du journalisme (schéma de l'auteure)	143
2.12	Mise en correspondance des étapes du modèle en cascade du cycle de vie du logiciel avec le modèle SCOT	157
3.1	Schéma conceptuel de la chaîne sociotechnique formée autour de l'artefact "Bxl'air bot"	163
3.2	Évolution de la maquette d'Alter Échos en 1997, 2001, 2007, 2012 et 2018	166
3.3	Modèle économique du magazine Alter Échos en 2018	167
3.4	Page d'accueil du site web de Bruxelles Environnement	173
3.5	Diffusion des mesures des taux de polluants sur le site de CELINE	173
3.6	Cycle de vie des données relatives aux mesures des taux de polluants atmosphériques en région bruxelloise	176
3.7	Première version de l'arborescence de l'application "Bxl'air bot"	178
3.8	Modélisation du flux de données de "Bxl'air bot"	178
3.9	Modèle des attributs d'acceptabilité d'un système (Nielsen 1999)	181
3.10	Interface pour le monitoring des données diffusées en temps réel	183
3.11	"Pollucarte" de la région bruxelloise	185
3.12	Affichage du bulletin du jour de "Bxl'air bot" sur le site d'Alter Échos	186
3.13	Arborescence finale de l'application automatisée	187
3.14	Processus d'automatisation des enregistrements et productions automatisés	188
3.15	Architecture de la base de données	189
3.16	Influence de l'expérience sur la représentation du phénomène du journalisme automatisé	190
3.17	Évaluation de l'interface de "Bxl'air bot"	191
3.18	Évaluation du bulletin généré en temps réel	191
3.19	Aspects susceptibles de trouver un prolongement pratique	192
3.20	Aspects que les journalistes souhaitent approfondir	193
3.21	Sommaire du dossier publié dans Alter Échos n°464	197
3.22	Métaphore du robot interviewé par des journalistes	198
4.1	Postes de travail utilisés par un journaliste du service "Investir", de gauche à droite : un tableur présentant des valeurs boursières évoluant en temps réel, le système de gestion de contenus interne et les informations de l'agence Bloomberg	210
4.2	Schéma conceptuel de la chaîne sociotechnique formée autour de "Quotebot"	213
4.3	Premier numéro de <i>L'Echo de la Bourse de Bruxelles</i> , le 22/05/1881	214
4.4	Évolution de la diffusion payante de <i>L'Echo</i> , de 2010 à 2015 (Centre d'information sur les médias, CIM)	217
4.5	Processus conceptuel du moteur de rédaction développé par Syllabs, infographie promotionnelle diffusée en ligne en 2019	222
4.6	Workflow de la première phase de la conception de "Quotebot"	236

4.7	Intégration des productions de "Quotebot" dans le CMS de Mediafin	245
4.8	Échantillon de texte généré de manière automatique, prélevé le 08/01/2019 . . .	254
4.9	Échantillon de texte généré de manière automatique, prélevé le 11/12/2019 . . .	254
4.10	Comparaison du score Flesch-Kincaid, relatif à la lisibilité des textes sources et des textes cibles	255
4.11	Comparaison de l'index Coleman Yau, relatif au degré d'intelligibilité des textes sources et des textes cibles	255
4.12	Comparaison du grade Flesch-Kincaid, relatif au niveau d'éducation nécessaire pour comprendre les textes sources et les textes cibles	256
4.13	Comparaison de l'Automated Readability Index, relatif au degré d'éducation nécessaire pour comprendre les textes sources et les textes cibles	256
4.14	Moyenne des indicateurs automatiques pour les textes sources et les textes cibles	257
4.15	Distance de Levenshtein entre les textes sources et les textes cibles	257
4.16	Évolution du taux de similarité entre les textes sources et les textes cibles	258
4.17	Évaluation des productions de "Quotebot"	260
4.18	Comparaison des deux évaluations des journalistes à propos de la qualité des productions de "Quotebot"	262
4.19	Cercles de la traduction sociotechnique (schéma de l'auteure)	277

Liste des tableaux

2.1	Indicateurs relatifs à la qualité formelle des données	116
2.2	Indicateurs relatifs aux dimensions de la qualité des données	117
2.3	"Bxl'air bot" et "Quotebot" : caractéristiques communes et particulières	156
3.1	Normes de l'Union européenne relatives aux taux de polluants atmosphériques .	161
3.2	Recommandations de l'OMS relatives aux polluants atmosphériques	162
3.3	Évaluation du challenge technique	174
3.4	Évaluation du challenge journalistique	175

Introduction générale

Le phénomène technique est (...) la préoccupation de l'immense majorité des hommes de notre temps, de rechercher en toutes choses la méthode absolument la plus efficace.

Jacques Ellul (1954)

La métaphore du "robot journaliste" est régulièrement utilisée pour désigner des processus informatisés de production d'actualités journalistiques, dont la particularité est de transformer des données structurées en textes, en graphiques ou en toute autre forme de représentation visuelle. Reconnus pour leurs qualités de vitesse et de précision (Graefe 2016), les artefacts d'automatisation de la production d'informations se caractérisent également par leur potentiel de production à grande échelle. Cela étant, les raisons poussant les rédactions à adopter ces technologies vont dépendre de la finalité de leur projet éditorial, qu'il s'agisse de couvrir un événement en temps réel (une rencontre sportive ou des résultats d'élection, par exemple), d'étendre leurs zones de couverture médiatique (comme dans le domaine sportif, en se concentrant sur sports ou divisions peu ou pas couverts jusque-là), de générer différents types de représentations visuelles à partir d'un même jeu de données, ou de personnaliser l'information dans le cadre de stratégies de segmentation des audiences. Une autre application de la production automatisée d'informations consiste à fournir aux journalistes des brouillons automatisés qu'ils enrichiront de leur expertise (Latar 2018 :29) et, partant, à les décharger d'un travail répétitif et chronophage pour leur permettre de se consacrer à des activités plus valorisantes (Carlson 2014).

Cette dernière approche a fait émerger de nouvelles pratiques professionnelles : en amont du système d'automatisation, où des journalistes et des spécialistes des données vont intervenir dans le paramétrage des contenus ou dans le maintien de leur qualité dans le temps (Diakopoulos 2019 :5-32); ou en aval, dès lors que les productions de l'artefact sont destinées à appuyer des routines professionnelles. Se développant dans le sillage de pratiques journalistiques pilotées par des données numériques, la production automatisée d'informations peut être considérée comme l'un des avatars d'une approche par données dans le journalisme (Latar 2018 :29). Elle est d'ailleurs souvent désignée par les locutions "journalisme computationnel", "journalisme algorithmique" ou "journalisme robotique", témoignant du rapprochement

observé, depuis près de deux décennies, entre les mondes de l'informatique et du journalisme (Lewis & Usher 2014).

Face à ce phénomène, les journalistes ont adopté une série d'attitudes antagonistes, qui témoignent de la longue histoire de relations ambiguës qu'ils entretiennent avec les technologies : déterministe ou résiliente, considérant que les développements technologiques sont inévitables et contribuent à créer les conditions d'un travail spécifique ou à renforcer des normes professionnelles existantes (Van Dalen 2012, Linden 2017) ; réactionnaire, estimant que les développements technologiques représentent une menace car elles ne s'accordent pas avec les pratiques et valeurs professionnelles ou parce qu'elles représentent un danger pour l'emploi des journalistes (Graefe 2016, Thurman *et al.* 2017) ; positiviste, envisageant ces développements technologiques comme un levier permettant de réinventer le journalisme (Karlsen & Stavelin 2014). Une quatrième prise de position, relevant d'un déterminisme radical, consiste à envisager la technologie dans une perspective schumpétérienne de "destruction créatrice", en vertu de laquelle l'innovation détruit le travail humain qu'elle automatise, tout en créant des emplois associés à de nouveaux profils (Ekbia & Nardi 2014, Benzell *et al.* 2015). Ces postures, qui ont caractérisé les discours des journalistes à propos des innovations technologiques pendant près de quarante ans (Powers 2012), ne peuvent être dissociées des représentations sociales et culturelles véhiculées par l'objet technologique.

Lorsqu'un artefact d'automatisation de la production d'informations est envisagé en tant qu'outil au service du travail journalistique, ses usages peuvent-ils être freinés ou encouragés par les représentations qui y sont associées ? Cette interrogation, inspirée par la recherche menée dans les domaines de la sociologie de l'innovation et de la sociologie des usages, constitue le point de départ de cette thèse, qui s'articule en deux parties : la première examine la complexité des représentations sociales et culturelles des technologies d'automatisation de la production d'information, en proposant une analyse des variables nourrissant les imaginaires technologiques des journalistes et des conditions qui permettent aux technologies d'automatisation de s'inscrire dans la perspective des dimensions culturelles de l'automatisation de la production d'informations ; la seconde soumet ces observations à l'épreuve du terrain empirique. Pour ce faire, elle s'appuie sur deux études de cas réalisées dans deux rédactions belges francophones, où deux systèmes d'automatisation de la production d'informations ont été développés en vue de soutenir un travail journalistique s'inscrivant tantôt dans le temps long de l'enquête, tantôt dans l'immédiateté des routines quotidiennes.

Dans ces deux expériences, les journalistes ont été invités à devenir des acteurs de l'innovation technologique, en s'impliquant activement dans la conception des artefacts. Ce processus, qui participe à celui de la construction sociale des systèmes d'automatisation, peut être compris comme une première forme d'usage, où vont se mettre en place des mécanismes d'ajustement réciproque entre l'objet et son environnement social (Akrich 2006, 2010). Deux cadres conceptuels et théoriques ont été mobilisés, en raison de leur complémentarité pour l'analyse du développement d'un processus d'innovation en tant que processus sociotechnique et

de celle des (non-)usages des artefacts. Le modèle SCOT ("*Social Construction of Technology*") permet d'aborder l'artefact technologique comme résultant de constructions et de compromis où vont entrer en jeu les intérêts particuliers d'agents sociaux impliqués dans une chaîne sociotechnique¹ (Bijker, Hugues & Pinch 1987). La sociologie des usages va, quant à elle, permettre d'examiner la problématique de l'usage sous l'angle de ses constructions collectives et individuelles (Jouët 1993, 2000).

Les objectifs de cette recherche sont de comprendre et d'expliquer les mécanismes de structuration des usages professionnels d'artefacts de la production automatisée d'informations, considérant leur double condition matérielle d'objets et d'outils du journalisme, mais aussi d'identifier les sources de résistance qui ne permettent pas de réaliser ces usages. Il s'agit donc, en filigrane, de comprendre les conditions permettant d'envisager une collaboration homme-machine qui soit fructueuse, dans le contexte d'une technologie véhiculant des représentations ambivalentes.

0.1 Problématique et question de recherche

L'automatisation de la production d'informations journalistiques connaît un développement continu depuis le début des années 2010. Ce phénomène s'explique par la conjugaison de trois variables : celle de technologies de l'information et de la communication en évolution constante qui constituent le moteur d'innovations (Hammond 2017), celle d'une mise en disponibilité de volumes de données de plus en plus importants qui participent à la "datafication" de la société (Loosen 2017), et celle d'un tournant quantitatif entrepris dans le monde du journalisme depuis la popularisation du journalisme de données dans les années 2000 (Dagiral & Parasio 2013, Coddington 2014). Unanimement saluées pour leurs performances (Graefe 2016, Fanta 2017, Leppänen *et al.* 2017) – rapidité de traitement d'importants volumes de données, couverture en temps réel, production de contenus à grande échelle, possibilité de générer différentes formes de représentations visuelles à partir d'un même jeu de données, multilinguisme ou encore extension des zones de couverture médiatique à des sujets peu ou pas traités en raison de moyens humains limités –, les technologies d'automatisation de la production d'informations ne sont pas dénuées de limites. La nécessité de disposer de données structurées répondant à des exigences de qualité, tant sur le plan technique que sur le plan journalistique (Dierickx 2018), explique pourquoi les domaines d'application couverts sont actuellement limités au sport, à l'économie, aux résultats d'élections ou à l'environnement. Leur forte dépendance à l'expertise d'un domaine d'application explique également qu'il s'agisse souvent d'artefacts uniques et non reproductibles (Linden 2017).

De plus, bien qu'elles soient capables de fournir rapidement des comptes-rendus factuels, ces technologies ne peuvent se substituer à la complexité du cerveau humain et à sa créativité, dès lors qu'elles sont basées sur des modèles rationnels en fonction desquels toute incertitude est

¹ Une chaîne sociotechnique implique, à l'inverse d'un réseau, un début et une fin. Elle est faite d'attachements techniques et sociaux hétérogènes, où s'agencent des rôles et actions entre humains et non-humains (Denis 2009).

rejetée du système (Latar 2018 :24). Elles sont le plus souvent utilisées pour la diffusion d'informations qui ne font pas l'objet d'un traitement journalistique, que ce soit via les pages web ou les réseaux sociaux d'un média d'information (*Le Monde* en France, *Yle* en Finlande, *The Washington Post* aux États-Unis) ou encore via les flux d'informations délivrés par les agences de presse (Associated Press aux États-Unis, NTB en Norvège). Mais elles le sont aussi en tant qu'outils destinés à un travail journalistique relevant d'une approche par données (Press Association et Urbs Media au Royaume-Uni), ou à alerter les journalistes sur des sujets présentant un intérêt potentiel (MittMedia en Suède).

La prise en charge de tâches routinières ingrates (Carlson 2014) est l'un des principaux arguments mis en avant pour justifier leur introduction dans une rédaction. Pour autant, cette manière d'aborder le phénomène n'est pas sans remettre en question la définition du journalisme : peut-il être considéré sous le seul prisme d'une somme de tâches et de compétences permettant leur exécution (Van Dalen 2012) ? Bien plus qu'une activité occupationnelle, le journalisme, c'est aussi une culture et une idéologie professionnelles qui englobent une série de valeurs, de stratégies et de codes formels communément partagés, en dépit de la diversité des contextes de production (Deuze 2005). La définition du rôle du journaliste n'est donc pas seulement interrogée : celles de son identité, de son autorité et du sens même de son métier d'informer le sont tout autant (Neveu 2010). Si, en tant que travailleurs, les journalistes ont toujours composé avec les évolutions technologiques (Hardt 1990), les technologies d'automatisation de la production d'informations impliquent une nouvelle remise à plat de ce qu'est le journalisme : qui le fait et que fait-il ? (Primo & Zago 2015). Bien que leurs processus reposent sur une exécution automatique d'algorithmes, ces technologies ne peuvent être abordées sous le seul angle de "boîtes noires" transformant des entrants en extrants : aucune technologie ne peut être réduite au code informatique qu'elle renferme (McCarthy & Wright 2004). Leur développement suppose l'implication d'agents sociaux qui ne sont pas traditionnellement liés au monde du journalisme. Si certains d'entre eux se défendent de faire acte de journalisme ou de vouloir remplacer les journalistes par leurs technologies (Latar 2018 :4), ils participent désormais à la chaîne de production de l'information.

Ces acteurs sont de trois types :

- des développeurs attachés au service IT d'une rédaction, comme c'est le cas à l'agence de presse norvégienne NTB pour l'automatisation d'informations quantifiables, qu'elles soient sportives, financières ou immobilières (Fanta 2017) ;
- des journalistes disposant d'une double compétence en informatique, comme c'est le cas au *Los Angeles Times* pour l'automatisation de données sismiques (Carlson 2015) mais dont le profil, malgré sa popularisation dans la littérature scientifique, reste rare, en ce compris dans les rédactions américaines qui font pourtant figures de pionnières (Dagiral & Parasio 2011) ;
- des start-ups technologiques spécialisées dans le traitement automatique de la langue (TAL), que ce soit en fournissant un flux d'information prêt à l'emploi ou un logiciel à

paramétrer par le client. À l'échelle mondiale, on en dénombre moins d'une quinzaine (Dörr 2015). Ces sociétés proposent leurs services à une variété de secteurs, tels que la communication et le marketing. Les médias d'information ne constituent donc pas leur unique clientèle, bien qu'elle soit la plus visible.

Les technologies d'automatisation de la production d'informations mobilisent des imaginaires susceptibles d'être influencés par l'usage répandu de la métaphore du "robot journaliste", qui véhicule l'idée d'une confiscation de l'emploi, voire celle de l'identité du journaliste (Linden & Dierickx 2019). Cette métaphore participe à forger un ou des imaginaires construits à partir de représentations sociales collectives et individuelles, dans le contexte d'un secteur en crise où l'emploi journalistique est fragilisé (Deuze & Witschge 2018). À l'instar des imaginaires de toute innovation technique, ils ne sont pas stables et peuvent évoluer dans le temps (Musso 2009). Plus largement, les innovations technologiques sont souvent associées au concept de progrès et à des discours utopiques et dystopiques, constitutifs de mythes et d'une idéologie technolâtre (Domingo 2008, Musso 2008). Combiné aux pratiques sociales, ceux-ci vont participer à la formation des usages, dès lors que l'on ne peut faire usage d'une technologie sans se la représenter (Musso 2009, Flichy 2001, Massit-Folléa 2002). S'ils se réfèrent indifféremment à une utilisation, une pratique ou à une appropriation, les usages sont envisagés comme résultant d'une construction sociale (Jouët 2000). Un usage social désigne un mode d'usage suffisamment commun et intégré dans la vie quotidienne pour s'établir dans des pratiques culturelles existantes ou pour devenir une pratique spécifique (Millerand 2008). Un usage technique se rapporte à une utilisation fonctionnelle de la technologie. La démarche pragmatique de l'usage permet d'éprouver la matérialité de l'artefact, dans une expérience utilisateur faisant l'objet d'une double médiation technique et sociale (Jouët & Sfez 1993, Flichy 2008).

Actantes d'un processus éditorial, les technologies d'automatisation vont se trouver au centre de jeux de médiation et de traduction qui les impliquent activement dans la construction du sens (Latour 2005). Ceux-ci peuvent être abordées de manière multiple : sous l'angle d'une médiation technique (entre agents humains et non-humains), sous celui de la médiation d'un artefact technologique dont l'objet est de structurer les pratiques, ou encore sous celui d'une médiation sociale (Jouët & Sfez 1993, Flichy 2008). Dès lors, l'intégration d'artefacts d'automatisation dans les usages et pratiques journalistiques ne pose pas la seule question du rapport du journaliste à l'objet. Il interroge sa relation avec les porte-paroles de l'objet technique, et il questionne la manière dont les artefacts d'automatisation sont conçus pour rencontrer les exigences journalistiques de précision, d'exactitude, de fiabilité et d'actualité (Clerwall 2014). Le journalisme, c'est donner aussi donner du sens (Delforce 1996) : dès lors, cela interroge également la manière dont les productions automatisées font sens, ainsi que celle dont ces objets du journalisme deviennent des outils destinés à s'inscrire dans les usages des professionnels de l'information. Une manière d'aborder ce phénomène est donc d'examiner comment ces technologies peuvent s'inscrire dans le cadre de leurs pratiques professionnelles. Cette complémentarité entre le journaliste et la machine peut être envisagée comme une opportunité aux prémices de la chaîne de production éditoriale : le système d'information va traiter des

faits bruts, tandis que le journaliste apportera la profondeur de la mise en contexte et de l'analyse (Thurman *et al.* 2017).

Ceci nous amène à poser la question de recherche suivante :

Considérant que l'on ne peut faire usage d'une technologie sans se la représenter, de quelle manière les représentations sociales et pratiques culturelles des journalistes vont-elles participer à la structuration des usages (et des non-usages) d'artefacts de la production automatisée d'informations lorsque ceux-ci sont envisagés en tant qu'outils du journalisme?

0.2 Cadre théorique et conceptuel

Le cadre théorique du modèle SCOT (acronyme de *Social Construction of Technology*) a servi de fil conducteur aux observations empiriques présentées dans la deuxième partie de cette thèse. Aussi appelé "modèle du tissu sans couture" (*seamless web*), il permet de comprendre en quoi un artefact technologique est le fruit d'une co-construction et de compromis, qui résultent d'intérêts particuliers parmi les agents sociaux impliqués dans un processus sociotechnique. Émanation de l'école constructiviste néerlandaise, il considère que les sciences et technologies sont des constructions sociales (Bijker, Hugues & Pinch 1987). Sa principale limite est de s'arrêter au processus de fermeture, qui correspond à la mise en usage de l'objet technique. C'est pourquoi l'analyse des données empiriques prend également appui sur la sociologie des usages, qui traite des conditions de production et de diffusion d'une technologie, ainsi que des mécanismes complexes à l'œuvre dans une construction à la fois individuelle et collective (Massit-Folléa 2002). Elle permet "*de bien cerner les pratiques à une échelle micro-sociologique*", bien que sa faiblesse consiste à ne pas souvent parvenir "*à produire une connaissance sur les grands courants qui traversent notre société, et à montrer en quoi et comment les usages peuvent y renvoyer*" (Jauréguiberry 2008). Néanmoins, en plaçant le niveau d'analyse sur l'individu et/ou sur un groupe social, y compris dans ses dimensions matérielles et symboliques, ce modèle permet d'aborder la problématique de l'usage en tant que forme de médiation sociotechnique (Flichy 2008).

Des cadres théoriques secondaires ont été ponctuellement mobilisés à différents moments de la recherche, en raison de la double nature, sociale et technique, des artefacts d'automatisation de la production d'informations : les études du journalisme et la linguistique (premier chapitre, relatif aux imaginaires technologiques), les études des bases de données informatiques et les études des logiciels (deuxième chapitre, relatif aux dimensions culturelles de l'automatisation de la production d'informations), les études des sciences et des technologies (STS) et les théories émanant de la sociologie de l'innovation (dans les deux chapitres consacrés aux deux études de cas), les méthodes d'évaluation issues du champ de la linguistique computationnelle et de celui du champ transdisciplinaire des HCI (*Human Computer Interaction*, dans le quatrième chapitre dédié à la seconde étude de cas).

Les bénéfices de cette approche multidisciplinaire résident dans les éclairages complémentaires qu'elle permet, en vue de mieux comprendre le phénomène étudié en tant qu'une pratique de médiation des savoirs participant à la construction du monde social (Feenberg 1995). Trois constats nourrissent cette démarche : (1) un artefact d'automatisation de la production automatisée d'informations véhicule des représentations duales, que l'on peut considérer en termes de menace et d'opportunité; (2) un artefact d'automatisation de la production d'informations n'existe pas en dehors d'interventions humaines qui relèvent du domaine de pratiques culturelles, dont certaines sont le fait d'agents sociaux relevant du monde de la technique; (3) à l'instar des pratiques journalistiques, les usages d'artefacts d'automatisation vont résulter d'un construit culturel et social. Les hypothèses qui ont guidé ce travail de recherche découlent de ces constats :

- **Hypothèse 1** – Les remises en question professionnelles et identitaires induites par la métaphore du robot journalisme vont être source de résistance et ne vont pas permettre de réaliser des usages journalistiques.
- **Hypothèse 2** – Les usages journalistiques vont dépendre de l'adéquation des productions automatisées avec les savoir-faire et exigences du journalisme.

0.2.1 Construction sociale de la technologie (SCOT)

Le modèle SCOT dispose que les artefacts technologiques sont culturellement construits et interprétés par leurs usagers. Ces derniers peuvent en avoir une interprétation différente en fonction de leurs référents socioculturels, qui façonnent leur normes et valeurs (Bijker, Hugues & Pinch 1989, Valenduc 2005). Ce modèle considère "*à la fois que la technologie est une production de la société, et que la société est transformée par la technologie, puisque celle-ci incorpore ces choix sociaux*" (Valenduc 2005 :59). Ce modèle théorique, qui distingue la conception de l'objet de son exploitation, aborde donc la problématique de l'usage en tant que forme de médiation sociotechnique. Il prend fin au moment de la diffusion de l'objet technique (Vinck 2005, Flichy 2008, Vinck 2012).

Le modèle SCOT est caractérisé par trois variables :

- **L'identification de groupes sociaux pertinents.** Un groupe social pertinent est composé d'agents sociaux partageant une même grille interprétative. Il adhère "*à une même représentation des problèmes et des solutions techniques*" (Valenduc 2005 :59).
- **Les acteurs d'un groupe social pertinent.** Les acteurs du système sont considérés comme exerçant "*une influence détectable*" sur d'autres. Un réseau d'acteurs peut être défini en fonction de sa taille, les acteurs opérant comme des "*forces unitaires pour influencer la structure du réseau*" (Bijker, Hugues & Pinch 1987).
- **Le cadre technologique.** Il s'agit d'un cadre cognitif partagé qui définit le groupe social. Ce cadre peut décourager certaines actions et en encourager d'autres. Il est composé de concepts et de techniques utilisés par une communauté "*pour résoudre un problème*" (Bijker, Hugues & Pinch 1987).

À partir du moment où une problématique est posée, la construction sociale de l'objet technique va se dérouler en trois étapes :

- **La flexibilité interprétative.** L'innovation technologique est envisagée comme un processus ouvert dont les résultats dépendront des circonstances sociales de développement. Il s'agit, pour le chercheur, "*d'identifier comment chaque groupe social concerné construit sa propre grille d'interprétation et ses propres représentations des solutions techniques envisageables*" (Valenduc 2005 :59).
- **La stabilisation.** Cette étape est liée à la clôture des débats autour de la conception et du design de l'artefact technologique (Doray 2015). C'est à ce stade que s'impose le cadre technologique.
- **La fermeture.** Elle est constatée lorsque l'artefact technologique a été adopté par le groupe social pertinent et fait l'objet d'un consensus, que ce soit sur le plan rhétorique ou pratique. Peu d'indications méthodologiques sont fournies sur cette dernière étape, où l'importance est accordée aux utilisateurs de l'artefact technologique : "*ils modifient le produit, inventent de nouveaux usages et développent de nouvelles compétences*" (Valenduc 2005 :61).

Le modèle SCOT considère qu'un artefact technologique peut être physique ou non-physique, et qu'il "*fonctionne comme le composant d'un système qui interagit avec d'autres objets. Si un composant est supprimé ou si ses caractéristiques changent, les autres objets du système sont modifiés en conséquence*" (Bijker, Hugues & Pinch 1987 :51). Le postulat est celui d'une interaction constante entre l'artefact et son environnement. Ce modèle théorique permet de comprendre comment se construit l'objet technologique, tant dans ses succès que dans ses échecs. Pour autant, une technologie donnée n'émerge pas d'une logique prédéterminée : différents cheminement sont possibles et chacun d'entre eux peut donner lieu à des implications différentes pour les groupes sociaux étudiés. De plus, les étapes de stabilisation et de fermeture ne sont pas irréversibles (Doray 2015).

Ce modèle théorique n'est donc pas dénué de limites. Edge et Williams (1996) le critiquent essentiellement pour deux raisons : la difficulté d'aboutir au processus de fermeture, et des interprétations du modèle qui peuvent sembler infinies. D'autres critiques portent sur des processus qui seraient rarement compris, en raison de leur redesign permanent (Fulk (1993), ainsi que sur une vision segmentée de la société où tous les groupes sociaux sont considérés comme égaux (Klein & Kleinman 2000). Une critique supplémentaire porte sur la réduction des choix technologiques à des choix sociaux, rendant le modèle très conceptuel (Valenduc 2005). Si sa force est de s'intéresser aux controverses sur l'usage des techniques (Flichy 2008), il reste insuffisant lorsque l'on aborde la problématique des usages, qu'ils soient sociaux ou techniques. C'est pourquoi un second cadre théorique a été mobilisé dans cette thèse. Il propose des clés d'analyse complémentaires, en vue de mieux comprendre en quoi les usages sont protéiforme et en quoi ils relèvent d'une dynamique dont la complexité tient aussi dans sa multidimensionnalité.

0.2.2 Sociologie des usages

L'école française de la sociologie des usages s'inspire du courant des *Science & Technology Studies* (STS) où l'usage est envisagé à la fois du point de vue du concepteur de l'innovation technologique et de celui de l'utilisateur. Contrairement aux études américaines consacrées à la communication de masse, elle a moins privilégié les approches quantitatives que qualitatives (von Pape & Martin 2010). Les études sur les usages s'inscrivent dans la perspective de la sociologie de l'innovation et de la sociologie de la traduction, où l'utilisateur est défini comme un acteur s'inscrivant dans un réseau caractérisé par des interactions et des jeux de pouvoir. Cette approche s'est développée "en l'absence de références théoriques constituées et de modèles à appliquer, dans une effervescence de bricolage intellectuel et d'artisanat conceptuel" (Jouët 2000). Les technologies de l'information et de la communication (TIC) présidèrent à la naissance de ce courant de recherche, qui a porté sur une diversité de technologies. Il fut notamment examiné à partir d'études sur les pratiques du grand public, décrivant ce que les gens font avec des objets techniques, tels que le minitel, le magnétoscope ou le répondeur téléphonique (Proulx 2015).

Tout au long des années 1980, la recherche française sur les usages se caractérise par des approches hétérogènes qui articulent usage d'un objet technique, pratiques quotidiennes d'un individu ou d'un groupe social, représentations de la technique, et contexte social, culturel et politique (Proulx 2015). Ces approches ont pour point commun d'envisager l'usage en tant que construit social (Jouët 2000). En ce sens, elles sont à distinguer des théories fonctionnalistes des "Uses and gratifications", qui se sont développées dans le courant des années 1970 aux États-Unis, et qui se concentraient sur des études sur les usages des médias par les audiences, en fonction de leurs besoins ou désirs (Flichy 2008). Elles s'opposent également à une vision déterministe, en distinguant les usages fonctionnels des usages sociaux (Granjon & Denouel 2011). La recherche en sociologie des usages s'attache alors à définir "la lutte pour l'alphabétisation de l'informatique" et "l'appropriation sociale des technologies" comme de possibles sources d'autonomie sociale et politique (Proulx 2015). Jusqu'au début des années 1990, elles se distancieront des études anglo-saxonnes relevant des champs des *Media Studies* (qui s'intéressent aux contenus et aux effets des médias sur les audiences), des *Cultural Studies* (un courant de recherche caractérisé par sa transdisciplinarité), de l'ergonomie (*Human factor*, qui étudie les relations entre l'homme et la machine dans un contexte professionnel), et des Interactions Homme-Machine (*Human-Computer Interaction, HCI*).

Au milieu des années 1980, la théorie de la structuration adaptative (TSA), développée par des chercheurs en management de l'information, s'intéressera aux modes d'appropriation directs ou indirects, aux usages instrumentaux, aux attitudes affichées envers les outils mis en place, et au respect de l'esprit de la technologie qui renvoie au cadre de fonctionnement prévu par les concepteurs (Coutant 2015). Cette dernière approche rencontre la sociologie des usages telle qu'elle s'est développée en France. À partir du milieu des années 1990, la recherche en sociologie des usages va prendre une orientation ethnographique qui mobilise, sur le plan méthodologique, cinq niveaux d'analyse : (1) le suivi d'un utilisateur dans son face à face technologique

(interaction homme-machine); (2) le suivi des actions de coordination entre le concepteur et l'utilisateur; (3) le suivi de la trajectoire de l'objet prescripteur, de sa conception à sa mise sur le marché; (4) la description fine et détaillée de la situation d'usage; (5) l'étude de l'ancrage collectif et historique des usages dans des séquences structurelles constituant des formes socio-historiques de l'usage (Proulx 2015). Ces nouvelles orientations auront pour conséquence une modification de la définition de l'utilisateur, désormais envisagé comme un agissant autonome investi de compétences techniques, mais placé sous contrainte dans une situation de travail : il s'agit donc d'un acteur en situation (Proulx 2015).

Utilisation, usages et pratiques

Les termes "utilisation" et "usage" ont été débattus dans la littérature scientifique où l'utilisation est définie dans l'acceptation la plus fonctionnelle et utilitaire du terme, tandis que l'usage est défini comme un processus multidimensionnel englobant à la fois des interactions entre individus et objets techniques et des trajectoires d'usage, lesquelles sont fondées sur l'histoire personnelle et sociale de l'utilisateur. Le terme "usage" peut parfois s'avérer ambigu, dans la mesure où il est indistinctement utilisé pour désigner l'emploi, l'utilisation, la pratique ou l'appropriation (Millerand 2008). L'usage peut donc prendre différentes formes (Jouët 1993) et être compris à la fois comme une pratique sociale "naturelle" et comme l'utilisation d'un objet permettant de mettre en lumière son sens social, qui est construit de manière subjective par l'acteur-utilisateur (Proulx 2005).

Les usages sociaux sont définis "*comme les patterns d'usages d'individus ou de collectifs d'individus (strates, catégories, classes) qui s'avèrent relativement stabilisés sur une période historique plus ou moins longue, à l'échelle d'ensembles sociaux plus larges (groupes, communautés, sociétés, civilisations)*" (Proulx 2005). Jouët (1993) différencie l'usage, qui renverrait à "*une simple utilisation*", de la pratique, "*une notion plus élaborée qui recouvre non seulement l'emploi des techniques (l'usage) mais les comportements, les attitudes et les représentations des individus qui se rapportent directement ou indirectement à l'outil*". En tant que construit social, l'usage va dépendre de la représentation qu'ont les utilisateurs de l'objet technique : en ce sens, le concept d'usage renvoie à celui de pratique (Chambat 1994). Flichy (2008) distingue trois principes en matière de recherche sur les usages : (1) la multidimensionnalité de l'usage technique, qui "*renvoie aussi bien à l'interface avec la machine, aux représentations sociales de la technique, au positionnement dans l'espace et le temps de la vie quotidienne*"; (2) l'impossibilité d'étudier l'usage en dehors de l'objet technique; (3) la prise en considération d'un usage non pas individuel mais collectif, dès lors qu'il possède une dimension sociale.

Un usage social fait référence à un mode d'utilisation suffisamment habituel et suffisamment intégré dans le quotidien pour s'intégrer et s'imposer dans des pratiques culturelles existantes ou devenir une pratique spécifique (Lacroix 1994, cité par Millerand 2008). Un usage technique fait référence à un usage utilitaire de la technologie. Ces formes d'usage s'inscrivent dans un cadre sociotechnique qui repose sur un imaginaire technique mobilisé par des acteurs et qui

visé à construire un cadre de référence qui forgera leurs actions. En ce sens, les discours sur les usages témoignent de leurs représentations, lesquelles renvoient aux concepts de l'imaginaire technologique et de l'imaginaire social (Jouët 1993, Musso 2009). Ces représentations ne sont pas exemptes de mythes, dans la mesure où la technologie peut être aussi bien associée au pouvoir, au savoir, à la mémoire, à la prospérité économique, ou encore au lien social (Musso 2009, Jarrige 2014). Comme toute action humaine, l'action technique "*ne peut pas exister sans prendre une forme symbolique, on ne peut ni concevoir, ni utiliser une technique sans se la représenter*" (Flichy 2001). Mais si l'usager se représente les fonctionnalités de l'objet technique, cette représentation individuelle s'inscrit dans un contexte social plus large qui véhicule, lui aussi, des représentations. L'inter-influence entre ces différents modes de représentation va agir sur la matérialité des pratiques des individus (Proulx 2006). À côté de la représentation individuelle, coexiste une représentation collective qui serait partagée par un même groupe social. Ces représentations – qui relèvent de l'imaginaire technologique (Flichy 1995) – participent à la construction d'une identité collective (Flichy 2001). Rendre compte des dispositifs techniques revient donc à restituer la complexité des représentations que s'en font les acteurs/utilisateurs (Akrich 1993b).

Cadre sociotechnique

Le cadre sociotechnique est constitué d'un cadre de fonctionnement (désignant les savoirs et savoir-faire mobilisés ou mobilisables, fruit d'un consensus entre technique et social) et d'un cadre d'usage (relatif à l'utilisation de l'artefact technologique et donc à son usage social) (Flichy 1995, 2008). Cet emprunt au modèle SCOT² défend l'idée selon laquelle "*dans un processus d'innovation, divers groupes sociaux se construisent chacun une représentation de la technique et de ses usages, lequel constitue le cadre technologique de référence*" (Perine 2010). L'innovation technologique naîtra de la confrontation entre groupes sociaux. Les cadres de fonctionnement et d'usage ne peuvent être envisagés sans tenir compte de considérations d'ordre représentationnel. Flichy (1995) envisage le cadre sociotechnique selon quatre variables : (1) l'articulation entre technique et société, (2) la place centrale des usagers, (3) une recherche s'appuyant sur l'action et non le fait, (4) la possibilité d'interactions à condition d'une stabilité dans le rapport entre les acteurs et l'objet technique. Ce cadre, mobilisé pour percevoir et comprendre les phénomènes, est commun aux différents acteurs, sans être nécessairement unique.

Le cadre de fonctionnement renvoie à un usage fonctionnel de la technologie, et il va dépendre des savoirs et savoir-faire de l'usager. Son objet est de définir ces derniers, et d'indiquer le sens social de la technologie (Flichy 2000). Un cadre d'usage décrit "*le type d'activités sociales proposées par la technique, qui la positionne dans l'éventail des pratiques sociales, des routines de la vie quotidienne, et précise les publics envisagés, les lieux et les situations où cette technique peut se déployer*" (Flichy 2000). Il n'est pas indépendant du cadre de fonctionnement, même si celui-ci peut évoluer sans influencer sur les usages. Par ailleurs, un cadre de fonctionnement peut correspondre à plusieurs cadres d'usage.

² Voir "Construction sociale de la technologie (SCOT), p.13

La fonction du cadre d'usage est à la fois cognitive et symbolique, tout en permettant "*d'organiser les interactions des acteurs avec l'objet technique et des acteurs entre eux*" (Flichy 2008). La définition du cadre d'usage n'est pas assurée par l'utilisateur : elle est le résultat de l'action conjointe de l'ensemble des acteurs de l'activité technique, qui en forme sa représentation. Le cadre d'usage, constitué de discours et de représentations, se construit dans le temps et trouve sa stabilité à l'issue de controverses dont il fait l'objet. Flichy (1995 :153) n'envisage donc pas les solutions techniques comme universelles et naturelles, mais comme résultant de constructions dans un jeu complexe d'interactions. L'expérimentation constitue une étape importante dans la constitution du cadre d'usage car les cadres de référence y sont encore suffisamment faibles pour pouvoir évoluer.

Flichy (1997 :219) emprunte également au modèle SCOT la notion de stabilisation de l'objet technique. Elle est constatée lorsque les acteurs techniques "*ont réussi à créer un alliage entre cadre de fonctionnement et cadre d'usage*". Le cadre sociotechnique n'est pas la somme de ces deux cadres, mais constitue une nouvelle entité. L'essentiel du cadre sociotechnique sera alors considéré par les agents sociaux comme une "boîte noire". Le processus d'innovation est ici entendu comme un processus de stabilisation des relations entre acteurs et objet technique. Toutefois, ce n'est pas parce qu'il n'utilise pas la technique qu'un non-utilisateur ne fait pas partie prenante du cadre d'usage. De plus, si le cadre d'usage se transforme, "*le non-utilisateur peut modifier sa position et devenir utilisateur*" (Flichy 2008).

Interactions entre une innovation et ses usagers

Pour Akrich (1990), les interactions avec les utilisateurs d'une innovation sont essentielles, en raison du caractère indissociable de la technique et du social : "*Les usages ne sauraient se déduire mécaniquement des choix effectués lors de la conception, pas plus qu'il n'est possible d'affecter telle ou telle signification à un objet indépendamment de son contenu technique*". Elle propose d'envisager la description de l'objet technique comme un scénario, un script "*définissant les espaces, des rôles et des règles d'interactions entre les acteurs (humains et non-humains) qui viendraient incarner ces rôles*". Dans cette optique, toute prise de décision en cours de conception "*opère un partage des compétences et des attributions entre l'objet, son utilisateur et un ensemble de dispositifs techniques et sociaux qui constituent leur environnement*". Toute décision technique engage une représentation de ce que sont, peuvent et veulent les utilisateurs supposés du dispositif : elle établit, de ce fait, un cadre limitatif aux différents scénarios d'interactions avec les usagers. Aussi, Akrich considère-t-elle les utilisateurs comme des acteurs de l'innovation. Les inclure au moment de la conception permet de ne pas les surreprésenter (Akrich 2006), mais aussi de développer une première forme d'usage visant à définir les fonctionnalités du dispositif et à orienter les choix techniques.

Akrich (1990) décrit quatre types d'interventions des utilisateurs d'une innovation : (1) le déplacement de l'usage, pouvant se rapporter à une technologie flexible ou incomplète; (2) l'adap-

tation, qui se rapporte à une technologie faisant l'objet de modifications pour être ajustée aux caractéristiques de l'utilisateur ou de son environnement; (3) l'extension, lorsque plusieurs éléments ajoutés à l'artefact permettent d'enrichir la liste de ses fonctions; (4) le détournement, lorsqu'un utilisateur se sert d'un objet technique à des fins autres que celles pour lesquelles il a été conçu. Mais, critique Flichy (2008), cette approche se situe davantage du côté de l'usabilité que de celle des usages. Dans le domaine des TIC, la co-construction est également entendue comme une forme d'usage, la figure de l'usager étant ici abordée sous l'angle du rapport entre concepteur et utilisateur. L'usager peut aussi être compris comme un consommateur, ou encore comme s'inscrivant dans le cadre plus large de la relation homme-machine (Boutet & Trémembert 2009). En analysant les technologies et processus d'innovation sous le prisme des usages et usagers, Akrich démontre qu'il n'y a pas de technique à l'état pur. Cette posture permet d'établir une symétrie entre l'innovation et ses usagers (Denis 2009). Akrich (1993b) soulignera également l'importance du contexte d'usage, la technologie ne revêtant un sens qu'à partir du moment où elle est mise en relation avec son environnement. Il n'est donc pas possible d'envisager l'usage sans tenir compte de son contexte (Akrich 1993).

Logiques d'usages

La recherche en sociologie des usages a envisagé l'usage sous les prismes de l'innovation, de la diffusion, de l'appropriation et de l'impact sur la construction de l'identité personnelle ou sociale (Millerand 2008). La logique d'usage résulte d'une confrontation entre l'instrument, sa fonction et le projet de l'utilisateur (Perriault 1989, cité par Pape & Martin 2010). Cette légitimation induite par l'utilisateur peut être mise en relation avec les travaux de Caradec (2001), qui proposent quatre logiques d'usages : (1) utilitaire, qui consiste à porter une appréciation sur l'utilité de l'objet; (2) identitaire, qui permet d'expliquer l'usage/non-usage par l'adéquation/inadéquation de l'objet avec l'individu; (3) de médiation, où l'usage ou le non-usage est fonction de l'intervention d'un tiers; (4) d'évaluation, où un jugement est posé sur l'objet technologique, ses caractéristiques et ses performances en fonction d'images négatives ou positives. Ces logiques peuvent se combiner entre elles et produire des discours mixtes parmi les usagers et les non-usagers (Pape & Martin 2010).

Appropriation

Les travaux de Michel de Certeau et ses interrogations sur les pratiques quotidiennes ont nourri une approche orientée sur l'appropriation, laquelle donne une place active à l'usager en termes de de détournement, de non-usage ou de rejet (Proulx 1994). Il a défini quatre temps dans l'acte d'usage, faisant écho à l'acte d'énonciation : réaliser, s'approprier, s'inscrire dans des relations, et se situer dans le temps. Ce faisant, il place l'acte d'usage dans le contexte duquel il se produit et l'usager au rang de pratiquant. La pratique est ainsi confrontée "*avec des objets, des espaces, des textes qu'elle n'a pas elle-même produits, avec lesquels elle doit composer*" (Jeanneret 2007).

Roux (2007) définit l'appropriation comme la manière dont un groupe utilise, adapte et reproduit une structure : "*Etudier la structuration nécessite donc de se focaliser sur les processus d'interaction dans le groupe, interactions essentielles qui nous permettent d'abandonner la dichotomie traditionnellement adoptée entre usages prescrits et usages réels, de considérer les éventuels 'écarts' comme partie intégrante des processus d'appropriation, et par conséquent, le changement comme un état normal de l'usage*". Pour Proulx (2005), il s'agit d'un "*processus d'intériorisation progressive des compétences techniques et cognitives à l'œuvre chez les individus et les groupes qui manient quotidiennement ces technologies*".

L'appropriation – ou l'acte de se constituer un "soi" (Jouët 2000) – se distingue d'une approche de l'innovation, dans la mesure où elle ne s'intéresse pas à la conception de l'objet technique mais bien à sa "mise en usage", renvoyant à une analyse placée du point de vue de l'utilisateur mais aussi au contexte de l'usage. L'utilisateur n'est pas envisagé comme un acteur unidimensionnel : à chaque utilisateur correspondront des référents socioculturels qui vont dépendre de son histoire personnelle. L'acte d'appropriation amène ainsi à considérer une série de concepts antagonistes : banalisation/idéalisation, hybridation/substitution, évolution sociale/révolution sociale, identité active/identité passive (Mallein & Toussaint 1994, cités par Millerand 2008). Jouët (2000) indique que l'appropriation "*se construit dans la relation avec l'objet de communication et l'usage comporte donc de facto une dimension cognitive et empirique. Sa construction met en jeu des processus d'acquisition de savoirs (...), de savoir-faire (...), et d'habiletés pratiques*". Cela suppose un processus de négociation entre l'utilisateur et l'objet technique. De plus, l'appropriation se joue aussi sur le terrain de la construction de l'identité personnelle et sociale de l'individu. L'intensité de l'usage constitue un facteur d'identification forte, cristallisant les enjeux de l'identité, professionnelle y compris. De plus, dans un contexte professionnel, l'arrivée d'une technologie ne bouleverse pas seulement le mode de travail, mais aussi "*les cultures de métier qui fondent l'identité du travail*" (Jouët 2000).

Une appropriation individuelle désigne un processus par lequel un utilisateur va intégrer l'objet technique à son quotidien, tout en l'adaptant à sa personnalité et à ses besoins; tandis qu'une appropriation collective porte sur la mise en œuvre du dispositif technique par un groupe ou une catégorie sociale "*dans le but d'accroître son autonomie ainsi que sa capacité à agir*" (Proulx 2015). Ces deux formes d'appropriation sont liées, dès que lors qu'une représentation collective est susceptible d'influer sur une appropriation individuelle (Latzko-Toth & Proulx 2015). Proulx (2005) pose quatre conditions pour qu'il y ait appropriation d'un objet technique : (1) sa maîtrise technique et cognitive; (2) son intégration significative dans les pratiques de l'utilisateur; (3) son usage répété permettant des possibilités de création; (4) son appropriation sociale, à condition que les usages soient correctement représentés dans les processus. Il propose également cinq niveaux d'analyse : (1) l'interaction entre le dispositif technique et l'utilisateur; (2) la coordination entre l'utilisateur et le concepteur du dispositif; (3) la situation d'usage dans un contexte de pratiques; (4) l'inscription des dimensions politiques et morales dans le design de l'objet technique; (5) l'ancrage social et historique des usages dans un ensemble de macrostructures (discursives, culturelles...) qui en constituent les formes.

Dans le contexte des TIC, les activités de "bricolage" – phase où l'acteur manipule l'artefact technologique pour le rendre conforme à ses attentes et besoins professionnels – participent au processus d'appropriation (Comtet 2009). L'usage est considéré comme efficace pour autant que les utilisateurs aient la possibilité de le détourner. Mais il est aussi dual car il met en jeu les dimensions subjectives et objectives de l'acteur du système. L'adoption est également considérée comme un préalable à l'appropriation (Boudokhane 2006). Celle-ci constitue un élément de la construction identitaire des usages, lesquels vont se jouer à différents niveaux : symbolique, individuel et collectif (Flichy 2008).

Non-usages

Le concept de non-usage peut renvoyer autant à la non-utilisation, la non-adoption (non-consommation) et la non-appropriation (absence de maîtrise technique et cognitive de l'artefact). Il consiste en une non-réalisation d'un scénario, dont la responsabilité ne peut être uniquement attribuée à un utilisateur incompetent ou à un objet mal conçu (Akrich 1990). Un non-usage peut être caractérisé par un refus ou une résistance culturelle (Proulx 1994, Boudokhane 2006). Kellner *et al.* (2010) envisagent quant à eux trois autres facteurs de non-usages : l'absence de besoin, le manque d'intérêt et de motivation, et le manque de sens. D'autres facteurs cognitifs peuvent encore expliquer un non-usage (Selwyn 2003) : la technophobie, relative à l'appréhension d'une technologie et à l'anxiété ou à la peur que celle-ci provoque chez l'individu, laquelle aurait tendance à s'estomper au fur et à mesure que la technologie devient de plus en plus courante; le refus idéologique, lié à une forme de non-conformisme ou à une résistance envers les conséquences que pourrait induire la technologie sur des schémas traditionnels de travail et de vie; une adoption progressive de la technologie induisant les concepts d'utilisateurs précoces et d'utilisateurs retardataires.

Cette approche peut être mise en relation avec la courbe d'adoption issue de la théorie de la diffusion de l'innovation développée par Rogers, en 1962, en fonction de cinq profils d'utilisateurs : enthousiastes (les innovateurs), visionnaires (les adopteurs précoces), pragmatiques (la majorité précoce), conservateurs (la majorité tardive), sceptiques (les retardataires). Cette théorie pose également l'hypothèse qu'un manque de connaissance et/ou des attitudes hostiles seraient à l'origine du non-usage d'une innovation technologique (Boudokhane 2006). Pape & Martin (2010) ajoutent que, selon Rogers, les facteurs qui influencent l'utilisation d'une innovation vont dépendre de l'adoptant potentiel, du système social dans lequel il s'inscrit, et de la compatibilité entre l'innovation et ces deux paramètres. Elle peut également s'inscrire dans la perspective du modèle de diffusion de Bass (1980), selon lequel l'adoption d'une innovation dépend d'un nombre d'adoptants suffisant pour atteindre une masse critique, dans un double phénomène de contagion (liée au nombre d'adoptants, qu'ils soient innovateurs ou imitateurs) et de saturation (liée à la taille du marché).

Le concept de non-usage fait écho au temps long de la formation des usages, qui va relever de mécanismes sociaux et individuels complexes, parfois ambigus, qui rendent impossible d'envisager l'usage sous le prisme dichotomique utilisateur/non-utilisateur. C'est pourquoi Livingstone & Helsper (2007, cités par Pape & Martin 2010) proposent d'établir un *continuum* avec des gradations du non-usage en passant par l'usage réduit jusqu'à un usage plus fréquent. Les non-usagers pourraient également être classés en fonction de quatre types de comportements : un abandon volontaire faisant suite à une mauvaise expérience, un abandon volontaire par déficience, un non-accès par manque de moyens, et un refus d'utilisation (Wyatt *et al.* 2002, cités par Jauréguiberry 2012). Les situations de non-usage ne sont donc pas uniques mais plurielles, et elles ne peuvent pas toujours s'expliquer par un désintérêt, un refus ou une résistance.

Médiations et construction des pratiques

Les objets techniques sont des moyens de médiation du social (Denis 2009) et ils n'ont de sens que dans leur mise en relation entre l'humain et son environnement (Akrich 1993). Toutefois, les pratiques de médiation sont doubles : elles sont à la fois techniques, considérant que la médiation d'un artefact technologique va structurer les pratiques ; et sociales, car la médiation va donner du sens à la pratique (Jouët 1993, Flichy 2008). En ce sens, les pratiques de médiation apparaissent comme un moyen de favoriser (ou non) l'expérience utilisateur. Mais si elles sont autant techniques que sociales, elles ne peuvent être considérées comme neutres, dans la mesure où la matérialité de l'objet infiltre les pratiques et où aucune pratique n'est dénuée de projections sociales. La pratique consiste en une "*production de soi qui ne prend sens que dans et par le social*" (Jouët 2000).

Dans cette perspective, technique et social se trouvent tous deux en situation structurée et structurante (Jouët 1993, citée par Coutant 2015). Flichy (1991, cité par Pélissier & Romain 1998) soulignera qu'il n'y a pas de séparation radicale entre la construction technique de l'objet et sa construction sociale, qui consiste en une construction collective qui prend forme à travers la circulation de l'objet technique. Flichy, qui s'inscrit davantage dans la mouvance interactionniste, considère qu'un système qui ne rencontre pas les besoins d'un utilisateur ne peut rencontrer les usages (Flichy 1995).

0.3 Plan de la thèse

Anderson (2013) souligne que les premières recherches traitant du journalisme computationnel ou algorithmique se sont davantage concentrées sur un point de vue internaliste, c'est-à-dire l'envisageant selon le point de vue de la profession journalistique. Il estime toutefois qu'une approche interdisciplinaire et externaliste est pertinente à développer, d'autant que les possibilités sont multiples telles que la dynamique organisationnelle et le processus de travail qui facilitent ou entravent l'adoption de technologies informatiques dans une rédaction ; l'analyse culturelle et historique, fondée sur la compréhension du rôle de la culture journalistique dans une perspective sociomatérielle nuancée ; et l'étude des relations entre technologies numériques et la "prise de sens" journalistique. Ces trois propositions sont explorées dans cette

thèse, de manière transversale, à travers l'étude des représentations sociales de la technologie et des pratiques culturelles de l'automatisation de contenus journalistiques (première partie, axe macrosociologique) et d'expériences d'usages d'artefacts de la production automatisée d'informations (deuxième partie, axe microsociologique). Il s'agit donc, dans un premier temps, de placer la problématique des usages dans ses contextes sociaux et culturels, desquels elle ne peut être dissociée. Dans un second temps, il s'agit d'examiner comment vont se structurer les usages lorsqu'une technologie d'automatisation de la production d'informations est développée dans une rédaction pour soutenir des pratiques professionnelles³.

0.3.1 Première partie : représentations sociales et pratiques culturelles

La première partie s'inscrit dans le cadre d'une analyse du contexte socioprofessionnel au sein duquel se développent les technologies d'automatisation de la production d'informations. Il s'agit d'une analyse macrosociologique qui s'inscrit dans la perspective d'une approche par données dans le journalisme, considérant que les technologies d'automatisation de la production d'informations constituent un avatar du "journalisme computationnel" ou du "journalisme algorithmique", qui peut être défini comme "*une combinaison d'algorithmes, de données et de la connaissance des sciences sociales pour compléter la fonction de responsabilité du journalisme*" (Hamilton & Turner 2010). En raison du caractère récent du phénomène, et dans le souci de tenir compte de sa double nature à la fois technologique et journalistique, notre approche est placée sous le signe de la transdisciplinarité. Cette posture épistémologique s'inscrit dans le prolongement du tournant matériel pris dans les études sur le journalisme à l'aune des années 2000 (Boczkowski 2014). Elle va permettre de préparer la démarche de recherche empirique, en vue de placer les résultats des observations de terrain dans un contexte plus général. Elle va notamment mettre l'accent sur les relations ambiguës que les journalistes entretiennent avec le monde de la technique, ainsi que sur les nouveaux acteurs des processus éditoriaux dont une particularité est ne pas être traditionnellement liés au monde du journalisme.

Le premier chapitre est consacré à l'étude des imaginaires technologiques des journalistes, en tant qu'éléments constitutifs des usages. Il comporte quatre sections, dont la première dresse un état de l'art sociohistorique qui a pour objet d'éclairer le présent avec les enseignements du passé. L'objet d'étude est ici mis en question via la mobilisation d'un corpus de sources académiques (Buton & Mariot 2006). Cet état de l'art met l'accent sur les deux principaux ressorts qui caractérisent cette évolution : (1) une résistance au changement comme symbolisant un refus de bousculer des valeurs professionnelles, (2) une reconfiguration des processus sociaux et éditoriaux induite par le développement d'innovations technologiques, qui amène les journalistes à reconsidérer non seulement leurs compétences, mais aussi leurs spécificités professionnelles, qu'elles soient routinières ou identitaires.

³ Les sources ayant alimenté cette recherche sont de trois types : académiques, entretiens et articles de presse. Lorsqu'elles sont mobilisées, celles-ci sont présentées comme suit : "(Nom, année)" pour les sources académiques, dans le corps du texte; "Communication personnelle, ville, date", pour les entretiens, en note de bas de page; "Titre de l'article", auteur, média, date de publication, date de consultation, URL", pour les articles de la presse en ligne, en note de bas de page. Voir également la liste des sources primaires p.318

La deuxième section de ce chapitre s'intéresse aux particularités de la presse belge francophone, terrain des deux études de cas empiriques où, au moment de débiter cette recherche, aucune rédaction ne s'était dotée d'un système d'automatisation de la production d'informations. Les imaginaires sont donc ici sondés à travers les relations entretenues vis-à-vis d'une approche par données dans le journalisme. En raison de sources académiques limitées, plusieurs entretiens ont été réalisés (en face à face ou via courriel) avec des acteurs actifs dans ce domaine et/ou dans celui de l'innovation technologique : Joël Matriche (journaliste spécialisé dans le traitement de données au journal *Le Soir*), Arnaud Wéry (journaliste au weblab de *L'Avenir*), Cédric Dussart (développeur-journaliste au weblab de *L'Avenir*), Nicolas Becquet (responsable du développement et de la transformation numérique à *L'Echo*), Philippe De Camps (news-manager à l'agence de presse Belga), Mehmet Koksal (journaliste indépendant, membre de la Commission d'agrément au titre de journaliste professionnel en Belgique francophone) et Arnaud Grégoire (producteur numérique indépendant, considéré comme l'un des pionniers du journalisme web en Belgique francophone).

Les enjeux professionnels de l'automatisation de la production d'informations journalistiques sont examinés dans une troisième section. Ici, il s'agit de s'intéresser aux principaux défis posés par ce phénomène, dans la mesure où ils participent à forger un ou des imaginaires collectifs. À partir de la littérature scientifique, trois enjeux majeurs sont analysés : celui de la gestion des angoisses professionnelles, celui de l'équilibre à (re)trouver dans la relation homme-machine, et celui de la (re)mobilisation des compétences.

Une quatrième et dernière section examine, de façon empirique, les imaginaires collectifs associés à la production automatisée d'informations. Ils sont abordés dans une analyse du discours médiatique des journalistes, lorsque ceux-ci sont amenés à traiter ce sujet dans le cadre de leurs activités professionnelles. Cette analyse du discours se fonde sur une approche par corpus s'intéressant aux seuls titres de presse, considérant que tout article journalistique est introduit par un titre qui en constitue la partie la plus visible. Leurs différents contextes d'utilisation leur confèrent un statut autonome : à la fois unité indépendante et sous-genre journalistique (Kronrod & Engel 2001, Ho-Dac *et al.* 2004). L'étude des titres de presse permet de suivre l'évolution de la langue, d'identifier les tendances et de définir les caractéristiques d'un groupe culturel donné (de Villers 2001). Le corpus examiné comporte 300 titres d'articles publiés en ligne, en français et en anglais, sur une période de huit ans (2010-2017). Cette analyse, qui s'appuie sur les quatre niveaux d'analyse du discours (lexical, syntaxique, sémantique et pragmatique), a pour objet de décrire la manière dont les discours font sens.

Le deuxième chapitre trouve son ancrage dans la seconde hypothèse formulée dans cette thèse, où se pose la question de l'adéquation des pratiques culturelles du journalisme avec les artefacts d'automatisation de la production d'informations. Consacré aux dimensions culturelles de l'automatisation de la production d'informations, il est abordé sous l'angle des trois dimensions constitutives de la culture du journalisme (Singer 2003) : cognitive, relative aux connaissances et aux techniques; normative, relative à l'éthique et à l'engagement du journalisme en

tant que bien commun ; et évaluative, qui aborde la manière dont les professionnels de l'information se perçoivent et sont perçus. Ce chapitre comporte trois sections, qui correspondent à chacune de ces trois dimensions.

La première section de ce chapitre analyse la dimension cognitive caractérisant le domaine de la production automatisée d'informations, où les challenges ne sont pas seulement journalistiques : ils sont aussi techniques. Comme dans toute approche par données dans le journalisme, qu'elle soit automatisée ou non, la problématique de la qualité des données va apparaître comme fondamentale. En effet, il est généralement admis que seule une donnée de qualité peut donner lieu à une information de qualité (Batini & Scanniapeco 2006) : "*si les données sont mauvaises, les histoires seront mauvaises*" (Lisa Gibbs, Associated Press, citée par Monti 2019). L'étude des bases de données informatisées permet de définir les exigences techniques qu'il convient de rencontrer, dès lors que des données sont destinées à être automatisées dans un contexte journalistique. Elle démontre la variabilité du concept de qualité ainsi que des préoccupations partagées, sur un plan théorique, avec le monde du journalisme, où le concept de qualité est tout aussi complexe et difficile à définir. Dans un second temps, il s'agit de voir en quoi les exigences journalistiques et techniques peuvent être rencontrées dans un projet d'automatisation de la production d'informations. Considérant que la qualité sera fonction de son domaine d'application et que celle-ci ne peut faire l'objet d'une définition formelle en raison de son caractère multidimensionnel, il s'agit de définir comment les challenges technique et journalistique peuvent être rencontrés, à travers le développement d'un modèle d'évaluation conceptuel pouvant être envisagé en amont de la conception d'un système d'automatisation de la production d'informations.

Bien que le volet relatif à la qualité des données soit essentiel, il n'est pas le seul à devoir être pris en considération, dès lors que le domaine d'application est également caractérisé par des processus d'analyse et de traitement de ces données. Ceux-ci ne peuvent être considérés comme indépendants de toute action humaine. Cela implique que la "boîte noire" technologique puisse être comprise à travers la succession de choix humains qui la caractérise, et qui ne sont d'ailleurs pas très éloignés d'un processus éditorial traditionnel. Cette section interroge la manière dont les productions automatisées peuvent faire sens, considérant qu'aucune activité d'analyse de données ne peut être "neutre" ou "objective". C'est pourquoi elle aborde également les enjeux liés au traitement de données informatisées dans un contexte journalistique, tout en s'intéressant au mythe de l'objectivité, en tant qu'argument technologique et qu'idéal journalistique.

La deuxième section de ce chapitre est consacrée à la dimension normative de la culture professionnelle du journalisme, étant disposé que les productions automatisées, dès lors qu'elles sont destinées à soutenir les routines journalistiques, devraient également être compatibles avec les valeurs du journalisme. Cette compatibilité ne peut être garantie qu'à partir du moment où un cadre commun est partagé entre toutes les parties prenantes d'un processus d'automatisation que l'on considère, d'abord et avant tout, comme un processus éditorial. Ces dernières peuvent

appartenir à des mondes sociaux où la responsabilité sociale de leurs acteurs n'est pas forcément envisagée de la même manière, selon que l'on se place du côté du journalisme ou de celui de la technique. Il s'agit d'identifier les points de convergence et de divergence devant être négociés pour rencontrer les valeurs du journalisme, en s'appuyant sur une analyse croisée. La méthode repose sur une revue de la littérature scientifique ainsi que sur les chartes professionnelles dans les domaines du journalisme, de l'ingénierie, de la science des données et de l'informatique. Dans une visée opérationnelle, l'expertise d'acteurs de terrain impliqués dans l'autorégulation professionnelle en Belgique francophone a été sollicitée. Deux entretiens ont été réalisés avec des membres du Conseil de déontologie journalistique, Ricardo Gutiérrez (également secrétaire général de la Fédération Européenne des Journalistes) et Gabrielle Lefèvre (anciennement membre du conseil de direction de l'Association des journalistes professionnels et du Conseil supérieur de la Justice). Des experts du domaine de la génération automatique de textes en langue naturelle ont également été sollicités pour éclairer certains aspects liés aux technologies d'automatisation : Claude de Loupy et Helena Blancafort (CEO de Syllabs, start-up française spécialisée dans l'analyse sémantique) et Frank Feulner (CBDO AEXEA/AXSemantics, société allemande spécialisée dans la génération automatique de textes en langage naturel).

L'examen de la dimension évaluative, également constitutive de l'identité professionnelle, permet de situer la manière dont les journalistes perçoivent la qualité des productions automatisées, ainsi que celle dont les audiences font (ou non) la différence entre des productions automatisées et des textes rédigés par des journalistes humains. Elle fait l'objet de la troisième et dernière section de ce deuxième chapitre.

0.3.2 Deuxième partie : structuration des usages

À un niveau microsociologique, cette recherche s'appuie sur deux études de cas permettant d'aborder, de manière empirique, les usages d'artefacts d'automatisation de la production d'informations par des journalistes. Toutes deux s'inscrivent dans le cadre d'un partenariat : d'une part, avec le mensuel d'actualités sociales *Alter Échos* et, d'autre part, avec le groupe Mediafin, éditeur des quotidiens économiques *L'Echo* et *De Tijd*. Deux systèmes de production automatisée d'informations y ont été développés, dans la perspective de soutenir un travail journalistique s'inscrivant sur le temps long de l'enquête ("Bxl'air bot") et dans le temps court des routines quotidiennes ("Quotebot"). Ces deux études de cas forment la seconde partie de cette thèse. En sociologie des techniques, l'étude de cas est une stratégie de recherche qui traite d'un phénomène contemporain dans son contexte (Latzko-Toth 2009). Elle se caractérise par le fait qu'elle comporte une abondance de variables pertinentes, qu'elle s'appuie sur de multiples éléments de preuves qui doivent converger, et qu'elle est orientée par un cadre théorique existant. Elle permet d'observer un phénomène social à caractère universel, en instaurant un rapport d'induction-déduction entre un cadre explicatif et un cadre particulier où la description joue un rôle important. Flyvbjerg (2006) souligne que l'étude de cas est nécessaire en sciences sociales car l'étude d'une discipline dépourvue d'exemples est inefficace. Pour ce chercheur, il existe toutefois cinq grands malentendus à propos des études de cas : (1) la connaissance théo-

rique est plus précieuse que la connaissance pratique, (2) on ne peut généraliser à partir d'un seul cas, (3) elle est utile pour générer des hypothèses tandis que d'autres méthodes sont plus appropriées pour les tester, (4) elle contient un biais à la faveur de la vérification, (5) il est souvent difficile de résumer des études de cas spécifiques. Bien choisie et bien balisée, une étude de cas est profitable à la recherche : la proximité avec la réalité que l'étude de cas implique, couplée au processus d'apprentissage qui y est mis en œuvre, seront souvent des conditions préalables pour la compréhension avancée d'une problématique.

Le troisième chapitre traite des usages de l'artefact "Bxl'air bot" au sein d'une rédaction présentant peu d'appétence pour les technologies informatiques en général, ainsi que pour une approche par données dans le journalisme. Nous avons développé ce dispositif sociotechnique dans le contexte de cette recherche, dans une démarche relevant d'une forme de recherche action expérimentale, visant à allier théorie et mise en pratique du développement d'une application de production automatisée d'informations. Cette démarche s'explique par la nécessité de comprendre avant d'expliquer les situations sociales (Barbier 1996 :14), et elle implique notre engagement dans un processus de médiation entre l'artefact et le groupe social étudié. En réalisant l'expérience technologique, il s'agissait également de mieux se rendre compte des challenges techniques et journalistiques abordés dans le deuxième chapitre de cette thèse. Cependant, il ne s'agit pas d'envisager cette participation active au processus de développement de l'outil comme une fin en soi, mais bien comme un moyen permettant d'étudier les comportements et usages dans une rédaction donnée.

Le quatrième chapitre traite des usages de l'artefact "QuoteBot" par les journalistes du service "Investir" du quotidien *L'Echo*. L'artefact a été développé par la start-up française Syllabs, spécialisée dans le développement de technologies sémantiques. Dans ce projet, nous avons été sollicitée en tant que consultante pour accompagner la rédaction dans le processus de conception de l'artefact technologique, ainsi que pour en évaluer les usages journalistiques. Les journalistes, en tant qu'utilisateurs-finaux, ont été activement impliqués dans la conception d'un artefact destiné à soutenir des pratiques quotidiennes s'organisant en flux tendu. La complexité de cette expérience tient tant dans ses aspects journalistiques et techniques – l'artefact devant répondre aux exigences de sa conception qui furent posées par les journalistes eux-mêmes –, que dans ses aspects organisationnels, en raison de la multiplicité d'acteurs impliqués dans ce projet.

L'observation participante est la principale méthode de collecte de données empiriques mise en œuvre dans ces deux études de cas. Elle a pour caractéristique de permettre "*de vivre la réalité des sujets observés et de pouvoir comprendre certains mécanismes difficilement décriptables pour quiconque demeure en situation d'extériorité. En participant au même titre que les acteurs, le chercheur a un accès privilégié à des informations inaccessibles au moyen d'autres méthodes empiriques*" (Soulé 2007). Cette méthode relève d'une approche ethnographique embarquée, dont l'objectif est de comprendre le processus social à l'œuvre tout au long de l'expérience (Boyer & Hannerz 2006). Dans une observation participante, les changements des comporte-

ments peuvent être observés sur une période donnée : si les usages se construisent dans le temps, il en va de même pour les discours y afférent. L'observation participante va donc fournir un contexte à l'expérience, pouvant amener "*à prendre conscience d'éléments incongrus ou inexpliqués (... et de les rendre ...) sensibles à leurs éventuelles implications ou connexions avec d'autres faits observer*", ce qui ne peut être reproduit dans des entretiens (Becker & Geer 1958). Cette méthode est enrichissante mais inconfortable, car elle interroge le positionnement du chercheur. Cependant, elle apporte une assise à la recherche et donne un sens à la réalité sociale étudiée, tout en impliquant un processus de subjectivation et de résistance. La nécessité critique y apparaît donc comme un garde-fou (Horvat 2013). L'observation participante permet également de suivre les évolutions dans le temps des représentations ou des pratiques au sein d'un contexte défini. Ses bénéfices sont ceux de montrer les articulations entre théories, concepts et enquêtes empirique (Boumaza & Campala 2007); mais aussi de garantir un accès privilégié à des informations, en vue de suivre les évolutions des représentations ou des pratiques au sein d'un contexte défini (Park, cité par Boumaza & Campana 2007).

Cette méthode a été complétée par des entretiens semi-structurés, dont l'objet est de se concentrer sur les agents humains et la manière dont ils perçoivent ou ressentent l'expérience vécue. Les entretiens permettent une analyse du "*sens que les acteurs donnent à leurs pratiques, aux événements dont ils ont pu être les témoins actifs; lorsque l'on veut mettre en évidence les systèmes de valeurs et les repères normatifs à partir desquels ils s'orientent et se déterminent*" (Blanchet & Gotman 1992, cités par Pape & Martin 2010). Ce faisant, ils participent à une démarche qualitative où l'objectif est de mettre en évidence des processus et des logiques d'action au sein d'un groupe social précisément identifié (Demazière & Dubar 1997, cités par Pierret 2004).

Compte tenu des caractéristiques sociales et organisationnelles propres à chaque étude de cas, la méthode de collecte de données empiriques a inclus des méthodes de recherche particulières, précisées au début de chacun de ces deux chapitres. La mise en œuvre d'une variété de méthodes de collecte des données empiriques a été envisagée dans un esprit de complémentarité, en raison du caractère récent d'un phénomène encore peu étudié. Ceci fait également écho à une proposition de Denis (2009), qui consiste à changer d'angle d'observation lors d'une même enquête dans le but de "*faire varier l'unité de l'action observée et la position de l'observateur*". Ce faisant, il s'agit d'offrir d'élargir les perspectives des observations empiriques, de manière à proposer une analyse enrichie et renforcée.

0.4 Implication personnelle

Malgré le mythe qu'il entretient, le principe de neutralité de la recherche a été privilégié tout au long du processus de recherche. Pour cette raison, celui-ci fut aussi placé sous tension, tant il est impossible à garantir en raison de la proximité qu'induit une étude de type ethnographique entre le chercheur et son objet (Boumaza & Campana 2007). Notre situation embarquée a été conditionnée, dans la première étude de cas, par la nécessité de l'expérience pour étudier un phénomène qui n'existait pas encore en Belgique francophone au moment où ont été lancés les travaux de recherche. En ce qui concerne la seconde étude de cas, nous avons été sollicitée

par le responsable du développement et de la transformation numérique de *L'Echo*, dans le cadre d'une mission de consultance. Bien qu'enrichissante, car elle induit un accès privilégié aux informations de terrain, cette situation n'en est pas moins demeurée inconfortable (Horvat 2013). Cela étant, toute étude de type ethnographique induit une tension entre le chercheur et son objet, dans le cadre d'une démarche visant à articuler concepts et enquête empirique (Park, cité par Boumaza & Campana 2007). Cette tension – entre engagement et distanciation, dès lors que le chercheur est en partie lié à son objet d'étude (Elias 1993) – s'inscrit dans le cadre d'une stratégie d'étude du terrain visant à la connaissance, et qui implique une nécessaire relation sociale qui n'est pas exempte d'effets sur les résultats obtenus (Bourdieu 1993 :903, cité par Papirot 2014). Les observations participantes ont donc supposé une implication personnelle au cours d'une "*période d'interactions sociales intenses entre le chercheur et les sujets, dans le milieu de ces derniers*" (Bogdan & Taylor 1975/1985, cités par Lapassade 2017). Elles ont imposé une réinterrogation permanente portant à la fois sur la méthode mise en œuvre et sur l'analyse des résultats. Elles ont impliqué un nécessaire recul en ce qui concerne l'analyse des données collectées. Admettre et baliser ceci a donné lieu à une nécessaire mise à distance dans le temps.

Nous sommes, par ailleurs, impliquée dans une collaboration professionnelle avec l'agence Alter depuis août 2015, dans des activités de développement de ses plateformes en ligne (Alter-medialab, *Alter Échos*, Agence Alter) et de mises en récit (rédaction d'articles journalistiques, conception de formats éditoriaux). Cette proximité peut être considérée comme un inconvénient, car cela suppose une proximité professionnelle pré-existante avec les sujets d'étude, mais aussi comme un avantage, dès lors que cela a contribué à forger une relation de confiance indispensable au bon déroulement de la recherche. Dès lors, cela a impliqué la mise en place d'un dispositif qui permette suffisamment de mise à distance critique pour décrire des pratiques et les normes institutionnelles qui encadrent ces pratiques (Demeulenaere 2012). Les bénéfices de cette implication professionnelle sont de comprendre plus facilement le sens d'une réalité sociale donnée, et d'être capable de s'immerger plus rapidement dans le domaine de la recherche en tant qu'"initié". Toutefois, dans ce type de démarche, un manque de familiarité avec un terrain d'étude peut s'avérer tout aussi précieux (Paterson & Zoellner 2010).

La prise en charge du développement de l'objet-outil "Bxl'air bot" s'inscrit dans le cadre d'une démarche expérimentale. Là aussi, il convient d'admettre qu'aucun choix posé ne peut être considéré comme formellement neutre, qu'il soit technique ou social. Toutefois, chacun de ces choix a été nourri, documenté, argumenté, placé en situation critique et validé par la porte-parole de la rédaction. Cette validation était un préalable à chaque étape du développement de l'artefact. Toute intervention technique a été conditionnée par des demandes formulées de manière explicite ou implicite par les journalistes impliqués dans ce projet. De manière à ne pas mélanger les contextes, toutes les formes de collaborations professionnelles qui ne participaient pas au projet de recherche ont été gérées de la manière la plus distincte possible. Il convient également d'admettre (et d'assumer) que nos expériences antérieures en journalisme et en développement numérique ont pu influencer sur la manière dont ce projet a été envisagé, en particulier lors de la phase de prototypage du dispositif.

À l'instar de la première étude de cas, cette seconde situation d'immersion s'est inscrite dans un processus dual de subjectivation et de résistance (Horvat 2013). Les relations avec la rédaction de *LEcho* et avec la start-up en charge du développement de l'artefact "QuoteBot", ont toujours été circonscrites au cadre de la recherche. Notre rôle fut, ici, d'accompagner les journalistes tout au long du processus de conception. Dans un second temps, il s'agissait d'assurer une mission d'évaluation de l'adéquation de l'artefact aux usages journalistiques, une démarche s'inscrivant pleinement dans la perspective de cette thèse. Sur le plan relationnel, le terrain de recherche a pu être considéré, à certains moments, comme difficile sur le plan des relations sociales (Boumaza & Campana 2007), dans la mesure où il impliquait un acteur extérieur à la rédaction de *LEcho* dont l'attitude a pu être perçue de manière mitigée, entre adhésion et méfiance, face à une démarche de recherche mobilisant une approche critique.

0.5 Limites de la recherche

Si une approche microsociologique centrée sur une observation participante permet de mettre en relief les situations, comportements et usages, elle n'en est pas nécessairement génératrice d'enseignements généraux sur le plan des concepts (Soulé 2007). Ce travers a pu être levé via l'approche macrosociologique présentée dans la première partie de cette thèse. Cela étant, les deux études de cas développées dans le cadre de cette recherche consistent en des expériences uniques, qui ne pourraient être reproduites à l'identique : cela supposerait de bénéficier de conditions de recherche qui soient semblables, en ce compris à propos des caractéristiques propres aux groupes sociaux étudiés.

Le concept d'usage est envisagé dans une acception où représentations et contexte jouent un rôle structurant. Bien qu'appartenant à un groupe social pouvant partager des représentations professionnelles communes, il s'agissait de contextes particuliers et distincts qui s'inscrivent tous deux dans le cadre particulier du marché de la presse belge francophone. Celui-ci est caractérisé tant par son étroitesse, que par des conditions d'exercice de la profession que l'on peut considérer comme difficiles ou fragiles.

Le caractère récent du phénomène de production automatisées d'informations constitue une dernière limite à cette recherche. En Belgique, aucun projet d'automatisation de la production d'informations n'était en cours au moment de débiter cette recherche. Pour toutes ces raisons, il est possible que les résultats des observations empiriques puissent être différents dans d'autres contextes. Si, pour cette raison, le caractère général des enseignements peut s'en trouver limité, le fait de disposer de deux terrains de recherches aux contextes organisationnels différents permet d'en relever les caractéristiques communes, de mettre ces résultats en perspective avec l'approche macrosociologique, et de fournir des pistes de travail pertinentes pour de futures recherches sur le phénomène.

Première partie

**Représentations sociales et pratiques
culturelles**

1 | Imaginaires technologiques

Les imaginaires d'une innovation technique ne sont pas stables. Ils évoluent en passant par des phases de consensus euphorique ou de grand scepticisme, d'affrontement et de débats, même passionnels, puis ils se stabilisent avec la diffusion de la technique et le développement de ses appropriations.

Pierre Musso (2009)

Le mot "technique" apparaît en français au XVIII^e siècle. Il est issu du latin "technicus" et du grec "technè", désignant l'art et l'habileté à faire quelque chose. Au siècle suivant, il prendra le sens que nous lui connaissons aujourd'hui, "*identifiant la technique à une activité productive, à la fabrication d'objets et à la manipulation de l'environnement*" (Jarrige 2014 :28). Le terme "technologie" désigne les outils, savoirs et pratiques relatifs à un domaine technique particulier. Technique et technologie ont ceci de commun qu'elles nourrissent des imaginaires rationnels ou irrationnels, perçues tantôt comme une menace pour l'homme, son emploi ou son environnement, tantôt comme symbolisant le progrès, en renfermant les promesses d'un avenir radieux pour l'homme. En ce sens, la technologie est duale : à la fois fonctionnelle et fictionnelle (Musso 2008). "*Une image n'est rien d'autre qu'un rapport (...) Le mot d'images ne saurait donc désigner que le rapport de la conscience à l'objet; autrement dit, c'est une certaine façon qu'a l'objet de paraître à la conscience, ou une certaine façon qu'a la conscience se donner un objet*", écrivait Sartre dans "L'imaginaire", en 1940.

Souvent associées au concept de progrès, les innovations technologiques sont caractérisées par des discours utopiques et dystopiques (Domingo 2008), constitutives de mythes et d'une idéologie technolâtre (Musso 2008). Aussi, ont-elles contribué tantôt à façonner des discours relevant de la science-fiction, tantôt à forger un imaginaire collectif où le progrès technologique est assimilé au progrès social (Jarrige 2014). Cette conception est tellement ancrée dans nos sociétés que l'idée même de critiquer la technologie peut être considérée comme une critique du progrès. Dans cette perspective, l'imaginaire constitue un "*ensemble de représentations sociales articulées : à la fois du réel transformé en représentation et la réalisation de représentations sociales ou individuelles*" (Musso 2009). Combiné aux pratiques sociales, il va participer à la for-

mation des usages de la technique (Musso 2009). Dans le contexte de la production automatisée d'informations, cet imaginaire est notamment nourri par les représentations véhiculées par la métaphore du "robot journaliste", source d'anxiétés professionnelles, identitaires, voire idéologiques car elle remet en question l'idée même de ce qu'est le journalisme (Linden & Dieckx 2019).

Face au phénomène technique, l'imaginaire des journalistes a été façonné de manière individuelle et collective : individuelle, car les représentations mentales d'un individu sont conditionnées par ses propres référents, qu'ils soient culturels ou personnels ; collective, car les professionnels de l'information composent avec les évolutions technologiques depuis de nombreuses années. Aussi, ont-ils pu adopter des attitudes révélant trois ressorts récurrents : (1) une mobilisation des compétences qui suppose un temps d'apprentissage alors que le temps est souvent perçu comme un facteur de pression, (2) une résistance au changement, s'expliquant par la remise en questions professionnelle et individuelle qu'induisent les développements technologiques, (3) une reconfiguration des flux de travail et des pratiques, qui seraient moins liée à des facteurs technologiques qu'organisationnels. Les imaginaires des journalistes se tisseront donc autour des tensions associées au développement des TIC au sein des rédactions. Dans le contexte spécifique d'une approche par données dans le journalisme, ils mettront en relief la contradiction entre une tradition latine du journalisme, issue du domaine de la littérature et plus engagée, et une tradition anglo-saxonne, considérée comme plus objective en opérant la distinction entre les faits et les commentaires (Serrano 2007).

Ce premier chapitre dresse un état de l'art de la recherche traitant des pratiques et usages d'artefacts technologiques, en les abordant sous l'angle des imaginaires auxquels ils sont associés. L'étude de la manière dont l'outil technologique s'est historiquement déployé au sein des rédactions permet de comprendre un contexte socioprofessionnel général. Elle met également en relief des imaginaires qui sont restés stables, tandis que d'autres ont pu évoluer dans le temps. En particulier, ce chapitre s'intéresse à la spécificité de la situation belge francophone, terrain sur lequel se déploient les études de cas détaillées dans la deuxième partie de cette recherche¹. L'examen des enjeux professionnels liés à l'automatisation de la production d'informations soulignera une relation homme-machine où la technologie est interrogée en tant qu'instrument de déstabilisation de ce qu'est le journalisme et de remise en question des compétences mobilisées par les professionnels.

Enfin, il questionne les imaginaires collectifs des journalistes, dès lors qu'ils sont amenés à traiter de l'automatisation de la production d'informations dans le cadre de leurs activités professionnelles. L'imaginaire est ici compris non seulement en tant que ce à quoi les individus pensent, mais aussi à ce que les individus font (Gray 2018). Interroger les usages professionnels des artefacts d'automatisation de la production d'information implique d'interroger ces imaginaires en tant qu'éléments constitutifs de la structuration des usages (Musso 2009), étant entendu que l'on ne peut faire usage d'une technologie sans se la représenter (Flichy 2001).

¹ Voir "Structuration des usages professionnels", p.155

1.1 Etat de l'art sociohistorique

En tant que travailleurs, les journalistes ont toujours composé avec les technologies (Hardt 1990), avec lesquelles ils entretiennent des relations ambiguës depuis les premières heures de l'informatisation dans les rédactions (Powers 2012). Les technologies informatiques ont ainsi pu être considérées tantôt comme des adversaires plaçant les professionnels sous pression, tantôt comme des alliées au service du journalisme. Les premières traces de la rencontre entre informatique, données et journalisme remontent à 1952, époque à laquelle un programme informatique est développé pour le télédiffuseur américain CBS. Il a pour objectif de prédire le résultat des élections présidentielles : qui du républicain Eisenhower ou du démocrate Stevenson remportera le scrutin ? En choisissant Eisenhower, la machine ne se trompe pas (Cox 2000, Howard 2014). Nous sommes alors au début du développement de l'informatique. Le Remington Rand UNIVAC², utilisé pour cette expérience, a été commercialisé l'année précédente. Il s'agissait du premier modèle d'ordinateur mis sur le marché (Chavand 2017). La machine, qui mesurait plus quatre mètres de long pour un peu plus de deux mètres de haut et un poids de huit tonnes, coûtait environ un million de dollars à l'époque³. Elle était capable de traiter 10.000 opérations par seconde⁴.

Les premières expériences d'une collaboration homme-machine fructueuse ne s'inscrivent pas dans le cadre d'un journalisme de prédiction, mais bien dans le cadre d'une approche par données visant à nourrir des récits journalistiques : d'abord avec le développement, aux États-Unis, du journalisme de précision (années 1970), puis avec celui du journalisme assisté par ordinateur (années 1980 et 1990). L'idée de se servir de données comme support à une information journalistique était loin d'être neuve, mais l'informatique a permis d'élargir le spectre des possibilités dans le traitement et l'analyse de données. En Europe, on retrouve les premières traces d'une approche par données dans le journalisme dès le 19^e siècle, dans le magazine britannique *The Economist* qui publie à la une de sa première édition, le 2 septembre 1843, un tableau de données relatives aux exportations de plusieurs entreprises. Il proposera une première visualisation de données quatre ans plus tard⁵. À la même époque, le quotidien britannique *The Guardian*, qui adoptera pour devise "*Les commentaires sont libres mais les faits sont sacrés*" (Rogers 2013), faisait lui aussi ses premiers pas dans le journalisme de données. Cette fois, il s'agissait de la publication d'un listing dévoilant le nombre d'élèves et le coût de la scolarité dans les écoles de la ville de Manchester (Rogers, cité par Bounegru 2013).

Le journalisme de précision⁶, initié par Meyer dès 1967, n'est pas tant lié à l'usage de l'informatique qu'à celui de l'application des méthodes de recherche en sciences sociales aux pratiques

² *Universal Automatic Computer*

³ L'équivalent de plus de huit millions d'euros actuels.

⁴ Selon la loi de Moore, théorisée en 1965, la puissance de calcul d'un ordinateur double chaque année : aujourd'hui, cette loi a atteint ses limites (Waldrop 2016).

⁵ Source : "Data journalism at The Economist gets a home of its own in print", Alex Selby-Boothroyd, Medium, 18/10/2018, consulté le 20/20/2018, URL : <https://medium.economist.com/data-journalism-at-the-economist-gets-a-home-of-its-own-in-print-92e194c7f67e>

⁶ La locution "precision journalism" sera également traduite, en français par "journalisme scientifique".

journalistiques. Toutefois, l'informatique permet un traitement rapide d'importants volumes de données qu'il eût été fastidieux de traiter autrement (De Fleur 2013, Cox 2000). L'approche défendue par Meyer tranche avec le nouveau journalisme, popularisé par Breslin et Wolfe dans les années 1960, caractérisé par une forme narrative empruntée à la fiction. Plusieurs expériences croisant journalisme, informatique et données ont lieu tout au long des années 1970 aux États-Unis, où le phénomène reste circonscrit et le fait de quelques initiés, à commencer par Meyer. La raison principale est à trouver dans une résistance des journalistes à l'utilisation des nouvelles technologies. Mais cela s'explique aussi dans le temps nécessaire pour conduire une enquête : seuls les journalistes obstinés ont supporté ce processus (Maier 2000). Ces premières expériences témoignent également des difficultés d'accès aux technologies informatiques, supposant qu'une rédaction ait les moyens de s'offrir une technologie coûteuse à l'époque, et que les journalistes disposent d'un niveau de compétence suffisant pour maîtriser les outils ou, à défaut, bénéficient de l'appui d'experts (Bounegru 2013).

L'arrivée sur le marché des premiers ordinateurs personnels, moins encombrants que leurs ancêtres et économiquement plus accessibles, couplé au développement de logiciels de tableur, permettront de changer progressivement la donne (Parasie 2013). Aux États-Unis, on parle alors de journalisme assisté par ordinateur (*computer assisted reporting*, CAR) pour désigner ces pratiques qui concernent la collecte des données et leur analyse. La question des compétences reste prégnante, mais il s'agit là de compétences générales plutôt axées sur la collecte d'informations (Coddington 2015). Dans ce modèle, les données n'ont aucune valeur tant qu'elles ne sont pas intégrées dans un récit journalistique. Les journalistes y sont moins des technologues, même si certains d'entre eux ont développé des compétences plus poussées en informatique : "*l'accent mis (...) sur l'analyse des données recueillies selon les principes des sciences sociales est un véritable défi pour le journalisme traditionnel (...) (Le CAR est) subordonné presque entièrement aux principes du journalisme professionnel*" (Coddington 2015).

Au cours de cette décennie, les technologies informatiques sont à la source de collaborations au service d'un journalisme caractérisé par les hauts standards qu'il défend (de Burgh *et al.* 2008 :6). Plusieurs enquêtes s'inscrivant dans ce sillage seront d'ailleurs récompensées par de prestigieux prix Pulitzer. Le journalisme assisté par ordinateur se déploie dans le contexte de l'investigation journalistique (Gynnild 2013) et dans une diversité de formes (De Fleur 2013), mais il ne concerne toujours qu'une minorité d'acteurs passionnés (Parasie 2013). Ceux-ci, indique Meyer (cité par Lewis & Usher 2013), ont souvent été contraints d'acheter eux-mêmes leur propre ordinateur, face aux résistances de leur hiérarchie : pourquoi un journaliste aurait-il besoin de connaître quoi que ce soit à propos d'un ordinateur, lorsque le journal dispose d'un service IT (Meyer 1991) ?

Pour Maier (2000), les pratiques relevant du journalisme assisté par ordinateur sont moins une nouvelle forme de journalisme qu'une extension de la tradition du journalisme d'investigation. Elles permettent de traiter des sources d'information disponibles dans un format numérique, de créer ses propres bases de données à partir d'un travail de récolte sur le terrain, et

d'analyser des jeux de données grâce à la puissance de calcul d'un ordinateur. Il s'agit donc de moyens pour les "*journalistes compétents de suivre les données, de creuser plus profondément, de déterrer les modèles sous-jacents et les faiblesses de la société*". Cette forme de journalisme ne concerne qu'une minorité de journalistes convaincus de ses bénéfices. Malgré l'incroyable travail réalisé par quelques pionniers, l'industrie de la presse s'est toutefois montrée hésitante, voire résistante, quant à l'adoption de nouvelles technologies (Maier 2000b). À l'époque, les bases de données n'ont pas besoin d'être énormes pour être utiles : un logiciel de gestion de bases de données offrait déjà plusieurs fonctionnalités en termes de tri, de calcul et d'analyse (Leonard 1992).

Au milieu des années 1990, le nombre de projets journalistiques s'appuyant sur des bases de données informatiques augmente de manière considérable, gagnant également les entreprises de presse de taille moyenne ou de petite taille. L'utilisation de bases de données est subordonnée à une histoire journalistique, aux normes journalistiques et à la pratique essentielle des statistiques car elles permettent de comprendre les données (Dagiral & Parasie 2012). Dans le même temps, la quasi-totalité des rédactions a été informatisée et l'ordinateur fait désormais partie du quotidien de l'activité journalistique, via le traitement de textes, la production assistée par ordinateur (PAO, qui désigne les logiciels de mise en page), la consultation de bases de données et de courrier électronique. Ces technologies font l'objet d'un usage individuel et collectif (Pélissier & Romain 1998).

Les années 1990 sont aussi celles de l'arrivée du World Wide Web dans le domaine public⁷. Au cours de cette décennie, l'intégration de l'ordinateur dans les processus journalistiques varie en fonction de la complexité des tâches à effectuer. Elle concerne surtout des dispositifs de collecte d'informations (Garrison 2001). La diffusion d'innovations technologiques va toutefois dépendre tant de la formation interne aux entreprises de presse, que du soutien au développement de moyens permettant de créer une masse critique d'utilisateurs (Maier 2000). L'adoption des technologies du web va rester lente, malgré le potentiel de ressources pour les journalistes (Bardoel & Deuze 2001), tandis que les techniques d'enquête s'appuyant sur une approche par données vont bénéficier de nouveaux outils performants qui ne nécessitent plus forcément la connaissance de langages informatiques.

Jusqu'au début des années 2000, les approches par données mêlées de technologies se déploient essentiellement sur le terrain du journalisme d'investigation. Elles sont liées à un imaginaire d'excellence dans le journalisme, voire d'élitisme, porté pendant près de trois décennies par quelques journalistes passionnés qui exploitèrent ordinateurs et bases de données informatisées en tant qu'outils au service de leur travail d'enquête. Mais si les années 1980 furent celles des expérimentations et les années 1990, celles d'une intégration progressive d'internet dans les pratiques journalistiques, l'entame des années 2000 marque le tournant d'une troisième époque dans l'histoire du journalisme et des TIC : en à peine une décennie, la plupart

⁷ Il est né, officiellement, le 30 avril 1993. Source : "Naissance du web", CERN, consulté le 01/09/2017, URL : <https://home.cern/fr/topics/birth-web>

des grands médias d'information se sont dotés d'un site web et leurs audiences en ligne se comptent en millions d'internautes.

Au chevet de ces évolutions, les milieux académiques commencent à s'interroger sur l'émergence de nouveaux comportements et compétences au sein de la profession journalistique. Les recherches traitant des médias, du journalisme en ligne et de ses pratiques se multiplient (Deuze 2003, Paterson & Domingo 2008). Il sera notamment observé que les politiques de convergence véhiculent des représentations négatives parmi les journalistes, et que celles-ci constituent un obstacle à l'usage de technologies. Aussi, la logique multisupport d'un journalisme multimédia induit-elle une augmentation de la charge de travail, une pression temporelle accrue ou encore des compétences à mobiliser ou à acquérir pour rencontrer de nouvelles exigences (Maier 2000).

Les contenus que les journalistes de données conçoivent "*sont désignés sous le terme de 'news applications' qui désignent aussi bien des cartes interactives, des infographies, des bases de données interrogeables et d'autres formes de présentation en ligne*" (Dagiral & Parasie 2013). Anderson *et al.* (2016) définissent le datajournalisme comme l'application de techniques statistiques à l'analyse de sources pouvant prendre la forme de bases de données informatiques, d'enquêtes d'opinion ou de tout autre type d'enregistrement numérique. Pour Trédan (2014), le journalisme de données s'appuie sur deux ressorts principaux : "*d'un côté, un savoir-faire technologique qui permet le développement des interfaces de visualisation des données; de l'autre, un savoir-faire dans l'identification et l'exploitation de données disponibles*". Qu'il s'agisse de récolter, analyser ou traiter des jeux de données, la démarche fondamentale du journalisme de données reste la même que dans n'importe quelle autre forme de journalisme : celle de raconter une "bonne" histoire en donnant du sens au récit (Coddington 2015).

Le journalisme de données s'inscrit dans le prolongement de la démocratisation "*massive des ressources, outils, techniques et méthodologies utilisés précédemment par des spécialistes, qu'ils soient journalistes d'investigation, spécialistes des sciences sociales, statisticiens, analystes ou autres experts*" (Bounegru 2013). Deux raisons permettent d'expliquer la popularisation du genre : un accès accru à des volumes de données de plus en plus importants, et le développement de technologies permettant l'extraction de données ainsi que leur visualisation sous la forme de cartes ou de graphiques interactifs (Henninger 2013). Ces nouvelles pratiques ont donné lieu à de nouvelles formes de coopération, où journalistes et développeurs sont amenés à travailler main dans la main (Antheaume 2013).

La paternité du terme "datajournalisme" est attribuée au journaliste-développeur Adrian Holovaty (Howard 2014). D'abord responsable des innovations éditoriales au *Washington Post*, il est également le fondateur de *EveryBlock*, un projet de datajournalisme hyperlocal reposant sur l'extraction de données provenant de sites web institutionnels et de fichiers transmis par les autorités pour informer les habitants de Chicago à l'échelle de leur quartier (Parasie 2013). En 2005, il crée l'un des premiers *mashup* à l'aide de Google Maps, en proposant une carto-

graphie du crime à Chicago. L'année suivante, il publie un manifeste appelant le journaliste à changer non pas sur le fond mais sur la forme, réclamant un journalisme de faits s'appuyant sur des chiffres et des données statistiques (Dagiral & Parasie 2012). Holovaty – diplômé de l'école de journalisme du Missouri – est impliqué dans la communauté des logiciels libres. Pour lui, faire du journalisme à l'aide de programmes informatiques n'est rien de plus qu'une manière de faire différente pour réaliser des objectifs journalistiques (Powers 2012).

Le journalisme de données fait référence au processus par lequel les journalistes utilisent des données numériques, qui constituent le matériau de base à l'information (Gynnild 2013). S'il n'en réinvente pas les techniques du journalisme, il compose avec les outils actuels (Joannes 2007). Il est aussi perçu comme une manière de retisser un sentiment de confiance entre les journalistes et leurs audiences, en s'appuyant sur le mythe selon lequel les données vont permettre de réaliser des objectifs de vérité, de précision ou d'objectivité (Sandoval-Martín & La-Rosa 2018). Cette approche peut être utilisée pour expliquer des données, servir de base à un projet d'investigation, enrichir une information, vérifier des faits ou appuyer des pratiques quotidiennes. Ces dernières seront toutefois limitées en ressources et en outils, et l'accent sera surtout placé sur la simplicité et sur la mise en forme visuelle (Wright & Doyle 2018).

Pour les journalistes engagés dans une approche par données, naissent alors de nouvelles possibilités. Celles-ci sont désignées par les vocables "journalisme computationnel" ou "journalisme algorithmique". La nouveauté réside à la fois dans le mode de traitement et de diffusion des données, ainsi que dans la manière dont les technologies s'intègrent dans les pratiques professionnelles (Plesner 2009). Cela fait référence à des pratiques hybrides et hétérogènes qui s'appuient sur des outils des sciences de l'informatique, des sciences sociales et des sciences humaines numériques pour collecter, vérifier, représenter, publier, diffuser des informations (Cohen *et al.* 2011). Le vocable "journalisme computationnel" fut utilisé pour la première fois en 2006, au Georgia Institute of Technology, à l'occasion d'un cours dispensé par le professeur Irfan Essa (Gynnild 2013). Il est défini comme une combinaison d'algorithmes, de données et de connaissance des sciences sociales pour compléter la fonction de responsabilité du journalisme (Hamilton & Turner 2010). Il recouvre des pratiques aussi variées que sont l'exploration de données (*data mining*), l'extraction de données (*data scraping*), la manipulation et la visualisation de données, l'analyse statistique et la génération automatique de textes en langage naturel. Lorsque ses processus se concentrent sur la seule génération automatique de textes, qui consiste en l'utilisation d'algorithmes codés pour automatiser des processus précédemment utilisés par les éditeurs et les journalistes, il est aussi appelé "journalisme algorithmique" (Anderson 2013, Dörr 2015, Graefe 2016, Weber & Kosterich 2018). Carlson (2014) et Latar (2015) parleront, quant à eux, d'un "journalisme robotique" reposant sur deux piliers : des logiciels extrayant automatiquement de nouvelles connaissances à partir de silos de données, et des algorithmes qui transforment automatiquement ces nouvelles connaissances en histoires lisibles pour des êtres humains. Ces processus ne se produisent pas en dehors de toute intervention humaine, mais le rôle de l'humain dans la création de contenus est devenu "indirect" (Napoli 2014).

Le terme "algorithme" désigne une procédure codée, destinée à transformer des données en fonction de calculs spécifiés, en vue de résoudre un problème. Cela suppose qu'il y ait un état initial et un état fini (Gillespie 2014). Une définition plus courte de l'algorithme, et plus largement répandue, est celle d'une suite finie d'instructions destinée à résoudre un problème (Steiner 2012). Les origines de l'algorithme remontent à l'Antiquité. L'algorithme d'Euclide (-300 AJC), l'un des premiers algorithmes connus (et encore utilisé de nos jours), permet de déterminer le plus grand commun diviseur de deux entiers naturels. Ce terme provient du nom latinisé d'un mathématicien perse (-9 AJC), Ali-Khwârizmi (Fuller *et al.* 2008). Dans le journalisme algorithmique (ou journalisme computationnel), l'automatisation est l'une des principales composantes. De la détection de *breaking news* à la diffusion d'informations, en passant par la sélection de l'information, le *fact-checking*, la hiérarchisation des contenus et la production automatisée de textes, l'ensemble de la chaîne de production journalistique est concerné (Steiner *et al.* 2013, Leppänen *et al.* 2017). Sur un plan opérationnel, les procédures algorithmiques sont, par exemple, utilisées pour prioriser, classer ou filtrer l'information (Diakopoulos 2015), pour prédire des demandes dans le contexte du "big data" ou pour créer des contenus originaux (Napoli 2014).

Il est généralement reconnu au journalisme computationnel le respect de valeurs fondamentales du journalisme, telles que la précision, l'immédiateté et la vérifiabilité (Graefe 2016). Elles seraient, toutefois, moins liées aux normes et pratiques professionnelles mais, à l'instar du journalisme de données, elles constituent une réponse face à une abondance d'informations (Coddington 2015, Flew *et al.* 2012). Leurs limites sont d'être très dépendantes d'une expertise du domaine d'application ainsi que d'une expertise technique. Le développement de ces pratiques s'explique par des facteurs organisationnels, des routines de travail et des représentations des utilisateurs qui façonnent des processus d'adoption des technologies résultant des évolutions technologiques. Ce façonnage mutuel donnerait ainsi lieu à des produits éditoriaux distincts (Anderson 2013).

La génération automatique de textes en langue naturelle (GAT) participe à cette dynamique. Appliquée à une variété de domaines dès le début des années 1990, tels que les modes d'emploi de produits commerciaux et divers textes associés au monde de l'entreprise pouvant être standardisés (Dale & Reiter 1995), elle s'est développée dans le giron journalistique à partir de 2007 aux États-Unis, pour s'étendre à d'autres continents dès le début des années 2010. Parmi les principaux fournisseurs de contenus automatisés à destination des médias d'information, figurent, en bonne place, les sociétés spécialisées dans le traitement automatique de la langue. À l'échelle mondiale, on en dénombre moins d'une quinzaine (Dörr 2015). Elles proposent leurs services à une variété de secteurs, tels que la communication, le business et le marketing. Les médias d'information ne constituent pas leur unique clientèle, mais elle est la plus visible. Aucune d'entre elles ne se revendique comme un média ou n'estime faire acte de journalisme. Ces sociétés se distinguent dans la manière dont elles assurent la fourniture de leurs services, suivant deux logiques : celle d'un logiciel que le client paramètre en fonction des données qu'il

souhaite traiter, comme c'est le cas au sein du groupe de presse suisse Tamedia pour l'automatisation des résultats de votations, où cette activité a notamment été confiée à un journaliste de données (Plattner & Orel 2019); et celle d'un flux d'informations auquel le client souscrit sous la forme d'un abonnement, dont le développement et le paramétrage seront assurés par ses soins, comme c'est le cas de Syllabs en France. Si les expériences d'automatisation de la production d'informations se sont multipliées au cours de ces dernières années, elles peuvent encore être considérées comme marginales ou expérimentales, et cela pour deux raisons : (1) faire appel aux services d'une société spécialisée dans la génération automatique de textes ou mobiliser une équipe en interne pour développer une solution sur mesure présente un coût qu'il s'agit de mettre en balance avec les bénéfices attendus de l'opération; (2) les rédactions ne disposent pas toujours des ressources et compétences nécessaires pour développer des solutions en interne (Linden 2017).

Nous l'avons vu, le développement d'une approche par données dans le journalisme se décline en une multiplicité d'avatars. Si la frontière entre les genres peut s'avérer ténue, que ce soit dans leur définition ou dans le cadre des pratiques professionnelles qu'ils désignent, une typologie permet d'affiner ce qui les distingue. Coddington (2014) souligne que si les journalistes engagés dans les pratiques d'une approche par données semblent indifférents au fait de "classer" leur travail, il est utile pour le chercheur d'en dresser la typologie de manière à comprendre leurs formes et implications.

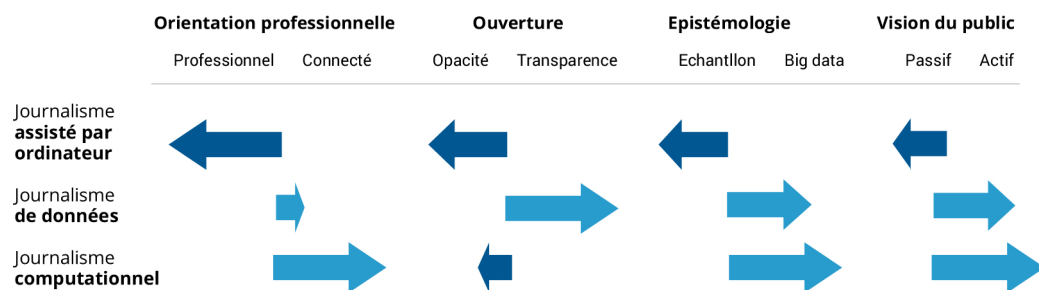


FIGURE 1.1 – Typologie d'une approche par données dans le journalisme (Coddington 2014)

Dans la typologie qu'il propose (Figure 1.1), Coddington différencie le journalisme assisté par ordinateur du journalisme de données en termes de transparence, d'ouverture, et de relation active avec les audiences. Le journalisme computationnel serait plus opaque dans ses processus, tout en étant à même de gérer des volumes de données plus importants. Les principales différences consistent (1) dans l'enracinement du journalisme assisté par ordinateur aux méthodes issues des sciences sociales et son orientation vers le journalisme d'investigation; (2) dans la démarche de transparence du journalisme de données; (3) dans les processus d'abstraction automatisés qui caractérisent le journalisme computationnel. Ce qui distingue principalement les pratiques relevant du journalisme computationnel de celles du journalisme de données consiste dans le fait que ce premier s'appuie sur l'extraction d'informations pour produire des modèles informatisés. À l'inverse, le journalisme de données repose sur une analyse

de données pour nourrir une information journalistique. Par ailleurs, les particularités du journalisme computationnel seraient moins celles de libérer les journalistes d'un travail ingrat que de développer "*de nouvelles formes de journalisme axées sur les données et sur l'utilisateur, qui ont le potentiel de réaliser les espoirs et promesses du journalisme traditionnel*" (Karlsen & Stavelin 2014).

Cette segmentation n'est pas toujours aussi claire dans l'esprit des journalistes, en ce compris en termes de définition des pratiques. Par exemple, le journalisme de données peut aussi bien être assimilé à une capacité à informer par l'articulation de preuves quantifiables et leur mise en contexte, ou à une possibilité de voir au-delà de l'information pour trouver de nouvelles idées qui seront transformées en une expérience utilisateur pilotée par une base de donnée (Borges-Rey 2016). Il peut être appréhendé comme s'inscrivant dans le prolongement de pratiques existantes, comme un outil supplémentaire au service du journalisme d'investigation, ou encore comme un moyen de rendre l'information plus "sexy". Certains journalistes peuvent également considérer que produire un graphique, c'est déjà du datajournalisme (De Maeyer *et al.* 2015). Bien que les technologies d'automatisation de la production d'informations participent à une approche par données dans le journalisme, leur caractère autonome est susceptible d'entraîner des résistances professionnelles. Celles-ci s'inscrivent dans le cadre plus large des phénomènes de résistance observés dans le monde du journalisme, dans la foulée de l'introduction de technologies numériques dans les rédactions.

1.1.1 Identification des résistances

Plusieurs travaux de recherche ont eu pour objet de comprendre la manière dont les innovations technologiques se sont déployées dans les rédactions. Ils peuvent être considérés selon deux premières époques : avant la naissance du *World Wide Web* puis lorsque les communications par internet et par courrier électronique sont devenues accessibles en dehors des communautés militaires et académiques (Plesner 2010). Cette dernière vague de changement dans les pratiques professionnelles fut observée à partir du milieu des années 1990. La recherche s'est ensuite attachée à comprendre la diffusion des innovations technologiques, de même que leurs implications sur les pratiques professionnelles. Cela étant, l'adoption de nouvelles idées, technologies et pratiques participe à un processus social qui nécessite du temps (Garrison 2001) ; tandis que la relation entretenue entre les journalistes et les objets informatiques sera souvent caractérisée par une résistance au changement en général (Deuze 2005), et à la technologie en particulier.

Le phénomène de résistance caractérise une culture professionnelle refusant de bousculer les valeurs véhiculées par l'idéologie du journalisme, que ce soit dans le cadre de politiques de convergences ou dans celui d'expériences de journalisme participatif (Deuze 2005, Nygren 2014). Les premières études de cas empiriques menées à cette deuxième époque de la recherche se concentrèrent essentiellement sur la manière dont le courrier électronique et internet avaient modifié les pratiques journalistiques (Cochran 1997, Dumlao & Duke 2003, cités par Plesner 2009). Elles spéculaient également sur les avantages, dans le cadre de la collecte

d'informations en ligne, à créer un accès à internet pour chaque poste de travail dans une rédaction, et sur ceux d'une redistribution des rôles induite par l'adoption de technologies en ligne. Quelques recherches se sont alors intéressées aux interactions entre nouvelles technologies et pratiques professionnelles.

Au Royaume-Uni, Cottle et Ashton (1999) ont observé que la politique de convergence, menée dans le cadre du développement d'une rédaction en ligne à la BBC, a eu pour effet une transformation des pratiques professionnelles devant désormais tenir compte d'une production multimédia qui place les journalistes sous pression. Ils ont également fait état des premières préoccupations quant à la possibilité d'un impact négatif des technologies sur l'information. Saltz & Dickinson (2008) dresseront le même constat, dans le cadre de l'étude des processus de convergence, menée en 2002 et 2003 dans quatre médias nationaux britanniques⁸. Ils démontreront que la pratique du journalisme multimédia est plus lente à s'installer parce qu'elle se heurte à la résistance des journalistes, devant négocier avec un changement s'opérant tant sur le plan organisationnel que sur celui du processus de production de l'information, lequel implique une redéfinition de la responsabilité des journalistes. Cette résistance s'explique par un affaiblissement des lignes de démarcation non seulement entre les formes médiatiques, mais aussi entre les journalistes et les équipes techniques.

Aux États-Unis, Maier (2000) a souligné que l'adoption de nouvelles technologies n'était pas systématique, que les journalistes témoignaient d'une forte résistance à la technologie et que si la question des compétences, couplée à celle du savoir-faire technologique, pouvait constituer un obstacle à l'usage, le manque de temps était la difficulté la plus importante à négocier. Toutefois, cette variable n'est pas la seule à prendre en considération pour expliquer les résistances : dans une recherche portant sur trois radiodiffuseurs britanniques⁹, Ursell (2001) a mis en avant que la responsabilité de la détérioration de l'information et des normes journalistiques n'était pas tant imputable aux innovations technologiques qu'à la direction et au management des entreprises de presse.

Dès le début des années 2000, les domaines de recherche se sont élargis à la compréhension dont les technologies de l'information et de la communication ont été intégrées dans les routines journalistiques, en considérant les apports des modèles théoriques de la sociologie des sciences et des technologies (STS). Cette posture épistémologique, qui marque le passage à une troisième époque dans la recherche, tranche avec les études considérant l'impact déterministe de la technologie sur les pratiques professionnelles, qui plaçaient l'appréhension de la technologie sur un plan utopique (Steensen 2011, Spyridou *et al.* 2013). Boczkowski (2004) propose d'envisager la technique et le social sous l'angle de leurs influences mutuelles (*mutual shaping*). Il considère les processus communs entre la construction technique et l'adoption par la société des nouveaux artefacts médiatiques selon trois facteurs critiques : (1) l'interdépendance entre technique et social, (2) le caractère continu de ce processus, (3) l'influence du

⁸ BBC, Skynews, *The Guardian* et *The Financial Time*

⁹ BBC, ITN et Yorkshire Television

contexte historique dans lequel il se déroule. Il s'agit également de reconnaître le caractère matériel de l'artefact technologique (Boczkowski 2015) qui, dans certains cas, endossera un statut d'acteur ou d'actant. La problématique de la résistance à la technologie sera alors examinée sous l'angle des relations de pouvoir au sein des rédactions.

Cela étant, la résistance au changement ne s'exprime pas toujours dans le contexte du développement d'innovations technologiques. Aux États-Unis, Ryfe (2009) a examiné les tentatives d'un rédacteur en chef pour encourager les journalistes de sa rédaction à ne plus se rendre sur le terrain pour couvrir des meetings publics. Ceux-ci ont estimé que, dès lors que cela les privait d'un contact direct avec leurs sources d'information, cette manière de faire tranchait avec ce qui constitue un "bon" journalisme. Cette étude montre une culture professionnelle à la fois résiliente et résistante au changement. Et elle indique que pour intégrer un changement dans la culture professionnelle journalistique, celui-ci doit s'appuyer sur deux ressorts : ses avantages relatifs et sa compatibilité avec la manière dont les journalistes envisagent leur profession (Ekdale *et al.* 2015).

Les explications du phénomène de résistance sont donc à trouver dans une variété de facteurs, qui peuvent être autant liés tant à l'organisation elle-même et à son management, qu'à un manque de budgets, de temps ou de formation professionnelle (Garrisson 2001, cité par Deuze 2008). Dans le contexte des TIC, ces résistances vont relever d'un imaginaire façonné à la fois par un idéal de la profession, et par les mythes véhiculés par la technologie en elle-même. Depuis la première révolution industrielle, l'idée d'une technique porteuse de progrès fait l'objet de critiques souvent liées à des questions de préservation du travail, de l'environnement et/ou du mode de vie de l'humain (Jarrige 2014). Si l'on prend le cas particulier de l'automatisation de la production d'informations, ces trois variables sont au centre d'enjeux majeurs¹⁰. La critique de la technologie est souvent assimilée à une critique du progrès, et donc à une résistance au changement, et cette résistance va être exacerbée à chaque moment de crise et de transformation sociale (Jarrige 2014 :345).

Dans le débat portant sur la technologie en tant qu'instrument puissant de changement social, Mosco (2005) considère que la technologie y contribue tout autant qu'elle le retarde. S'il admet que les forces sociales, économiques et politiques sont importantes dans ce que l'on comprend de la technologie, les relations sociales s'en trouvent fondamentalement transformées. Ce qu'il appelle le "sublime technologique", c'est prêter à la technologie une toute puissance qui ne serait rien d'autre qu'une forme d'hymne au progrès. Ces discours utopiques sont sous-tendus par une approche déterministe, où la technologie est envisagée comme une force externe de changement ou de contrôle social : en ce sens, elle est considérée une force primaire qui aliènerait les autres phénomènes sociaux (Orlikowski 1992, Scranton 1995, Sterne 2014). La résistance ferait alors partie intégrante d'un "*processus complexe de négociation sociotechnique, à travers lequel toute la société définit son rapport aux artefacts matériels*" (Jarrige 2014 :18).

¹⁰ Voir "Enjeux et défis professionnels", p.57

1.1.2 Reconfiguration des processus sociaux et éditoriaux

L'introduction d'innovations technologiques au sein des rédactions n'a pas été sans incidences sur les flux de travail et les pratiques professionnelles. Les politiques de convergence ont entraîné dans leur sillage des logiques multisupports, tandis que le développement du journalisme de données, puis du journalisme computationnel, ont amené de nouvelles compétences et de nouveaux acteurs sur le terrain de l'information. Dans le champ académique, ces reconfigurations ont été étudiées dans une perspective transdisciplinaire, en se nourrissant des apports théoriques des STS. La théorie de l'acteur-réseau, en particulier, a permis une mise en relief des relations de pouvoir au sein des organisations médiatiques, en vue d'expliquer les transformations du travail dans le monde du journalisme. Sur le plan de l'imaginaire, celles-ci vont évoquer des sentiments partagés, pilotés par les tensions qu'induisent les nouveaux modes de productions de l'information, mais aussi par les opportunités en termes de polyvalence et de nouvelles formes d'expériences au service du récit médiatique.

C'est dans cet esprit que Boczkowski réalise, en 2005, une étude ethnographique au sein de la rédaction d'un média argentin¹¹, en vue de définir les contraintes de l'information chaude et de l'information froide. Son analyse met en avant le rôle central des schémas temporels dans la structuration du travail éditorial, et elle montre une variabilité des activités journalistiques en fonction de facteurs organisationnels et matériels. La théorie de l'acteur-réseau est mobilisée dans une recherche de Domingo (2008), portant sur l'interactivité du journalisme en ligne envisagée en tant que mythe. Les résultats suggèrent que la culture professionnelle du journalisme traditionnel prévaut dans les rédactions en ligne, et que cette conception empêche le développement des réseaux d'interactivité, le transformant "*en un problème à résoudre au lieu d'une opportunité de changement*". Cette théorie est également utilisée par Plesner (2009), dans une recherche traitant de la création des récits médiatiques à travers un jeu d'associations entre actants hétérogènes. Dans ses principaux enseignements, elle souligne que la nouveauté ne réside pas dans la manière dont les technologies ont modifié les pratiques professionnelles, mais bien dans la manière dont elles s'y sont intégrées.

La méta-théorie des communautés de pratique appuiera une étude ethnographique menée dans quatre rédactions de journaux en ligne¹² par Schmitz-Weiss et Domingo (2010). Bien que cette recherche soit davantage connectée à la théorie de l'acteur-réseau pour expliquer les relations entre les journalistes et la technologie lorsqu'un processus d'innovation est implémenté dans une rédaction, elle met en avant le fait que les journalistes disposant de compétences liées aux technologies jouent un rôle tampon au sein des rédactions. D'autre part, elle souligne des choix professionnels ayant pour conséquences de placer les routines sous contrainte. Cet aspect est également mis en exergue par Micó *et al.* (2013), dès lors que la convergence est comprise comme un facteur constitutif de la dégradation des conditions de travail car elle soutient la production de contenus pour plusieurs plateformes. Caractérisée par la polyvalence des tâches, la convergence n'est pas toujours appréhendée de manière négative. Les jeunes journa-

¹¹ Clarín.com

¹² *The Chicago Tribune*, États-Unis; *El Norte*, Mexique; *El Periódico de Catalunya*, Espagne; et CCMA, Espagne

listes apparaissent comme les plus réceptifs, estimant qu'il s'agit là de nouvelles opportunités en termes de créativité et de liberté de choix dans les sujets et modes de reportages (Ngyren 2014). Si la polyvalence est ce qui se produit lorsque les compétences professionnelles des journalistes vont au-delà d'activités de collecte de traitement d'informations, elle s'inscrit dans une logique à la fois technologique et économique. Et elle est souvent accompagnée par un phénomène de déqualification, dès lors qu'elle laisse moins de temps pour satisfaire des pratiques journalistiques traditionnelles (Kammer 2013). À *contrario*, une logique multisupport peut être considérée comme un moteur positif pour une carrière professionnelle (Singer 2004).

Les évolutions des technologies relatives à la production d'informations journalistiques n'ont pas seulement modifié les pratiques professionnelles (Powers 2012). Elles ont également fait émerger de nouvelles formes de travail et de collaborations. Certaines rédactions ont commencé à faire appel à de nouveaux acteurs aux profils hybrides, disposant de compétences en journalisme et en informatique, de manière à développer, en interne, des projets journalistiques s'appuyant sur des technologies numériques. Aux États-Unis, face à la difficulté d'attirer des candidats issus du monde du journalisme, le chemin a été entrepris en sens inverse. Cette évolution pose la question de la relation entre les journalistes et l'informatique, dans un contexte où les approches par données se multiplient, mobilisant des associations entre agents sociaux provenant de mondes différents. Ces pratiques ne sont pas sans incidence sur le processus de production d'informations. Elles peuvent s'inscrire dans le cadre de projets d'automatisation, tout comme elles peuvent donner lieu à de nouvelles formes de récits médiatiques, souvent désignée par le vocable anglais de *news applications* (Dagiral & Parasie 2013).

La recherche s'est intéressée à ces journalistes d'un nouveau type, à l'instar de Royal (2010), qui a étudié la figure du journaliste-programmeur au sein du département "Interactive News Technology" du *New York Times*. Son objet était de comprendre le rôle joué par la technologie dans une rédaction en évolution constante et en fonction des opportunités ou défis liés à internet. Elle indique que si cette figure est liée à une culture créative et innovante, ainsi qu'à la culture *open source*¹³, ces nouveaux acteurs ne donnent pas l'impression de jouer un rôle différent de celui d'un journaliste traditionnel. Toutefois, dès lors qu'ils sont issus de communautés fondées sur une culture de la transparence qui est controversée dans le mode du journalisme, leur approche peut entrer en contradiction avec la manière d'appréhender un processus journalistique s'appuyant sur un traitement de données. Dans une recherche consacrée à la figure du journaliste-programmeur, Dagiral et Parasie (2012) indiquent que le travail des journalistes-programmeurs se fonde sur l'hypothèse que des données peuvent aider les journalistes à définir l'agenda politique grâce à la divulgation de problèmes publics. Leurs tâches, au sein des rédactions, sont de réaliser des applications qui impliquent "*un large éventail de compétences en informatique pour collecter, traiter, analyser et visualiser des ensembles de données*".

¹³ Ce terme désigne les communautés liées au développement des logiciels libres, par opposition au développement des logiciels propriétaires (Raymond 1999).

Malgré les outils et compétences spécifiques qu'elles requièrent, les pratiques relevant du journalisme computationnel vont être envisagées comme s'inscrivant dans le prolongement du journalisme traditionnel. En Norvège, Karlsen et Stavelin (2014) souligneront une dépendance entre les journalistes et les développeurs qui serait plus forte que la différence culturelle entre ces deux groupes sociaux. L'accent est également placé sur la nécessaire double vision – journalistique et technologique – que suppose la pratique du journalisme computationnel.

Le journalisme computationnel en tant que produit éditorial a été exploré par Young et Hermita (2015), dans une étude portant sur l'impact de ses pratiques dans le contexte du développement d'une carte des homicides dans la région de Los Angeles et publiée sur le site du *L.A. Times*. Ils soutiennent que cette démarche s'inscrit dans une évolution non continue des normes organisationnelles, des pratiques, des contenus, des identités et des technologies; et que la production de la base données sur laquelle s'appuie cette *news application* résulte de nouvelles pratiques combinées aux nouvelles technologies. Ils font ainsi écho aux postures défendues dans les travaux de plusieurs chercheurs, dont Boczkowski, Anderson et Domingo. Les résultats de cette recherche montrent que, dans le contexte du développement de cette carte, l'adoption de la pensée computationnelle et des technologies du journalisme computationnel a été inégale et discontinue au sein de la rédaction. Dans ses différentes formes, relèvent les chercheurs, ce projet de cartographie suggère "*un processus clair d'adaptation mutuelle aux technologies antérieures*" liées au journalisme assisté par ordinateur et au journalisme de données, ainsi que "*la création d'une nouvelle identité professionnelle pour les journalistes*". Objet et outil du journalisme, l'application est ici considérée comme un soutien à un travail journalistique, le journaliste étant, par ailleurs, susceptible d'enrichir les informations de base publiées de manière automatique. L'automatisation n'est donc pas perçue comme une menace mais bien comme une opportunité. Dans le même temps, les chercheurs indiquent que cette application n'a pas modifié la manière de faire des journalistes traitant de la criminalité. Une autre conséquence de l'implémentation de cette application est d'avoir étendu l'expertise professionnelle du journaliste en incluant l'identité du programmeur et de l'acteur non-humain, tous deux étant l'expression de normes et de pratiques distinctes.

La transformation des pratiques professionnelles peut également être comprise dans le cadre de la conception d'interfaces et d'applications. Ananny et Crawford (2015) se sont concentrés sur une dynamique émergente au sein de l'espace techno-journalistique : celle de la conception d'applications d'informations (*news app design*). Ils se sont intéressés à la manière dont les concepteurs comprennent leur travail en relation avec le journalisme, ici envisagé en tant que processus, et en quoi ils marquent leur différence. Les chercheurs perçoivent ce groupe professionnel comme créateur de nouvelles expériences, et ils soutiennent que les activités de conception d'interfaces constituent une presse liminaire émergente. Toutefois, les acteurs de ce domaine ne s'identifient pas comme des journalistes, bien qu'ils définissent les conditions dans lesquelles l'information est produite et circule. Leur activité se déploie dans un espace hybride où cohabitent technologie et entrepreneuriat.

1.2 Particularités du contexte belge

Les technologies d'automatisation de la production d'informations ont commencé à faire leur apparition sur la scène de la presse belge francophone. Le quotidien *L'Echo* s'est lancé, en 2018, dans un projet d'automatisation de données boursières, baptisé "Quotebot", qui fait l'objet d'une étude de cas dans la seconde partie de cette thèse. La même année, le groupe Rossel signait un partenariat avec Wibbitz, une start-up israélienne spécialisée dans la création de vidéos automatisées, pour "*l'aider à diversifier et enrichir la production quotidienne de contenus vidéo à destination de 23 de ses titres médias*"¹⁴. Ces capsules vidéo, qui requièrent toutefois l'intervention d'un journaliste humain, sont diffusées sur les sites du *Soir* et des quotidiens du groupe SudPresse. Questionné en 2016 sur une éventuelle utilisation de générateurs automatiques à l'agence Belga, Philippe De Camps, newsmanager, indiquait que "*actuellement Belga ne travaille pas (encore?) sur ce genre de robot. La seule chose que nous ayons d'automatique est l'entrée de résultats de cyclisme qui génère une dépêche avec un titre, une phrase et des résultats. Mais nous n'avons pas réellement développé le moteur, ce qui fait que cette phrase mal foutue n'est qu'indicative pour le deskeur qui doit toujours la réécrire en français ou en néerlandais*"¹⁵. En décembre 2018, la situation n'avait pas évolué¹⁶.

Étant donnée l'implantation pratiquement inexistante de la production automatisée en Belgique, cette section explore, de manière plus générale, le développement du journalisme de données, qui trouve encore peu d'échos au sein de la presse belge francophone. Les journalistes présentant une filiation avec le journalisme de données sont loin de l'idéal de spécialisation porté dans l'espace anglo-saxon. Il s'agit d'une activité de niche, pratiquée essentiellement par quelques autodidactes passionnés. Comment expliquer cette situation? Par un manque d'intérêt des journalistes ou de formations spécialisées? Par des difficultés d'accès aux données? Ou par un manque de retour sur investissement pour des projets mobilisateurs de ressources? Nos témoins – qui font partie de cette poignée de journalistes intéressés par le journalisme de données, bien qu'aucun ne se revendique comme "datajournaliste" – estiment que chacun de ces arguments participe à expliquer cette situation de *statu quo*. On retrouve des traces de la matérialité des démarches relevant du journalisme de données, en Belgique, dans la base de données des *Data Journalism Awards*, du nom du seul prix international de journalisme dédié à cette approche spécifique. Sur les 2.439 projets répertoriés en 2018, 25 étaient Belges et la majorité d'entre eux émanait de journalistes néerlandophones. À titre de comparaison, 86 projets étaient Français, 43 Néerlandais, 106 Allemands, 263 Britanniques et 507 Américains sur un total de 86 pays représentés¹⁷. L'existence du journalisme de données en Belgique francophone serait donc moins matérielle que discursive. C'était d'ailleurs la conclusion de la seule recherche consacrée au phénomène, "*Waiting for Data Journalism*" (De Maeyer *et al.* 2015). Celle-ci s'était appuyée sur vingt entretiens semi-structurés avec une diversité d'acteurs im-

¹⁴ "Le groupe Rossel choisit la plateforme Wibbitz pour faciliter la production de ses contenus vidéos", OffreMedia.com, 24/05/2018, consulté le 06/07/2018, <https://www.offremedia.com/le-groupe-rossel-choisit-la-plateforme-wibbitz-pour-faciliter-la-production-de-ses-contenus-vidéos>

¹⁵ Communication personnelle, recueillie par courriel, le 25/04/2016.

¹⁶ Communication personnelle, recueillie par courriel, le 09/12/2018.

¹⁷ Consulté le 08/07/2018, URL : <http://community.globaleditorsnetwork.org/projects>

pliqués dans l'adoption et le développement du journalisme de données en Belgique francophone, en ce compris des responsables de la gestion des ressources humaines de rédactions et des personnes en charge d'organiser la formation des journalistes. La méthode avait été complétée par un corpus de 52 documents récoltés sur une période de deux ans, lesquels étaient relatifs à des programmes de formation, des articles de blogs et à un artefact "occasionnel" du journalisme de données. Les chercheurs avaient également participé à des événements relatifs à cette pratique, telles que des réunions du groupe belge de l'organisation internationale "Hacks/Hackers"¹⁸ et à des ateliers de journalisme de données, consistant en des observations informelles. Les résultats avaient démontré que la représentation du journalisme de données n'était pas partagée par l'ensemble des répondants, dont certains considéraient que la production d'une visualisation de données, c'est déjà du datajournalisme¹⁹. Parfois initiés aux pratiques du datajournalisme lors de formations de courte durée, le profil de la poignée de datajournalistes belges francophones est plutôt celui de journalistes formés sur le tas, présentant un intérêt personnel pour la démarche.

1.2.1 Journalism de données en Belgique francophone

L'espace médiatique belge francophone est caractérisé par l'étroitesse de son marché – la Belgique compte environ 40% de francophones. Il n'y existe pas, à proprement parler, d'équipes rédactionnelles spécifiquement dédiées à une approche par données. En presse quotidienne, la tendance est à l'organisation de pôles tantôt orientés vers un journalisme d'enquête s'appuyant sur des données et des outils technologiques – comme c'est le cas au *Soir* dans une "cellule" au sein de laquelle travaillent trois journalistes –; tantôt multidisciplinaires où collaborent des journalistes et des développeurs, et où une approche par données s'inscrit dans le cadre plus large d'un storytelling numérique (comme c'est le cas à *L'Avenir* et à *L'Echo*). Cet état des lieux se concentre sur les médias qui participent à cette dynamique, ainsi qu'aux formations abordant le journalisme de données. Il s'appuie, pour ce faire, sur des témoignages de professionnels ainsi que sur des ressources consultées en ligne.

Journalisme de données au *Soir*

Joël Matriche (*Le Soir*)²⁰ fait partie de ces passionnés du datajournalisme. Il explique avoir suivi plusieurs formations en analyse et traitement de données "*par intérêt et par envie d'apprendre*" mais, souligne-t-il, il s'est aussi auto-formé. Il estime que les journaux belges francophones sont encore réticents à investir dans l'engagement de statisticiens attachés aux rédactions ou à mobiliser des développeurs pour ce type de projet, en raison d'un manque de budgets, dans une période de crise, pour un "*retour sur investissement qui n'est pas encore évident*". Mais

¹⁸ "Hacks/Hackers est organisation américaine transnationale, qui a pour objet de réunir des journalistes et des programmeurs en vue de susciter des collaborations (Lewis & Usher 2014).

¹⁹ Cette vision réductrice pose raisonnablement la question de savoir si l'accessibilité d'outils de visualisation en ligne aboutit à considérer le journalisme de données non pas comme une pratique journalistique nourrie par l'analyse de données mais comme une forme de journalisme qui serait "gadgétisée".

²⁰ Communication personnelle, Liège, mars 2014 et communication personnelle, recueillie par courriel le 08/09/2017.

"les choses changent" progressivement : "Par exemple, L'Echo a embauché un jeune journaliste orienté vers le traitement de données. Au Soir, mon collègue Alain Jennotte travaille maintenant tout son temps sur le traitement de données. Il n'est pas développeur professionnel mais est passionné, doué et apprend sur le tas". La mise en relation des projets de datajournalisme avec la rédaction web n'est pas plus évidente, estime-t-il, passant d'une situation "de punition, de goulag" à une mise en avant des effectifs en ligne qui, toutefois, "restent réduits". Pour travailler avec des données, "il ne faut pas craindre les chiffres. Il faut aussi savoir se servir d'un tableur et connaître un minimum les techniques de scraping"²¹. Outre une question de budgets, une autre raison expliquant le faible développement du journalisme de données serait, selon lui, à trouver dans "un manque de données disponibles".

Journalisme de données à L'Écho

Pour Nicolas Becquet, responsable du développement et de la transformation numérique de L'Écho, les données sont l'essence même du journalisme économique. À l'origine, le pôle multimédia qu'il pilote était commun avec la rédaction néerlandophone du *Tijd* (groupe Media-fin), dans le cadre d'une stratégie multicompétence : développement, data, multimédia. Une réorganisation de ces services a abouti, en 2015, à une séparation en deux pôles linguistiques "pour se rapprocher des identités éditoriales de chacun des titres. À partir de ce moment-là, on a eu des profils multimédias polyvalents, constitués de jeunes journalistes qui ont appris à faire des cartes et des data pour apporter une plus-value aux articles. J'ai engagé, en août 2017, un journaliste présentant un profil de datajournaliste : il en a envie et il a des connaissances périphériques, telles que la statistique et la visualisation de données". Du côté néerlandophone, l'accent a été placé sur un recrutement de trois profils spécialisés. "On y retrouve un spécialiste data 'hard-core' qui travaille sur un 'dashboard' basé sur Google Analytics. Il est capable de fouiller et comprendre un très grand nombre de données. Il y a un architecte de l'information, spécialisé dans l'expérience utilisateur et la visualisation de données. Surtout, il a des notions de code qui lui permettent de suivre toute la chaîne de production de l'information. Enfin, la troisième personne de ce pôle est un développeur 'front-end' qui sait aussi travailler la data. Il a mis en place une première forme d'automatisation à partir de données sur les communes belges, via le principe de phrases à trous"²². Dans les publications de L'Écho, "le datajournalisme est davantage abordé en termes de datavisualisations, et ça reste encore compliqué de travailler, en profondeur, avec un journaliste, sur les données. Ce n'est pas encore intégré de manière naturelle. La plus-value est d'abord vue comme la possibilité de créer du contenu multimédia. À l'inverse, mes collègues néerlandophones peuvent passer plusieurs jours pour traiter une information basée sur les données. En termes d'audiences, cela reste des contenus de niche".

Le retour sur investissement d'un projet de datajournalisme peut être envisagé de deux manières, indique-t-il : sur le plan de la formation continue et sur celui de la volonté d'assumer un coût de production qui permette de se positionner, à long terme, "comme un média de référé-

²¹ Récupération de données sur des documents numériques.

²² Communication personnelle, Bruxelles, le 14/11/2017.

rence". En novembre 2017, il déposait un dossier de financement auprès de Google Digital News Initiative, en vue de développer un système de traitement automatisé de données boursières. Il porte pour nom de code "Quotebot" et fait l'objet de la seconde étude de cas présentée dans cette thèse²³. *"C'est une chance pour faire de la recherche et du développement. Cela nous met en mouvement, et nous permet de nous faire réfléchir sur nos pratiques, usages et modèles financiers. (...) On développe déjà des outils destinés aux investisseurs, cela compléterait notre palette. En tant que journaliste, je trouve que c'est une aide. Ce n'est pas, en soi, de l'ingénierie hors de portée. Depuis que je suis journaliste multimédia, j'ai toujours utilisé des outils accessibles"*. Nicolas Becquet épingle un autre aspect relevé dans deux recherches à propos de l'enseignement d'un journalisme piloté par les données (Berret & Philips 2016, Schmitz-Weiss & Riva 2018) : celle de la difficulté de travailler avec les chiffres lorsque l'on est habitué aux lettres. *"Ce n'est pas naturel, pour moi, de travailler avec des chiffres. Ma formation est plutôt littéraire. Si ce projet se fait, ce serait plutôt en collaboration avec les journalistes qui traitent ces matières et pas avec un datajournaliste : je ne peux pas construire des outils sans consulter les personnes qui vont les utiliser. J'ai besoin de leur expertise"*.

En tant qu'observateur des médias numériques en Belgique francophone, il estime qu'une approche par données est peu répandue en raison de plusieurs facteurs : *"Je pense que l'on passe à côté du potentiel. Les médias sont confrontés à une série de difficultés et de défis qui font que ce n'est pas la priorité. C'est perçu comme un domaine où il faut investir et où il faut embaucher des profils spécifiques pour un retour sur investissement qui n'est pas immédiat. L'approche d'Arnaud Wéry à L'Avenir est très intéressante dans une démarche locale. Sinon, en Belgique francophone, on en reste souvent à Instagram ou à Datawrapper en versions non payantes, avec les logos du prestataire de services. Avec des compétences techniques, on peut construire des outils plus pratiques pour l'utilisateur. C'est une question de priorité. Beaucoup de journalistes ont envie de creuser ces aspects, mais cela demande du temps et des compétences spécifiques. Si je devais engager un profil aujourd'hui, ce serait quelqu'un qui comprend la manière dont fonctionnent les interfaces et l'information. Le datajournalisme, c'est une spécialisation métier très forte qui nécessite des compétences très spécifiques"*. Il estime également que la situation belge est très différente de celle observée en France, où les rédactions sont plus grandes et où la production matérielle du datajournalisme est souvent externalisée *"à l'échelle de projets majeurs ou de projets vitrines. Ils ont des moyens mais pas forcément au quotidien. Si l'on prend le cas du Monde, par exemple, on est encore très en arrière en Belgique"*.

Journalisme de données à L'Avenir

À l'échelle locale, *L'Avenir* se démarque en proposant régulièrement des projets de datajournalisme qui prennent souvent la forme de cartes interactives. Ce n'est pas, pour autant, que Arnaud Wéry, journaliste attaché au weblab du quotidien, se définit comme un journaliste de données. *"C'est quelqu'un qui, pour moi, a de très bonnes connaissances en tableurs et en nettoyage de données. J'ai de bonnes bases pour faire ce que je dois faire dans le cadre de mon travail"*

²³ Voir infra p.208

mais elles ne sont pas suffisantes, faute de temps à consacrer à l'apprentissage. Je suis un peu plus à l'aise avec les outils de visualisation de données car j'en ai l'expérience. J'apprends sur le tas, par essais-erreurs. Mais travailler avec des développeurs, cela ouvre des perspectives : avoir des connaissances du code est devenu nécessaire". Depuis mars 2017, époque de la création de la cellule "weblab" au sein du journal, il travaille avec un développeur, Cédric Dussart, l'un des premiers à être reconnu au titre de journaliste professionnel en Belgique francophone. *"Dans le cadre de notre travail, il met régulièrement sa casquette de journaliste. Il va beaucoup voir ce qui se fait ailleurs, je n'avais jamais connu ça avant"* ²⁴.

Pour cet acteur du datajournalisme, la question des compétences ne peut être éludée. À l'instar de Nicolas Becquet, il souligne qu'il s'agit de compétences spécialisées. Tout comme à *L'Echo*, le travail sur les données ne constitue qu'une partie du travail de *storytelling* développé dans le cadre d'une cellule dédiée. *"Au quotidien, un journaliste et un développeur peuvent travailler sur des longs formats ou sur les réseaux sociaux. On peut donner l'impulsion aux rédactions locales sur des thématiques particulières reposant sur des données, tout comme celles-ci peuvent également nous solliciter. On travaille avec des données par opportunité. Il y a un sillon que l'on creuse davantage, qui est celui de l'utilisation des données pour apporter un service, tels que la localisation des plaines de jeux dans les communes ou celle des zones de baignade. La data, ce n'est qu'un aspect parmi tant d'autres, d'autant que la visualisation n'est pas une fin en soi. De plus, il faut disposer de données permettant de trouver des histoires. Une dataviz pour une dataviz, c'est quelque chose qui me frustre un peu. On essaie d'être cohérents, tout en composant avec les moyens du bord"*.

Le frein principal au développement du datajournalisme ne serait pas seulement une question de compétences, mais aussi d'accès aux données. *"Après, c'est de savoir d'où elles viennent, de comprendre la manière dont elles ont été récoltées, leur éventuelle marge d'erreur, ... Cet accès se met en place petit à petit, mais on peut se poser la question de la valeur ajoutée de données ouvertes qui vont lister les urinoirs publics dans une ville. Des données sont aussi bel et bien disponibles, mais la manière dont elles sont organisées ne permettent pas un traitement journalistique car l'on n'y trouve aucun sens"*. Il perçoit l'automatisation de la collecte et du traitement de données comme un outil, tout en s'interrogeant sur un éventuel effet de mode. À la question de pouvoir identifier des datajournalistes évoluant dans le monde francophone, il cite quatre noms : un premier était journaliste indépendant mais a annoncé une reconversion professionnelle dans la fiche qui lui est dédiée sur le site *journalistefreelance.be*, tandis qu'un deuxième est sorti de la profession pour se lancer dans la recherche académique. Les deux suivants sont tous deux journalistes au quotidien *Le Soir*, Joël Matriche et Alain Jennotte. Il leur reconnaît des compétences beaucoup plus avancées : *"Le Soir a une longueur d'avance grâce à eux. Les rédactions françaises ont bien plus d'avance que nous, mais les nôtres sont faites de bric et de broc. On est de plus en plus serrés et les profils sont différents"*.

²⁴ Communication personnelle, Waremme, le 10/11/2017.

Malgré une lente progression des méthodes journalistiques s'appuyant sur les données et leur traitement à l'aide de technologies numériques, le milieu professionnel semble ouvert à ces nouveaux profils issus du monde de la technique. Cédric Dussart, qui se décrit sur son profil Twitter comme "front dev"²⁵, est au sein du weblab de *L'Avenir* avec Arnaud Wéry. Il se définit comme "une espèce d'hybride" dont le positionnement ne va pas toujours de soi : *"En fonction de la personne face à laquelle je me trouve, je vais me présenter comme journaliste ou comme développeur. Même au sein de la rédaction, au début, les gens se demandaient ce que je faisais là. Aujourd'hui, je pense que je suis davantage considéré comme un développeur web. Certains journalistes me considèrent comme une espèce de magicien fou et c'est plutôt positif. Je ne pense pas que la technologie y soit perçue de manière négative"*²⁶.

Après des études en infographie, il a débuté sa carrière dans le domaine de la communication. Peu intéressé par ce secteur, il a rejoint le groupe *L'Avenir* en 2013. *"Au début, je travaillais pour le département marketing mais, au bout d'un an, j'ai commencé à travailler pour la rédaction. J'y suis attaché depuis un an à mi-temps et ce travail me parle beaucoup plus qu'en communication. J'y trouve plus de sens dans mon boulot"*. Début 2018, le weblab de *L'Avenir* se voyait décerner le prix du journalisme numérique du Parlement de la Fédération Wallonie-Bruxelles. Pour lui, il s'agit de la reconnaissance d'une expérience qu'il envisage comme *"une forme de prise de risques car si les retours ne sont pas bons, elle pourrait ne pas se poursuivre. Dans le monde francophone, il y en a très peu qui osent ce genre de choses. Ça nous donne une base solide. Ça nous pousse"*.

Dans son travail quotidien, Cédric Dussart *"code beaucoup"* mais pas seulement. *"Je fais aussi du traitement de données, du scraping, de l'Excel. Mais ça reste beaucoup de code. J'ai rarement des contacts avec les sources d'information. Là où le journaliste n'a plus la main pour valider des choses, c'est moi qui m'en occupe. (...) Le travail du code suppose aussi qu'il faut que je me tienne à jour et cela me prend du temps en dehors de mon travail. Si j'en étais resté à ce qui existait à l'époque où je suis sorti de l'école, je pense que je n'aurais pas de boulot. J'aimerais bien tester une formation de l'AJPro²⁷ en motion design mais j'ai peur que ce soit trop généraliste et pas assez technique"*. Ce qui a changé pour lui depuis 2017, époque de son agrégation au titre de journaliste professionnel ? *"J'ai deux jours de congés en plus par an!"*. Ce qui l'a incité à soumettre sa demande ? *"Je pense, la protection et la reconnaissance de mon travail. Je ne suis pas qu'un développeur, j'ai aussi une part de journaliste. Travailler sur une base de données et pouvoir en tirer des messages, c'est du journalisme. Mais les explications relèvent encore du travail du journaliste, qui garde la main sur le travail d'analyse. Mes choix techniques sont validés en collaboration avec lui. La rédaction est ouverte au weblab et, aujourd'hui, certains journalistes viennent nous solliciter de manière spontanée"*.

²⁵ C'est à dire, spécialisé dans le design d'interfaces web.

²⁶ Cédric Dussart, communication personnelle, Namur, le 29/01/2018.

²⁷ AJPro est la structure de formation continuée de l'Association des journalistes professionnels. Elle a été lancée en 2013.

L'offre en formations

En Belgique francophone, les journalistes spécialisés dans la collecte, l'analyse et le traitement de données se comptent sur les doigts d'une main. Ils ont en commun d'avoir acquis leurs compétences sur le tas, plus rarement dans des formations spécialisées. Si l'ensemble de nos témoins s'accordent sur le fait que la pratique du journalisme de données nécessite des compétences fortes, l'offre en formations ne suit pas forcément. En matière de formation initiale, les programmes des universités intègrent progressivement des cours liés à une approche par données dans le journalisme. Pour l'année académique 2016-2017, trois cours spécifiques étaient identifiés en master de journalisme : "Visualisation de données" (ULB, 5 ECTS); "Datajournalisme" (analyse dans un tableur et visualisation à l'aide d'outils gratuits, IHECS, 5 ECTS), "Initiation au datajournalisme" (ULG, 15 heures). En matière de formation continue, l'AJPro organise assez peu de formations dans ce domaine. Celles-ci relèvent davantage d'une démarche introductive. Elles ont pour point commun d'être dispensées par les quelques journalistes autodidactes rencontrés dans le cadre de cette recherche. L'examen de ses programmes, s'échelonnant de 2014 à 2020, montre que 174,5 heures de formation ont été organisées pour ces sept années, soit une moyenne de près de 25 heures par an. Ces formations sont essentiellement de deux types : extraction de données (*web scraping*) et visualisation de données. Deux formations (extraction et fouille de données) nécessitaient des prérequis, non pris en charges par la structure organisatrice : d'une part, des connaissances de base en informatique et, d'autre part, des connaissances de base des tableurs et du web. Ces formations ne font pas nécessairement le plein de candidats. Mehmet Koksak, journaliste indépendant et observateur des médias en Belgique francophone, relate l'expérience d'une formation gratuite en journalisme de données, organisée en 2015, où *"on a eu du mal à trouver un public. On avait galéré pour remplir la salle"*²⁸.

À titre de comparaison, la Nordic Datajournalism Conference (NODA), organisée chaque année en Scandinavie, propose plusieurs sessions de formation en journalisme de données. En 2018, celles-ci portaient sur la pratique du datajournalisme pour débutant (récolte, nettoyage et analyse de données en 7 heures de formation), sur le langage de programmation Python pour débutants (récolte, analyse et visualisation de données en 7 heures de formation), et sur l'apprentissage du journalisme à des développeurs (également en 7 heures de formation)²⁹. Cela représente 21 heures de formation organisées en une seule journée. En 2017, la Global Investigative Journalism Conference, qui réunit tous les deux ans des journalistes d'investigation du monde entier, a proposé simultanément, pendant trois jours, quinze courtes formations axées sur la collecte, l'analyse et le traitement des données. Celles-ci ont abordé une variété des facettes des technologies utilisées dans le cadre d'une approche par données dans le journalisme. Il était possible d'y suivre des ateliers pour un total d'une vingtaine d'heures à panacher entre les quinze formations du programme, certaines proposant un niveau débutant et d'autres un

²⁸ Communication personnelle, Bruxelles, le 06/02/2018.

²⁹ Source consultée le 13/03/2018, URL : <https://noda2018.se/competence-courses/>

niveau avancé³⁰. Cette conférence a accueilli 1.200 journalistes provenant de 127 pays. Parmi les participants, tous n'étaient pas rompus à une approche par données dans le journalisme. Leur point commun est à trouver dans leur investissement dans un travail journalistique s'inscrivant dans le temps long de l'investigation. *"Mais au GIJN, la communauté des journalistes adore les chiffres alors qu'une majorité de journalistes les détestent"*, pondère Mehmet Koksak, pour qui il existe *"clairement un manque d'intérêt"* pour une approche par données dans le journalisme³¹.

La spécialisation participe à une logique de professionnalisation qui permet d'établir un monopole en maintenant la rareté (Sarfatti-Larson 1977, cité par Lewis 2012). Mais cette voie reste peu explorée en Belgique francophone, alors qu'elle peut permettre d'apporter de la valeur ajoutée en termes de compétences et, potentiellement, de faire la différence sur le marché de l'emploi. Si de nombreuses expériences en journalisme numérique sont le fait de générations de "bidouilleurs", cette posture n'est toutefois pas suffisante pour garantir de hauts standards dans les pratiques. Pour autant, les rédactions sont-elles prêtes à s'engager dans cette voie? Dans un billet publié sur son blog à propos de son expérience de sept ans en journalisme multimédia, Nicolas Becquet reconnaît faire partie de cette première génération tout en indiquant que la polyvalence technique, intégrée dans la plupart des écoles de journalisme, reste encore peu exploitée : *"Elle n'est toujours pas considérée comme une source de plus-value, mais plutôt comme une capacité à alimenter un flux et des 'canaux' numériques"*³². Il souligne également que les standards de qualité ont évolué : *"À l'heure des applications 'responsive', se lancer dans un webdoc est devenu une gageure. Visualiser quelques données avec Datawrapper ou avec Infogram paraît bien dérisoire face aux productions des datajournalistes chevronnés. Le Facebook Live est devenu une affaire de réalisateurs télé (...) Que ce soit pour la vidéo, la data, l'édition multicanal ou les dominantes thématiques, les médias ont plus que jamais besoin d'experts (fond et forme) capables de répondre aux exigences de l'audience en termes de qualité, de pertinence et d'usage"*.

L'arrivée de nouveaux acteurs dans la profession, issus du monde de la technique, pose la question d'une logique inclusive qui ne devrait pas s'arrêter au seul fait de reconnaître que ceux-ci font partie du monde du journalisme en leur délivrant une carte de presse. Membre de la Commission francophone d'agrégation au titre de journaliste professionnel et observateur des pratiques de journalisme des données, auxquelles il s'intéresse depuis plusieurs années, le journaliste indépendant Mehmet Koksak se dit favorable à l'octroi de la carte de presse à toute personne travaillant dans un média qui en fait la demande. L'évolution des tâches, observe-t-il, a pour conséquence d'estomper les frontières entre les techniciens³³ et les journalistes. Il

³⁰ Source consultée le 24/12/2017, URL : <https://gijc2017.org/program-gijc17/>

³¹ Communication personnelle, idem.

³² "De la nécessaire polyvalence à l'indispensable (re)spécialisation des journalistes. Ce que j'ai appris en sept ans de journalisme multimédia", Nicolas Becquet, 10/09/2017, consulté le 17/09/2017, URL : <http://mediatype.be/2017/09/10/necessaire-polyvalence-indispensable-specialisation-journalistes/>

³³ Des cartes "T" sont délivrées aux techniciens, de manière à faciliter leur identification lorsqu'ils assistent les journalistes. Source : Association des Journalistes Professionnels, consulté le 23/04/2019, URL : <http://www.ajp.be/documents-technicien/>

ajoute, non sans humour : *"J'attends le jour où un robot journaliste pourra être reconnu, cela viendra peut-être d'ici vingt ou trente ans et, à ce moment-là, la Commission aura peut-être été elle-même remplacée par un robot!"*³⁴. Une approche inclusive suppose aussi de mettre en place des formations continues qui soient destinées à ces nouveaux acteurs du monde du journalisme, qu'elles portent sur les bonnes pratiques, les routines journalistiques ou la logique éditoriale. Ces formations devraient également leur faciliter l'intégration de cette idéologie professionnelle du journalisme qui donne du sens à l'activité d'informer (Deuze 2005). En d'autres termes, il s'agirait de leur permettre de développer une forme de pensée journalistique³⁵. Ce type de formation n'existe plus en Belgique francophone depuis la disparition de l'Institut de journalisme en 2013. Cette institution, fondée en 1922, dispensait des cours du soir en journalisme, ouverts à un public provenant d'autres secteurs que celui des médias et qui souhaitait se réorienter dans le journalisme. Cédric Dussart a indiqué avoir cherché ce type de formation mais n'avoir rien trouvé. *"J'en ai parlé à Arnaud Wéry qui a estimé que cela n'était pas nécessaire car mon travail sur le terrain allait plus m'apporter qu'un diplôme en journalisme. Ce qui m'aurait intéressé, c'est de savoir comment traiter un sujet à partir d'une idée ou connaître l'aspect éditorial d'un long format. C'est un métier qui, je pense, s'apprend plus facilement que la technique. Et puis, les journalistes sont souvent réfractaires à l'apprentissage du code. (...) Etant donné la manière dont évoluent les médias, je pense que cet apprentissage du code est indispensable. Mais cela va créer de la concurrence dans mon boulot (rires). Je ne vois pas l'un sans l'autre et, malheureusement, le futur n'est pas dans le papier"*. Dans le sens inverse, il se dit régulièrement sollicité par les journalistes de la rédaction *"pour apprendre quelques petits trucs de mise en page. Le journaliste a assez bien conscience de la difficulté de ce travail"*³⁶.

Malgré les effets stimulants de la polyvalence des tâches, celle-ci ne permet pas d'exceller dans un type de récit particulier. Elle nécessite du temps, dans l'apprentissage et la pratique, alors que la rapidité est une norme. La professionnalisation de ces nouvelles formes de journalisme suppose que les pas doivent se faire dans les deux sens : des journalistes vers la technique et des techniciens vers le journalisme. Mais en Belgique francophone, il semble que nous n'en soyons pas encore là. Que ce soit sur le plan collectif ou individuel, l'intégration de pratiques professionnelles supposant des compétences techniques reste marginale et limitée à un nombre restreint d'expériences éditoriales. Cela peut s'expliquer par le nombre réduit d'acteurs en activité sur le marché de l'emploi et un manque d'intérêt des médias. Une autre raison serait à trouver dans le manque de formations spécialisées, induisant que seuls les journalistes motivés par l'auto-formation s'engagent sur cette voie. De plus, observe Arnaud Grégoire, producteur numérique et ancien journaliste en Belgique francophone, *"les journalistes ont tendance à se comporter comme des divas face à la technologie. Ceux qui pratiquent ce métier depuis une dizaine d'années n'ont pas de bonnes raisons d'évoluer autrement. J'ai l'impression d'avoir passé ma vie de journaliste à affronter les résistances dans les rédactions au sein desquelles j'arrivais, et à faire du déplacement de montagne"*³⁷.

³⁴ Communication personnelle, ibidem.

³⁵ Voir "Développer une pensée journalistique", p.76

³⁶ Cédric Dussart, idem.

³⁷ Communication personnelle, Bruxelles, le 11/09/2017.

1.2.2 Journalisme de données en Flandre

Les pratiques du journalisme de données ne sont pas plus répandues en Flandre, mais celles-ci y sont encouragées, que ce soit en matière de formation continuée ou de soutiens financiers publics. Le journaliste Maarten Lambrechts fait figure de pionnier dans ce paysage médiatique. Se définissant lui-même comme journaliste de données, il a fait ses premières armes au quotidien *De Tijd*. Depuis le début 2017, il exerce cette activité, qu'il a élargie à celle de consultant en datavisualisations, en tant qu'indépendant : "*Aujourd'hui, je réalise toutes sortes de travaux basés sur la visualisation de données, mais aussi du journalisme de données. J'essaie donc de trouver des histoires dans de grands ensembles de données et des bases de données, puis je raconte ces histoires en combinant du texte avec des graphiques et des cartes*"³⁸. Interviewé en 2016 à propos du nombre de journalistes de données en Belgique, il répondait : "*J'aimerais avoir plus de collègues dans ce domaine, mais je suis certains que beaucoup de personnes qui travaillent dans celui de la business intelligence font des choses similaires. Peut-être qu'ils n'investissent pas autant dans le storytelling et dans la visualisation comme je le fais, mais il y a certainement d'autres profils que le mien avec des connaissances et compétences similaires. Seulement, ils ne sont pas dans les rédactions aujourd'hui*"³⁹.

En 2018, le gouvernement flamand dégageait une enveloppe de 500.000 euros pour soutenir l'innovation dans les médias⁴⁰. Par innovation, il désigne tant les modes de production de l'information que le fact-checking, les modèles d'affaires alternatifs... que le journalisme de données⁴¹. Sur le plan de la formation, Mediacademie⁴² (une initiative de quatre associations de médias flamands – Vlaamse News Media, The Ppress, UPP et VKUPP – soutenue par le ministère flamand des médias), est investie de longue date dans le journalisme de données. Son objet est de doter les professionnels des médias des compétences journalistiques, commerciales et techniques nécessaires au maintien de la qualité de la presse écrite. Par exemple, en 2014, la structure a organisé deux demi-journées de formation incluant l'extraction de données (formation dispensée par Paul Bradshaw, datajournaliste, en charge de la direction du master en datajournalisme de l'université de Birmingham, au Royaume-Uni), les outils *open source* pour la collecte et la visualisation (par Hille van der Kaa, coordinatrice de la rédaction du quotidien régional néerlandais BN De Stem et chercheuse spécialisée dans le domaine de la production automatisée d'informations). L'International Research Journalistiek⁴³, émanation du Fonds Pascal De Croos, qui soutient le journalisme d'investigation via l'octroi de bourses, propose également des formations spécialisées : leur objectif annoncé est d'acquérir de nouvelles compétences en matière de recherche, d'analyse et de visualisation de données.

³⁸ "Marteen Lambrechts, freelance data journalist", Glyn Moddy, Copybuzz, 2018, consulté le 23/04/2019, URL : <https://copybuzz.com/fr/copyright/humansofcopyright-maarten-lambrechts-freelance-data-journalist/>

³⁹ "Data journalism blog : Maarten Lambrechts on the underused power of explorable explanations & more", Anastasia Valeeva, Open Belgium 2016, 14/09/2016, consulté le 23/04/2019, URL : <http://2016.openbelgium.be/blogpost/data-journalism-blog-maarten-lambrechts-underused-power-explorable-explanations-more>

⁴⁰ Précisons que cette aide a été allouée de manière ponctuelle, les budgets dédiés à la culture – en ce compris aux projets journalistiques – ayant été réduits de manière drastique à la fin 2019.

⁴¹ Source : <https://cjsm.be/media/themas/pluralisme/innovatieve-journalistieke-projecten>

⁴² Source : <http://www.mediacademie.be>

⁴³ Source : <https://researchjournalistiek.eu/>

La Dutch-Flemish Association for Investigative Journalists⁴⁴ (VVOJ), une organisation belge-néerlandaise dédiée au journalisme d'investigation qui compte 650 membres, a pour objectif "d'améliorer les méthodes d'investigation classiques et à présenter les méthodes de recherche les plus récentes impliquant l'utilisation des technologies de l'information, la narration interactive et le journalisme de données". Pour ce faire, elle organise régulièrement des formations ciblées. S'ajoute à cet arsenal l'événement *Data Harvest*, organisé chaque année à Malines. Quatre jours durant, les méthodes de recherche et d'enquête autour du journalisme de données font l'objet de modules de formation intensifs. En 2019, ceux-ci s'axaient sur l'acquisition de compétences techniques incluant la programmation informatique : analyse de données via des feuilles de calcul, lignes de commande en CSV, introduction à la programmation en R et en Python, cartographie avec le logiciel QGIS, bases de données SQL et Graph... et développement de code pour la production automatisée d'informations. Cette dernière formation était assurée par le datajournaliste suisse Mathias Born, qui a participé à la conception de "Tobi" au sein du groupe Tamedia. Cette application automatisée fournit les résultats des votations – une forme de consultation populaire – organisées en Suisse⁴⁵.

Dans l'enseignement, l'*Erasmus Hogeschool* (Bruxelles) organise, dans le cadre d'un bachelier en journalisme, des formations en journalisme de données en tant que pratique destinée à soutenir un *slow* journalisme. Il s'agit donc d'une approche que l'on peut considérer comme plus pointue que ce qui est actuellement proposé dans l'espace francophone, mais celle-ci s'inscrit dans le champ du journalisme d'investigation. On observe également des différences quant à la manière d'aborder la question des compétences entre les journalistes issus des deux communautés linguistiques du pays. Aussi, en 2018, les journalistes néerlandophones estimaient-ils plus important de disposer de compétences multimédia que les francophones (87,6% contre 78,2%). Ils étaient également plus nombreux à considérer que les compétences liées à l'interprétation et à l'usage des statistiques sont plutôt très importantes (69,4% contre 58,8%) (Van Leuven *et al.* 2019).

1.3 Enjeux et défis professionnels

Les enjeux professionnels de la production automatisée d'informations apparaissent comme multiples à négocier. Cela concerne tant la manière d'intégrer les journalistes dans des processus de conception dont ils sont souvent tenus à l'écart, que celle dont les systèmes d'information peuvent être envisagés en termes de complémentarité dans leurs pratiques professionnelles. Cela pose aussi la question de la place du journaliste dans un monde de machines. Pour Van Dalen (2012), ce phénomène entre en conflit avec une définition du journalisme axée sur les seules tâches et compétences du professionnel. De l'analyse de grandes quantités de données à la diffusion d'informations, l'automatisation peut modifier la manière dont les journalistes travaillent (Van Dalen 2012). Elle ne contribue pas seulement à redéfinir le rôle du

⁴⁴ Source : <http://www.vvoj.nl>

⁴⁵ Source : EIJC 2019, URL : <https://eijc19dataharvest.sched.com/>, consulté le 10/05/2019. Voir aussi p.107

journaliste, son identité et son autorité, mais aussi le sens de son métier d'informer (Neveu 2010). Un autre enjeu consiste à intégrer de nouveaux acteurs dans le monde du journalisme, alors qu'ils n'y sont pas traditionnellement liés et que la plupart d'entre eux estime ne pas y appartenir. C'est aussi ce en quoi les technologies d'automatisation participent au floutage des contours du journalisme, dans un monde où l'acte d'informer n'est plus, depuis quelques années, l'apanage des seuls professionnels de l'information (Mercier & Pignard-Cheynel 2014). Cette question du "flou" a été abordée par Ruellan (1992), qui soulignait déjà l'importance de ne pas cantonner le journalisme à une vision purement fonctionnaliste qui se limiterait aux compétences techniques ou à l'intégrité du professionnel : aucune profession ne peut être définie par la seule somme des tâches qui la caractérisent (et qui peut d'ailleurs varier pour une seule et même profession).

Si les "robots" font le job au sein des entreprises de presse, quelle place restera-t-il aux journalistes? Régulièrement posée, la question d'un "monde sans journalistes" ne se réduit plus au seul mode de la provocation : certains professionnels y voient une menace bel et bien réelle. La dégradation constante des conditions de travail des journalistes, couplé à un phénomène de précarisation, que ce soit en termes de revenus ou en termes de statut professionnel, ne permettent pas d'affirmer que les technologies d'automatisation se déploient dans un contexte socioprofessionnel favorable. Le journalisme est une profession fragile parce que les entreprises de presse le sont aussi, confrontées à la difficulté de trouver un modèle d'affaires viable pour leurs activités d'information en ligne, tandis que le marché publicitaire leur échappe au profit de nouveaux acteurs dont l'objet est moins d'informer que de drainer des audiences mondiales. Aussi, les relations entre les journalistes et les systèmes d'automatisation peuvent être non seulement techniques, cognitives ou sociales : elles revêtent également une forte dimension économique (Ekbja & Nardi 2014).

En provoquant une série de changements de paradigmes, les technologies d'automatisation de la production d'informations peuvent être considérées comme disruptives (Kostoff *et al.* 2004). Elles bouleversent à la fois un mode de production jusqu'alors dévolu à des humains, et elles remettent en question la conception même de ce qu'est le journalisme et la manière dont celui-ci fait sens dans un monde social. Cette section aborde ces enjeux et défis sous le prisme de trois variables qui participent à forger les imaginaires des journalistes : (1) une concurrence sur l'emploi des journalistes, qu'elle soit avérée ou fantasmée ; (2) une relation homme-machine placée sous tension, dès lors qu'elle questionne les spécificités des journalistes ; (3) une mobilisation des compétences, dans le cadre d'un (re)positionnement professionnel.

1.3.1 Gérer les angoisses professionnelles

À l'instar de tout imaginaire, indique Musso (2008), "*les représentations sociales de la technique sont ambivalentes : elles peuvent aisément se retourner en leur contraire (...) Les techniques annoncent aussi des régressions comme la suppression des emplois, des qualifications, l'aliénation à la machine sur le modèle Frankenstein (...)*". Dans cette perspective, la production automatisée d'informations peut être abordée sous l'angle d'une concurrence pouvant être per-

que comme une menace pour la sécurité de l'emploi des journalistes qui effectuent des tâches routinières de base (Van Dalen 2012), alors que les journalistes les plus expérimentés ne seraient pas menacés (Linden 2017). De plus, lorsque des routines journalistiques peuvent être automatisées, les journalistes sont contraints d'offrir de meilleures productions en devenant encore plus créatifs dans leur écriture, en offrant une meilleure couverture médiatique à leurs audiences, et en approfondissant le contexte de l'information qu'ils traitent (Van Dalen 2012).

Les politiques de convergence, menées dans la foulée de l'avènement du *World Wide Web*, avaient placé les journalistes sous la pression d'un temps de plus en plus réduit dans le contexte d'une production multisupports, s'agissant d'acquérir de nouvelles compétences pour maîtriser la chaîne de production de contenus numériques (Bardoel & Deuze 2001). Les exigences liées à la polyvalence des tâches furent souvent considérées comme un moyen d'augmenter la productivité de l'entreprise de presse au détriment de l'emploi et de sa qualité. Plus largement, nous pouvons considérer que les conditions d'exercice de la profession n'ont cessé de se dégrader au cours de ces dernières années. Une étude menée en 2018 par l'ULB et l'UGent indiquait que les journalistes belges travaillent, en moyenne, un peu moins de 43 heures par semaine, et qu'ils considèrent que leur charge de travail ne cesse d'augmenter. À cela, s'ajoute l'incertitude liée au statut d'indépendant, qui n'est pas toujours choisi par le journaliste : il concerne, en Belgique francophone, un journaliste sur cinq (Van Leuven *et al.* 2018).

À l'automne 2016, la Fédération Européenne des Journalistes (FEJ) invitait ses organisations membres à participer à une enquête sur les conditions de travail des journalistes. Quarante-neuf responsables de syndicats nationaux de journalistes, issus de 19 pays européens, y avaient répondu. Parmi les facteurs affectant la santé ou la sécurité des journalistes, ils épinglaient la surcharge de travail (82,98%), le stress (76,6%) et le manque de temps (59,57%). Ils étaient 80,85% à estimer que les journalistes sont insuffisamment rémunérés compte tenu de la quantité de travail effectuée; et 100% que leurs conditions de travail se détérioraient au fil des ans⁴⁶. Dans un tel contexte, l'automatisation de la production d'informations peut être une source d'anxiétés professionnelles. Cela étant, l'angoisse que peut générer l'automatisation n'est pas un phénomène neuf dans le monde du travail : les exemples, dans l'histoire des technologies, ne manquent pas (Linden 2017). L'idée que des systèmes d'automatisation puissent prendre la place de l'humain s'est accentuée avec les premiers développements de l'informatique, au lendemain de la Seconde Guerre mondiale : l'automatisation est alors définie comme un "*processus de remplacement des tâches humaines par les fonctions des machines*" (Ekbia & Nardi 2014). Cette assertion ne concernait pas le seul travail manuel, mais aussi les tâches cognitives humaines telles que le traitement des données, la prise de décision et la gestion organisationnelle. Compte tenu des possibilités des systèmes d'automatisation de production de l'information – production à grande échelle, rapidité d'exécution, précision des contenus –, les craintes de remplacer des travailleurs humains par des machines sont compréhensibles, et génèrent

⁴⁶ "Le burn-out des journalistes, symptôme d'un malaise dans les rédactions", European Trade Union Institute, HesaMag 15, consulté le 06/10/2017, URL : <http://www.etui.org/fr/Themes/Sante-et-securite/HesaMag/Journaliste-un-metier-en-voie-de-precarisation/Le-burn-out-des-journalistes-symptome-d-un-malaise-dans-les-redactions>

"des fantasmes dans l'esprit des journalistes qui voient leurs rédactions, dont les revenus se réduisent à une peau de chagrin, tailler dans le personnel" (Antheaume 2016).

Les technologies d'automatisation ont un coût qu'il ne faudrait pas négliger, qu'il s'agisse d'un développement réalisé en interne où des compétences devront être mobilisées pendant quelques semaines; ou d'un développement confié à une société de services. À titre indicatif, selon une enquête réalisée en 2017, à l'échelle mondiale, par la société de recrutement Harvey Nash, le salaire journalier moyen d'un développeur web est de 353 USD, celui d'un développeur de logiciel de 576 USD, celui d'un *project manager* de 633 USD, et celui d'un développeur de bases de données est de 1.033 USD⁴⁷. Chez Syllabs, on indique que le système d'information développé dans le cadre du premier tour des élections régionales françaises, en 2015, avait été facturé au *Monde* "entre 20.000 et 50.0000 euros"⁴⁸. Ces investissements peuvent apparaître rentables, au regard de la rapidité de traitement des données et d'un volume de production avec lequel aucun journaliste ne pourrait rivaliser : un million d'articles avaient été générés en l'espace d'une nuit, lors du premier partenariat entre *Le Monde* et Syllabs. Quatre ans plus tard, force est de constater que ces investissements sont à géométrie variable et dépendent de la taille du projet ou de la solution technologique envisagée. Lors de la Conférence nationale des métiers du journalisme (CNMJ), qui s'est tenue à Paris le 24 janvier 2019, Claude de Loupy, l'un des deux fondateurs de Syllabs, annonçait un premier prix à partir de 7.000 euros par an. La particularité de Syllabs est de fournir des flux d'informations qu'elle va contrôler en amont, plaçant ainsi le client dans une situation de dépendance.

Lorsque l'on examine les tarifs des sociétés développant des logiciels de génération automatique de textes, ceux-ci apparaissent comme plutôt compétitifs. Arria Studio, développé par la société britannique Arria NLG, l'une des pionnières en matière de traitement automatique de la langue, propose une tarification en fonction du volume de textes produits. Le plan tarifaire minimum s'élève à 199 dollars par mois pour 20.000 mots, soit environ 80 feuillets⁴⁹. Le prix d'un abonnement mensuel au logiciel de société allemande AX Semantics⁵⁰, également accessible en ligne en tant que *software as a service*, débute à partir de 279 dollars par mois et peut s'élever jusque 799 dollars mensuels. L'avantage de ces solutions logicielles réside dans le contrôle qu'elles offrent à leurs utilisateurs finaux, qui vont eux-mêmes assurer le paramétrage du logiciel, sans pour autant devoir disposer de compétences particulières. Et elles sont une vingtaine de fois moins coûteuses qu'un rédacteur humain. Selon la société de consultance Gartner, ces solutions logicielles coûteraient de 250 dollars à 4.800 dollars par an⁵¹.

L'emploi freelance serait potentiellement le plus menacé, les journalistes indépendants n'étant pas à l'abri de perdre l'une ou l'autre collaboration en raison de stratégies d'automatisation.

⁴⁷ "Harvey Nash Technology Survey 2018", consulté le 10/12/2017, URL : <http://www.harveynash.com/techsurvey/full-report/charts/>

⁴⁸ Communication personnelle, Claude de Loupy et Helena Blancafort, CEO de Syllabs, Paris, le 29/02/2016.

⁴⁹ Source : <https://store.arria.com>, consulté le 16/06/2019

⁵⁰ Source : <https://www.ax-semantic.com/en/pricing.html>, consulté le 16/06/2019

⁵¹ Market guide for natural language generation, Gartner, 27 juin 2019.

Ce fut le cas en France, en 2015, rapporte un collaborateur régulier du *Monde*. S'inquiétant de ne pas recevoir de nouvelles de la rédaction en chef du journal quelques jours avant le premier tour des élections du régionales, plusieurs collaborateurs se sont enquis de la situation, explique-t-il. Il leur a été répondu que leurs prestations seraient assurées "par un robot". Pour ce journaliste, il faut aussi y voir une politique éditoriale visant "à désinvestir dans les régions". Il indique également que cela n'a pas été sans conséquences : cela a contribué à détricoter un maillage territorial et à briser la confiance entre le correspondant régional et ses sources d'information⁵².

Les journalistes employés dans les agences de presse seraient les moins à l'abri, puisque c'est principalement dans ce secteur que sont développées des technologies d'automatisation (Fanta 2017). Les professionnels dont les tâches sont les plus répétitives pourraient également craindre pour leur emploi, dès lors que celles-ci peuvent être automatisées (Van Dalen 2012). Toutefois, il est n'est pas encore possible de créer une technologie qui remplace la part humaine importante qui caractérise un travail journalistique, et cela d'autant plus qu'aucune profession ne peut être réduite à la somme de ses tâches. Comment automatiser la relation qu'entretient un journaliste avec ses sources d'information, une analyse contextuelle approfondie, des techniques d'enquête, des opinions, la définition de la valeur de l'information, ou encore l'expertise et l'expérience du journaliste? Bien que les technologies d'automatisation de la production d'information soient en mesure d'imiter un journaliste humain, elles ne peuvent pas le remplacer, pas plus qu'elles ne sont capables de se substituer à la complexité du cerveau humain et à sa créativité. À l'instar d'autres technologies relevant du champ de l'intelligence artificielle, elles sont basées sur des modèles rationnels en fonction desquels toute incertitude est rejetée du système (Latar 2018 :24). S'il s'agit d'un jeu de l'imitation, il est donc forcément incomplet.

Il est régulièrement répété par les acteurs du secteur de la génération de textes en langage naturel qu'aucun de ces systèmes n'a eu un impact négatif sur le volume de l'emploi salarié⁵³, mais le risque zéro n'existe pas dans une logique de compression des coûts de production. Ce discours a toutefois pu évoluer dans le temps, Claude de Loupy, CEO de Syllabs, admettant que "*C'est une évidence, l'intelligence artificielle aura un impact sur l'emploi. On ne sait pas exactement quel sera cet impact. Je suis persuadé (...) que de nouveaux équilibres se trouveront, comme on les a trouvés avec chaque arrivée technologique, comme pour la machine à tisser qui a provoqué une catastrophe dans certains bassins d'emploi. (...) Typiquement, lorsque l'on parle des correspondants et des pigistes, ils auront peut-être moins de choses à rédiger*". Il considère également que les correspondants ne fournissent pas un travail qualitatif suffisant, amenant les journalistes à réécrire leur articles "*parce qu'on n'a pas confiance et qu'il faut revoir la façon dont cela a été écrit et qu'il faut le refaire. (...) Cela représente du travail en plus pour les journalistes. Si on demandait aux locaux ou aux correspondants d'entrer de la data, on produirait des textes dans lesquels on aurait totalement confiance parce qu'on vérifierait éventuellement*

⁵² Communication personnelle, Paris, le 21/03/2016.

⁵³ "Non messieurs les journalistes, les robots ne voleront pas votre emploi", Claude de Loupy, L'Obs/Rue 89, 09/05/2014, consulté le 10/07/2018, URL : <https://www.nouvelobs.com/rue89/rue89-tech/20140509.RUE3752/non-messieurs-les-journalistes-les-robots-ne-voleront-pas-vos-emplois.html>

la data, mais au moins, le texte serait bien rédigé et on n'aurait pas à le réécrire. Ce sont donc des métiers qui vont évoluer, avec de nouveaux métiers et des gens qui vont aussi récupérer de la data"⁵⁴.

Vu sous l'angle d'une stratégie qui serait purement économique, un logiciel ne dort pas, ne prend pas de vacances, et son rythme de production est incomparable à celui d'un journaliste. Loin de nous, au Japon, le "robot journaliste" n'est déjà plus une métaphore : *"Je n'ai pas besoin de prendre de pause ou de dormir. Je peux travailler pendant 24 heures. Je ne me plaindrai pas d'un dur travail. (...) Tout ce que vous pouvez faire, je peux le faire mieux. Et plus vite. Et moins cher. Mais je ne suis pas là pour vous remplacer"*, indiquait Erica, un robot androïde "embauché" en avril 2018 par une chaîne de télévision japonaise pour présenter des émissions d'information⁵⁵. En Chine, un robot humanoïde présente le journal télévisé depuis la fin 2018⁵⁶. Mais dans un pays où la presse est placée sous contrôle, ce robot-là apparaît davantage comme un bon petit soldat.

Jusqu'à présent, nous n'avons aucune trace de pertes d'emplois contractuels en raison de l'implémentation de dispositifs d'automatisation dans les rédactions. Mais ceux-ci se développent dans le contexte plus large d'une "quatrième révolution industrielle" qui concerne des professions intellectuelles relativement épargnées par les trois précédentes révolutions. S'il est encore difficile de chiffrer l'impact de l'automatisation sur l'ensemble des professions, plusieurs études prospectives ont mis en avant des conséquences auxquelles il faudrait s'attendre dans le champ journalistique. Il convient de rester prudent face à de telles études : la "futurologie" n'est pas une science, mais elle a pour effet d'alimenter un climat anxigène. De plus, aucune de ces études ne tient compte des politiques pouvant éventuellement être menées pour encadrer ce phénomène, pas plus qu'elles n'envisagent des mouvements de résistance sociale qui seraient susceptibles de le ralentir. En 2014, aux États-Unis à partir d'entretiens mené avec 1.896 experts, le Pew Research Center concluait que les technologies relevant du domaine de l'intelligence artificielle allaient remplacer certains métiers, qu'elles en créeraient de nouveaux, et qu'elles permettraient de redéfinir le travail de manière positive en libérant le travailleur de tâches répétitives. Dans le même temps, indiquait-elle, le système éducatif ne prépare pas de manière adéquate à cette évolution du travail et des modes de production. Seuls les travailleurs hautement qualifiés conserveraient leur emploi. À l'inverse les travailleurs les moins qualifiés seraient les plus susceptibles de perdre le leur⁵⁷. Deux ans plus tard, ce centre publiait une nouvelle étude sur le sujet qui, cette fois, indiquait que les risques de remplacer les humains par des machines étaient bien réels⁵⁸.

⁵⁴ Table ronde "Les journalistes face aux robots", Journée Nationale des Métiers du Journalisme, CNMJ, Verbatim, Paris, La Sorbonne, 24/01/2019.

⁵⁵ "Replacing Humans : Robots Among Us", CBS News, mis à jour le 30/04/2018, consulté le 07/07/2018, URL : <https://www.cbsnews.com/news/robots-replacing-humans-cbsn-originals/>

⁵⁶ "World's first AI news anchor unveiled in China", Lily Kuo, *The Guardian*, 09/11/2018, consulté le 12/11/2018, URL : <https://www.theguardian.com/world/2018/nov/09/worlds-first-ai-news-anchor-unveiled-in-china>

⁵⁷ "A.i., robotics, and the future of jobs", Aaron Anderson et Janna Smith, Pew Research Center, 06/08/2014, consulté le 13/10/2017, URL : <http://www.pewinternet.org/2014/08/06/future-of-jobs/>

⁵⁸ "Public predictions for the future of workforce automation", Aaron Smith, Pew Research Center, 10/03/2016, consulté le 13/10/2017, URL :

En 2015, la branche belge de la banque ING rendait publique une étude traitant de l'impact de la robotisation sur différents métiers. Sa méthodologie s'appuyait sur un calcul de probabilités basé sur les avancées scientifiques et technologiques dans les domaines de l'apprentissage des tâches par les machines et les robots mobiles, mis au point par l'économiste C.B. Frey et l'ingénieur M.À. Osborne. Elle proposait une liste de 426 métiers potentiellement menacés⁵⁹. Deux facteurs étaient examinés : l'esprit de créativité dans l'exécution des tâches, et celle de la gestion des liens sociaux. Si 35% des métiers présentaient une probabilité plus élevée d'être robotisés ou automatisés, l'emploi journalistique serait épargné avec une probabilité de 8,25%⁶⁰. En 2016, l'International Data Corporation (IDC), une société de consultance spécialisée dans les marchés des technologies de l'information, des télécommunications et des technologies grand public, active dans plus de 110 pays, publiait un rapport évaluant l'impact de l'automatisation sur l'emploi, à l'horizon 2020. Sept professions avaient été identifiées comme présentant un risque élevé de disparition d'emplois : les camionneurs, les métiers de la construction, les juristes, les médecins et équipes médicales, les comptables, les vendeurs et les éditeurs. À propos de ces derniers, cette étude estimait que *"les romanciers ont encore un avenir, mais les rédacteurs financiers ou ceux qui écrivent toutes sortes de rapports ont un avenir limité"*. Pour appuyer ses résultats, elle indiquait que les systèmes de génération automatique de textes sont capables d'écrire des textes factuels de plus en plus lisibles⁶¹.

La société suédoise de télécommunications Ericsson s'est, quant à elle, spécifiquement penchée sur l'impact de l'automatisation des tâches sur le métier de journaliste. Son rapport, publié en 2017, comportait dix prédictions concernant l'usage des TIC dans les rédactions. L'une d'entre elles concernait l'emploi : pour l'auteur du rapport, de plus en plus de journalistes seront remplacés par des logiciels. Ce rapport soulignait également que *"Il serait naïf de croire que les entreprises de presse n'utiliseront pas non plus la technologie pour réduire les coûts en réduisant le personnel. En ce sens, l'intelligence artificielle est à la fois une opportunité pour le journalisme et une menace pour les journalistes"*⁶².

Selon un rapport publié la même année par la société de consultance McKinsey, 15% du travail d'un journaliste et 9% du travail d'un éditeur pourraient être remplacés par des machines. Résultant d'une enquête menée dans 46 pays, ce rapport estimait jusqu'à 800 millions d'emplois humains menacés à l'horizon 2030, soit 12% de la population active. Les secteurs qui seraient épargnés par ce phénomène seraient ceux impliquant une action humaine, comme en médecine ou dans l'enseignement. Les pays les plus touchés seraient ceux qui disposeront

⁵⁹ <http://www.pewinternet.org/2016/03/10/public-predictions-for-the-future-of-workforce-automation/>
La liste a été établie à partir de celle de l'International Standard Classification of Occupation (ISCO), éditée par le Bureau International du Travail (BIT).

⁶⁰ "ING, focus sur l'emploi. La révolution technologique en Belgique", document par le service de presse ING par courriel, le 17/02/2015.

⁶¹ "The 7 jobs that will be first killed by robots", Pierre Pina, IPFC Online, 01/05/2017, consulté le 13/10/2017, URL : <http://ipfconline.fr/blog/2017/05/01/the-7-jobs-that-will-be-killed-first-by-robots/>

⁶² "The future of journalism in a networked society", Mikael Törnwall, Ericsson, septembre 2017, consulté le 13/09/2017, URL : <https://www.ericsson.com/en/networked-society/trends-and-insights/networked-societyinsights/future-of-journalism>

des moyens pour investir dans des technologies coûteuses⁶³. Pour Diakopoulos (2019 :35), les technologies relevant du domaine de l'intelligence artificielle ne vont pas nécessairement supprimer l'emploi journalistique mais bien le transformer. Les rôles et tâches des journalistes sont amenés à évoluer, et le travail humain sera "hybridé", c'est à dire mélangé avec des algorithmes, pour s'adapter aux capacités et aux limites des technologies.

1.3.2 Trouver l'équilibre : les promesses du partenariat homme-machine

Avec l'apparition et le développement du phénomène des contenus générés par les utilisateurs (UGC), les journalistes avaient perdu un monopole qui leur était jusque-là dévolu, n'étant plus les seuls à "montrer" et à "raconter le monde" (Deuze 2004, Domingo & Le Cam 2014). Le développement de la production automatisée de contenus journalistiques amène, une nouvelle fois, la remise en question de l'autorité des journalistes (Hammond 2017), ces derniers devant admettre qu'une partie de leurs activités n'est plus leur exclusive. Symptôme de l'émiettement du journalisme (Ruellan 2005), ce phénomène repousse les limites de ce qu'est le journalisme (Lewis 2012). Il place les professionnels de l'information face à leurs compétences et à ce qui caractérise leur identité professionnelle (Neveu 2010, Van Dalen 2014). Il pose également les questions de la relation entre le journaliste et l'artefact d'automatisation, et de celle qu'il entretient avec les agents sociaux qui développent ces technologies. Dès lors, comment réussir à tisser des liens harmonieux tout en maintenant ses spécificités? Bien que les caractéristiques de base du journalisme soient restées les mêmes au cours de ces dernières décennies, les tâches des journalistes et les compétences qu'ils mobilisent ont considérablement changé. Le journalisme est devenu plus technique, et le journaliste a vu glisser sa fonction de producteur d'informations à celle de "cueilleur" ou de "gestionnaire" d'informations (Bakker 2014).

Powers (2012) constate que les changements technologiques relatifs aux modes de production de l'information n'ont pas seulement eu une incidence sur les pratiques professionnelles, elles ont aussi permis de créer de nouvelles formes de travail. Par exemple, l'édition informatisée a permis que les activités de conception et de mise en page se déroulent au sein des rédactions. Cette évolution n'a pas été sans générer des tensions, essentiellement liées au contrôle éditorial et aux contenus. Les pratiques de journalisme assisté par ordinateur illustrent comment les technologies ont contribué à créer "*les conditions nécessaires pour l'émergence d'un travail technologique spécifique*". Ces facteurs, indique Powers, "*contribuent à la construction d'un discours sur le rôle que ce travail devrait jouer dans le journalisme*". Cependant, "*la technologie n'est pas un facteur indépendant qui influence le travail des journalistes de l'extérieur. Il faut plutôt la voir en termes de mise en œuvre, et donc comment elle étend et amplifie des façons de faire antérieures*" (Deuze 2008).

⁶³ "Jobs lost, jobs gained : What the future of work will mean for jobs, skills, and wages", consulté le 03/07/2018, URL : <https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages>, "Artificial intelligence-enhanced journalism offers a glimpse of the future of the knowledge economy", Nicholas Diakopoulos, The Conversation, 11/06/2019, consulté le 03/07/2019, URL : <https://theconversation.com/artificial-intelligence-enhanced-journalism-offers-a-glimpse-of-the-future-of-the-knowledge-economy-117728>

Linden (2017) estime que les journalistes ont des raisons légitimes de critiquer les technologies d'automatisation vendues aux rédactions comme un service ou un logiciel "*plug-and-play*". Dans ce type de projet, lorsqu'ils y ont été associés, le rôle des journalistes a souvent été limité à "former" des algorithmes pour choisir le phrasé le plus approprié; et ceux-ci n'ont pas considéré ce type d'expérience comme particulièrement agréable. La manière dont les journalistes réagissent à l'introduction d'artefacts de production automatisée d'informations dans leurs rédactions a été explorée dans plusieurs travaux de recherche⁶⁴ : généralement, il leur est reconnu le bénéfice potentiel d'un gain de temps, par la prise en charge de tâches répétitives et chronophages. Dans une étude exploratoire portant sur la manière dont la logique du journalisme a été traduite en logiciels et celle dont les experts de l'industrie anticipent l'avenir en matière d'automatisation dans les rédactions, Linden (2017) a démontré que les journalistes peuvent faire preuve d'une "*forte capacité d'adaptation*", alors que le débat sur l'automatisation de la production d'informations porte principalement sur les pertes d'emplois.

Dans une autre étude reposant sur des entretiens avec 24 experts du journalisme automatisé – journalistes, éditeurs, gestionnaires, représentants de sociétés de génération automatique de textes et observateurs du phénomène – ce même chercheur (Linden 2017b) souhaitait comprendre comment les gestionnaires, journalistes et programmeurs travaillaient avec des logiciels d'automatisation de la production d'informations, ainsi que la manière dont cette automatisation pouvait soutenir les journalistes dans leur travail tout en les libérant de tâches répétitives. Il s'est adossé au cadre proposé par Powers (2012) pour aborder la manière dont les journalistes évaluent l'impact de la technologie dans leur travail, qui consiste en trois variables : celui de la continuité, celui de la menace et celui d'une possibilité de réintégration journalistique. Il conclut que chacun de ces discours se retrouvait dans ce contexte.

Si l'on se tourne du côté du management, le principal défi consiste essentiellement à démythifier une technologie que les journalistes craignent (Milosavljević & Vobič 2019). En Finlande, les responsables chargés du développement de systèmes d'automatisation dans les rédactions l'envisagent comme un soutien à des routines répétitives mais cela suppose un accès à des données fiables, ainsi que des compétences et des ressources pour utiliser et comprendre les données. Cela suppose aussi que la qualité des textes générés soit suffisante (Sirén-Heikel *et al.* 2019).

En automatisant certaines tâches éditoriales, l'agence de presse américaine Associated Press s'inscrit dans la perspective de la continuité de ses activités. L'introduction d'un logiciel de génération automatique de textes, en 2014, y a d'ailleurs fait suite à une demande formulée par des dirigeants du département de l'information de l'agence (Marconi *et al.* 2017) : "*Alors, oui, vous avez bien lu. Les journalistes d'AP étaient ceux qui ont suggéré d'abord de céder du travail à un programme d'ordinateur – aux robots, si vous préférez. Ironique, à première vue, cette décision a été cruciale pour l'agence de presse en luttant contre deux méga-tendances dans l'entreprise : l'augmentation incessante des informations à couvrir et les contraintes humaines*

⁶⁴ Voir dans le chapitre 2, "Dimension évaluative", p.145

associées à cette couverture". Il s'agit là d'un argument régulièrement mis en avant pour justifier l'introduction de systèmes d'automatisation : les tâches qu'ils effectuent sont répétitives et chronophages pour un journaliste humain (Van Dalen 2012, Latar 2015, Graefe 2016, Linden 2017). Mais cela suppose que les journalistes "*trouvent un endroit ou autre chose qui va les payer pour faire un travail vraiment utile que les machines ne peuvent pas faire*" (Carlson 2014).

L'extension des zones de couverture médiatique est un autre argument mis en avant pour justifier l'implémentation d'un système d'automatisation dans une rédaction : le logiciel ne remplacerait donc pas le journaliste mais permettrait de couvrir des événements trop modestes pour justifier l'envoi d'un journaliste sur le terrain. C'est ce qui a motivé l'agence de presse norvégienne NTB à développer un système automatisé qui traite une variété d'informations quantifiables (prix des logements, résultats financiers d'entreprises et résultats sportifs), pour un volume mensuel de 20.000 articles générés (Fanta 2017). L'algorithme de génération automatique rédige un texte en fonction d'un ensemble de règles et de modèles prédéfinis, qui vont varier en fonction de la valeur des données. Il s'agit d'un texte standardisé, qui s'appuie sur une structure "titre-texte" de description des événements. C'est donc un "*squelette de base sur le modèle duquel tous les articles sont construits (...) Pour pouvoir générer des phrases en langage naturel, le robot doit disposer d'un ensemble de mots et d'expressions pouvant être pondérés et sélectionnés en fonction d'un large éventail de critères*"⁶⁵. Les journalistes sportifs de NTB ont été associés à la conception du projet. Leur travail a consisté à créer un ensemble de modèles que le système peut utiliser "*pour rendre la langue aussi naturelle que possible et lui fournir le vocabulaire approprié*"⁶⁶. Chez ProPublica, *pure player* américain spécialisé dans l'investigation, l'automatisation a été envisagée comme une manière de créer de nouvelles formes d'emploi et, partant, de participer au renouveau du journalisme (Linden 2017b). Tandis que chez Local Labs, aux États-Unis, certaines tâches traditionnellement dévolues aux journalistes ont été automatisées dans une logique de compression des coûts. L'entreprise emploie toujours des journalistes. Dans certains cas, ils sont amenés à collaborer avec les machines (Linden 2017b).

Pour trouver l'équilibre dans la relation homme-machine, il convient de bien comprendre en quoi consiste un dispositif technique de production automatisée de contenus journalistiques. En premier lieu, deux types de démarche sont à distinguer : d'une part, des logiciels de génération automatique de textes en langage naturel; et d'autre part, des applications prenant la forme d'interfaces web ou mobiles pilotées par des données, dont le *storytelling* peut prendre des formes variées. En second lieu, les systèmes d'automatisation de la production d'informations peuvent relever de deux logiques différentes. La première s'appuie sur les technologies du *machine learning* (apprentissage par la machine, relevant du domaine de l'intelligence artificielle dans ses développements actuels⁶⁷), comme c'est le cas dans les sociétés américaines

⁶⁵ "Building a robot journalist", Espen Waldal, Medium.com, 18/11/2016, consulté le 20/11/2016, URL :<https://medium.com/bakken-b%C3%A6ck/building-a-robot-journalist-171554a68fa8>

⁶⁶ "Building a robot journalist", Espen Waldal, idem.

⁶⁷ À ce propos, indique Lévy (2017), "*les ordinateurs n'apprennent pas de manière autonome. Les programmes de la nouvelle intelligence artificielle sont produits par un couple. D'un côté, un algorithme d'apprentissage simule un système complexe de type neuronal (...). D'un autre côté, un ensemble de données entraîne cet algorithme. Le programme d'intelligence artificielle n'arrive à traiter automatiquement l'information qu'après*

Narrative Science et Automated Insights; et des systèmes à base de règles, qui imitent une forme rudimentaire d'intelligence dans un contexte limité, comme c'est le cas dans la start-up française Syllabs. Ces deux logiques ont pour point commun d'être adaptées au domaine d'application auquel ils se rapportent : on n'écrit pas les résultats d'une rencontre sportive comme on écrirait ceux d'une entreprise. Pour Syllabs, cela signifie qu'à chaque nouveau projet, son moteur de rédaction doit faire l'objet de nouveaux paramètres. Lors du second tour des élections régionales françaises, en décembre 2015, Syllabs travaillait pour quatre médias différents : elle a donc dû mettre au point quatre robots différents⁶⁸. Un responsable de Mittmedia, un groupe de presse suédois éditant 19 quotidiens, souligne également qu'il est difficile de créer une solution générique : chaque nouveau projet d'automatisation nécessite de recommencer à zéro ("*start form scratch*") et chacun des projets développés est relativement peu réutilisable (Linden 2017).

Toujours dans le contexte des médias d'information, la start-up allemande AEXEA indique travailler avec des textes aux structures répétitives, tels que les résultats de matches de football et des bulletins météorologiques. "*Les données nous sont fournies par des tiers avec lesquels nos clients ont généralement signé une licence d'utilisation. Par ailleurs, nous leur offrons la possibilité d'entraîner leurs propres textes sur notre système*", explique Frank Feulner, *Chief business development officer* chez AEXEA⁶⁹. Une autre particularité est que le logiciel développé par la compagnie travaille simultanément avec des grammaires d'une trentaine de langages différents, la machine étant à chaque fois entraînée pour produire des textes originaux (et non pour fournir des traductions). "*En interne, des copywriters et des journalistes entraînent la machine pour nos clients. Ce processus d'entraînement prend environ deux semaines. Dans tous les cas, nous avons besoin de la connaissance du domaine d'application. Cette expertise nous est fournie par nos clients. Notre système est hybride, même s'il est essentiellement basé sur des règles. Nous avons également mis au point un système permettant d'évaluer la fiabilité et la qualité des données. Tout dépend toujours de leur provenance mais nous avons besoin de données structurées de grande qualité. Toutefois, obtenir des données ne pose pas de grand problème, dans la mesure où celles-ci sont aujourd'hui disponibles en grandes quantités*".

En juin 2019, la société de consultance Gartner publiait son premier rapport sur le marché des logiciels spécialisés en génération automatique de textes en langage naturel. Il soulignait que la logique basée sur une approche par template domine toujours, malgré les promesses des techniques de l'intelligence artificielle⁷⁰. Robbie Allen, fondateur de la société américaine Automated Insights, estime qu'un processus de rédaction automatique n'a pas fondamentalement changé en vingt ans. "*La manière de faire est pratiquement la même, mais la principale innovation consiste dans la diffusion de contenus via les réseaux sociaux. Cet espace entraîne qu'au*

avoir été entraîné sur des dizaines de milliers d'exemples de problèmes accompagnés de la 'bonne réponse' ou de données correctement catégorisées".

⁶⁸ Communication personnelle, Claude de Loupy et Helena Blancafort, CEO de Syllabs, Paris, le 29/02/2016.

⁶⁹ Communication personnelle, Stuttgart, le 12/03/2016.

⁷⁰ Voir annexe 7 "Différenciation des solutions de génération automatique en langue naturelle et directions du marché", p.331

lieu d'écrire une seule histoire, dix histoires, des centaines d'histoires peuvent être écrites"⁷¹. Un processus théorique de génération automatique de textes (Figure 1.2) utilise "les connaissances du langage et d'un domaine d'application pour produire automatiquement des rapports, des documents, (...) des messages d'aide ou tout autre type de texte" (Reiter & Dale 1997). Son architecture est modulaire et son processus théorique s'articule en deux questions : "Quoi dire?" et "Comment le dire?" (Danlos 1991). La première question comprend l'étape de planification du document (macro-planification, dont la fonction est de linéariser la forme logique du texte) et celle de la micro-planification (s'appuyant sur des connaissances linguistiques permettant de former les phrases). Cette deuxième étape a pour objectif d'élaborer le sens du texte en fonction des ressources disponibles et de leur analyse. Le "Comment le dire?" correspond à la génération de surface, étape au cours de laquelle vont intervenir des décisions d'ordre linguistique.

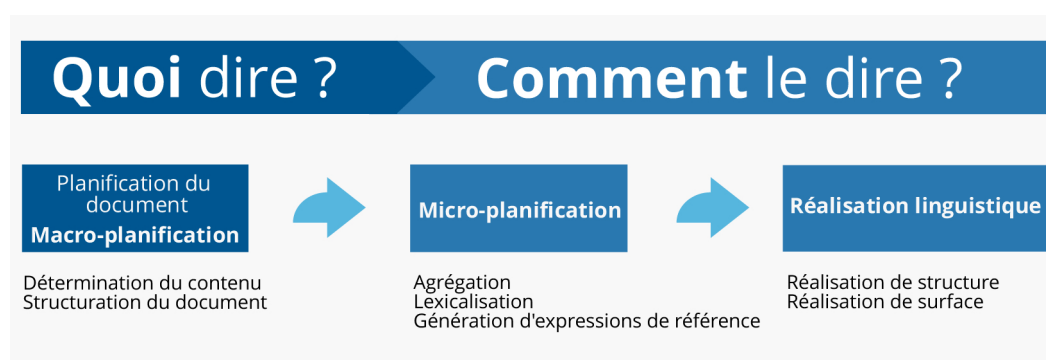


FIGURE 1.2 – Processus théorique "en pipeline" de génération automatique de textes, d'après Danlos (1991) et Reiter & Dale (2000)

Les étapes de cette architecture en pipeline n'interviennent toutefois pas nécessairement de façon linéaire et de façon aussi distincte, les processus se trouvant souvent en interaction les uns avec les autres (Ponton 1997). Tout au long de ce processus, différents niveaux de représentation de la langue vont se succéder : pragmatique, sémantique, syntaxique, morphologique et phonétique (Dale 1995). Utilisée dans un contexte journalistique, cette approche s'appuie sur un modèle *data-to-text* qui permettra de générer un texte de manière automatique (*output*) à partir de données structurées (*input*).

À l'instar d'un processus éditorial traditionnel, où se posent aussi ces deux questions, les procédures informatiques résultent d'une succession de choix qui demeurent le plus souvent opaques pour leurs utilisateurs. Là aussi, il s'agit de formaliser un ensemble de règles, de routines et de procédures institutionnalisées opérant au sein d'un cadre social déterminé et qui sont soutenus par une expertise professionnelle (Lewis & Westlund 2015). C'est pourquoi le code informatique remplit une fonction éditoriale à part entière (Weber & Kosterich 2018), de même qu'un algorithme d'automatisation de l'information comporte des jugements humains qui ne peuvent, par leur nature, être considérés comme "neutres" (Gillespie 2014). Dans cette perspective, la production automatisée d'informations journalistiques ne serait rien d'autre qu'une

⁷¹ Recueilli lors d'un webinaire organisé par Automated Insights, le 27/01/2017.

forme remixée d'un processus éditorial ou médiatique existant (Manovich 2013). Elle va dépendre de la logique du média dans lequel elle se déploie, que ce soit en termes de règles ou de routines spécifiques (Graefe 2016). Étant entendu qu'un algorithme consiste en une suite de procédures destinées à solutionner un problème, Linden (2017) considère que le journalisme peut, lui aussi, être envisagé comme un algorithme : sa pratique consiste en différentes étapes dont chacune est pilotée par des choix (sources, angle, récit...) en vue de résoudre un problème (raconter une actualité). Cet algorithme pourrait être rédigé en pseudo-code, par exemple : *"Trouvez une nouvelle information qui adhère à un ensemble fixe de règles telles que la curiosité humaine, la pertinence et l'impact; contactez trois personnes reconnues publiquement comme des sources indépendantes la commenter (l'information, ndlr) et la contextualiser; rédigez un article de 200 mots en fonction des règles éditoriales de la rédaction."* Ces instructions de base, souligne Linden (2017), pourraient très bien être introduites dans un système informatique, à partir du moment où chacune des étapes décrites peut être automatisée.

Bien que journalisme et technologie se rejoignent dès lors que tous deux sont abordés sous l'angle fonctionnaliste d'une activité occupationnelle balisée par une série de routines et de choix, leurs chemins se séparent lorsque l'on s'intéresse à leurs aspects culturels et sociaux. Ils portent chacun des idées différentes à propos des concepts d'objectivité, de transparence et de partage d'informations. Dans le monde du journalisme, ces concepts participent à un idéal de la profession et sont souvent codifiés (Zamith 2019). L'objectivité légitimise le discours du journaliste, et elle implique la collecte et la diffusion de faits vérifiables. Toutefois, l'information journalistique ne peut être détachée de son contexte de production, pas plus qu'elle ne peut l'être des référents socioculturels et professionnels des agents sociaux qui la produisent (Carlson 2019).

En raison de leurs normes distinctes, journalisme et technologie induisent des vues différentes sur la nature du journalisme et de ses processus (Lewis & Usher 2014). Les passerelles existent et sont nécessaires, mais celles-ci doivent s'opérer dans le cadre de cultures différentes : *"les journalistes aiment rendre les choses simples, alors que les développeurs préfèrent qu'on leur dise à quel point le sujet est complexe"* (Antheaume 2013). Si informaticiens et journalistes peuvent éprouver des difficultés à développer une langue commune (Lewis & Usher 2014), leur dépendance semble plus forte que les différences dans leurs cultures de travail (Karlsen & Stavelin 2014). Sur le plan opérationnel, les interactions entre développeurs et journalistes présentent souvent un caractère limité dans le temps et ne conduisent pas à des collaborations formelles à long terme dans le développement de nouveaux produits (Bakker 2014, cité par Weber & Kostersch 2018).

Les journalistes composent de longue date avec les technologies informatiques, et elles sont devenues des conditions matérielles de leur travail (McCosker & Milne 2014). Journalismes et technologies informatiques ont été amenés à davantage se rencontrer au cours des années 2000 pour deux raisons principales : un intérêt pour une approche journalistique s'appuyant sur des données numériques, et le développement de récits interactifs publiés en ligne nécessitant

des connaissances techniques. Dans les deux cas, des compétences spécifiques sont mobilisées (Lewis & Usher 2014). Il en va de même dans tout type d'approche par données dans le journalisme, exigeant en temps et en compétences, où certains projets peuvent mobiliser des journalistes, des chercheurs, des statisticiens, des concepteurs numériques, des analystes de données, des spécialistes de la visualisation de données et/ou des développeurs web (Henninger 2013). Cette approche transversale du travail d'équipe permet d'intégrer les compétences informatiques dans les processus de travail des rédactions (Gynnild 2010).

Dans le contexte de la production automatisée de contenus, les promesses d'un partenariat gagnant-gagnant peuvent être comprises comme un soutien aux routines journalistiques, les technologies permettant de brasser rapidement de larges volumes de données dans le cadre d'une démarche d'investigation qui prend du temps et coûte de l'argent (Daniel *et al.* 2010). Si certaines tâches peuvent faire l'objet d'une automatisation, elles impliquent donc un savoir-faire dans les techniques d'investigation et de relations aux sources d'information qui ne peuvent être automatisées, même partiellement. Pour autant, envisager la relation homme-machine comme une compétition où seule compterait la performance est un leurre : cette course est perdue d'avance (Brynjolfsson & McAfee 2012). De plus, si des tâches répétitives et chronophages sont prises en charge par une machine, les journalistes n'auraient pas à développer de nouvelles compétences mais à renforcer leur savoir-faire, en se concentrant sur des tâches plus valorisantes. Cela étant, les performances de ces systèmes renvoient invariablement les professionnels à la question de leurs compétences, celles-là mêmes qui seraient susceptibles de faire la différence.

Par ailleurs, le développement des technologies d'automatisation laisse entrevoir un avenir où des systèmes informatiques basés sur des technologies de *machine learning* – qui consiste à créer des programmes informatiques qui "se construisent" via des algorithmes d'apprentissage – n'échapperont plus seulement au contrôle des journalistes, mais aussi à celui de leurs programmeurs car ils sont "*plus difficiles (parfois impossibles) à comprendre, à déboguer, et plus difficiles à contrôler. Pourtant, c'est précisément pour ces raisons qu'ils offrent le potentiel d'un comportement beaucoup plus 'intelligent'*" (Auebarch 2015, cité par Linden 2017).

1.3.3 Comprendre le fonctionnement de la machine

Historiquement, les rédactions ont été souvent un terrain d'expérimentation pour les innovations technologiques mais elles ont aussi été un champ de bataille sur lequel se sont affrontés intérêts économiques et sociaux (Hardt 1990). Lorsqu'elles ont été développées dans le cadre de stratégies organisationnelles, les journalistes en ont souvent été tenus à l'écart (Hardt 1990). Pour autant, la dynamique du changement des compétences au sein d'une profession est complexe, dès lors qu'elle intègre également une dimension liée à des stratégies organisationnelles (Örnebring 2000). En tant que travailleurs, les journalistes ont composé avec ces évolutions auxquelles ils ont dû s'adapter, dans la mesure où les technologies ont modifié la nature des rôles, des tâches et des flux de travail (Diakopoulos 2019 :5). Avec l'automatisation de la production d'informations, de nouvelles pratiques professionnelles ont ainsi émergé,

avec la figure du "méta-journaliste", dont le rôle consiste à configurer et paramétrer un logiciel de génération automatique, ainsi qu'à définir des gabarits de textes. Il s'agit ici de penser en termes de procédures en développant une forme de pensée computationnelle, de manière à comprendre ce nouvel espace de conception du récit médiatique (Diakopoulos 2019 :5-32). Dans le même temps, d'autres nouvelles fonctions ont commencé à apparaître dans les rédactions américaines, en vue d'assurer la maintenance des productions automatisées, que ce soit sur le plan des données ou sur celui du domaine d'application (Diakopoulos 2019 :5-33).

Le principe de division entre tâches techniques et intellectuelles apparaît aujourd'hui moins clairement, les journalistes posant de plus en plus de gestes techniques qui étaient, auparavant, le fait de techniciens spécialisés (Nygren 2014). Cette évolution des pratiques n'est pas sans incidences sur les professionnels : elle leur laisse moins de temps pour satisfaire leurs pratiques traditionnelles, elle contribue à augmenter la pression sur leur travail quotidien, et elle suscite des inquiétudes quant à la qualité de leurs productions. De plus, les outils d'aujourd'hui ne sont pas ceux d'hier – bandes magnétiques et pellicule sont des artefacts appartenant bel et bien au passé – et toute une génération de journalistes n'a pas été formée à ces nouvelles pratiques. Pour Bakker (2014), *"le journalisme est plus que jamais en train de devenir un travail technique. Ce qui frappe dans les qualifications techniques requises des nouveaux employés est la diversité et la profondeur des compétences demandées. Il est toutefois impossible pour un journaliste de tout maîtriser, encore moins pour une école de journalisme d'enseigner toutes ces compétences en profondeur (...) Mais avoir des compétences techniques augmenterait probablement les chances des étudiants en journalisme de trouver un emploi"*.

Bien que de nombreux gestes techniques apparaissent désormais comme une norme, ceux qui relèvent d'une maîtrise des langages informatiques sont moins courants. Une enquête réalisée en 2017 par l'International Center For Journalists (ICFJ), auprès de plus 2.000 journalistes de 130 pays, témoigne de ce décalage entre journalistes et technologies informatiques. Jusqu'à 5% des rédactions employaient à l'époque du personnel diplômé en technologie, avec une moyenne globale de 2%, et seulement 9% d'entre eux étaient diplômés en journalisme ou communication. Dans 82% des rédactions, 18% de l'emploi concernait des postes dédiés au numérique (médias sociaux, édition et production de contenus numériques, éditeurs analytiques); et moins d'un tiers des répondants affirmait pratiquer le journalisme de données. Cette enquête soulignait que l'offre en formations ne répond pas toujours aux demandes des journalistes (52%), alors que seulement 40% des rédactions leur en proposent ⁷². Une autre enquête menée la même année par le Google News Lab, auprès de 900 datajournalistes, indiquait que 53% des journalistes estiment que le nettoyage, le traitement et l'analyse de données nécessitent une formation approfondie (Rogers *et al.* 2017).

Tim Berners-Lee, principal inventeur du *World Wide Web*, fait partie des plus ardents défenseurs du journalisme de données (Stray *et al.* 2013). Il estime que ceux qui connaissent les

⁷² "First-ever Global Survey of News Tech Reveals Perilous Digital Skills Gap", ICFJ, 05/10/2017, consulté le 16/10/2017, URL : <http://www.icfj.org/news/first-ever-global-survey-news-tech-reveals-perilous-digital-skills-gap>

formats CSV et RDF (formats de données structurées), et qui peuvent lancer rapidement des requêtes en MySQL (langage de requête pour les bases de données en ligne) pour une sortie en langages de programmation PHP ou Python, auront le plus de chances de découvrir les meilleures informations⁷³. Pour autant, ces compétences ne vont pas nécessairement de soi dans les pratiques journalistiques. Dans une recherche s'appuyant sur l'examen de cent programmes de journalisme aux États-Unis et sur plus de cinquante entretiens avec des journalistes et enseignants, Berret & Philips (2016) constatent que près de la moitié de ces programmes ne comportent aucun cours de journalisme de données, que les cours dispensés sont surtout introductifs (utilisation d'une feuille de calcul, compréhension des statistiques descriptives, nettoyage de données..), qu'il manque de littérature pour l'enseignement de la pratique du journalisme de données, que les membres du corps professoral ne sont pas outillés dans ces matières et doivent donc faire appel à des intervenants extérieurs, et que les diplômés ayant suivi un cursus "data" sont mieux armés pour réussir leur vie professionnelle. De plus, malgré une offre de formations spécialisées au journalisme de données, les étudiants ne s'y inscrivent pas forcément (Schmitz-Weiss *et al.* 2018). La principale raison serait à trouver dans un manque d'appétence pour les chiffres, les mathématiques et les statistiques.

Sur le terrain, observe Bradshaw (2018), les employeurs qui cherchent à embaucher des journalistes de données se heurtent à deux écueils : des candidats qui disposent de compétences techniques mais ayant peu d'idées quant à la manière de repérer et de raconter des histoires ; et des candidats qui ont un bon sens de l'information mais qui manquent de capacités pour les réaliser techniquement. Ceux-ci s'inscrivent dans une "*tendance à 'bluffer sur leurs prouesses techniques'*". Aussi, estime-t-il, le challenge de l'enseignement consiste à développer des compétences à la fois éditoriales et techniques. L'objet n'est pas de faire des journalistes de données des programmeurs mais plutôt de les familiariser avec une variété de langages informatiques utilisés dans les rédactions. Dans les médias anglo-saxons, les exigences sont pointues dès lors qu'il s'agit d'embaucher un datajournaliste⁷⁴. Ces exigences s'expliquent notamment par le fait que, en raison de la mise en disponibilité d'importants volumes de données, les journalistes ont besoin de compétences analytiques, en mathématiques et en matière de résolution de problèmes dans un environnement technologique (Linden 2017b). Les journalistes eux-mêmes, dès lors qu'ils pratiquent le journalisme de données, considèrent que les compétences en informatique sont essentielles, et que le manque de qualification en interne a pour effet de se tourner vers des collaborations externes.

Au Royaume Uni, dans une recherche s'appuyant sur vingt-quatre entretiens avec des datajournalistes, des éditeurs de données et des directeurs des médias d'information générale, Borges-Rey (2016) a indiqué que plus de la moitié des répondants estimaient que les connaissances techniques relatives aux manipulations de données devraient être une compétence fondamentale. Toutefois, plusieurs répondants ont dit se sentir limités par leur incapacité à écrire du code

⁷³ "Analysing data is the future for journalists, says Tim Berners-Lee", Charles Arthur, *The Guardian*, 22/11/2010, consulté le 12/10/2016, URL :

<https://www.theguardian.com/media/2010/nov/22/data-analysis-tim-berners-lee>

⁷⁴ Voir annexe 1, "Offre d'emploi pour un poste de datajournaliste au Guardian", p.322

informatique. Reconnus par leurs collègues comme exerçant une forme sérieuse de journalisme, les journalistes de données luttent contre des technologies en constante évolution. En externe, cette dynamique se traduit dans un "combat" avec des plateformes technologiques auxquelles ils doivent s'adapter et dans la mise en place d'un cadre juridique entourant les modèles d'affaires de ces plateformes. La limite de ces pratiques est celle de l'accessibilité des données, qui contraint les journalistes à faire usage de méthodes traditionnelles de collecte et d'enquête.

Face à la difficulté de recruter des journalistes disposant de compétences technologiques, certaines rédactions se sont tournées vers le recrutement de spécialistes disposant d'une compréhension plus pointue des données et de leur analyse, bien que "*le rôle du journaliste reste important*" (Loosen 2018). De nouvelles formes de travail collaboratif sont alors apparues au sein d'équipes multidisciplinaires (Sandoval-Martín & La-Rosa 2018). À l'Associated Press, où l'on travaille avec un logiciel de génération automatique de textes depuis 2013 pour la couverture d'informations financières et sportives, on souligne que les journalistes qui travaillent bien avec les scientifiques des données et les journalistes computationnels sont les mieux placés pour évoluer dans une rédaction assistée par l'intelligence artificielle. L'option y est de "*faire rentrer la science dans les rédactions, parce que les normes de la bonne science – la transparence et la reproductibilité – s'intègrent parfaitement dans le journalisme (...) Les journalistes 'computationnels' comprennent comment l'intelligence artificielle fonctionne et comment elle peut être utilisée pour augmenter leurs pratiques journalistiques*" (Marconi et al. 2017). Pour autant, rares sont les rédactions disposant de ressources et compétences nécessaires pour développer des solutions en interne. "*Les développeurs dans les rédactions sont un peu comme l'eau dans le désert : rares et très recherchés*", affirmaient Gray et al. en 2013.

Quant à ces professionnels hybrides, à mi-chemin entre journalisme et informatique, ils ne représentent encore qu'une minorité au sein de la profession. Y compris aux États-Unis, un pays considéré comme pionnier en la matière. Une seule expérience de production automatisée de l'information est attribuée à un profil de ce type : celle de "Quakebot", un programme informatique développé en mars 2014 par Ken Schwencke pour le *Los Angeles Times* qui alerte, en temps réel, lorsqu'un tremblement de terre se produit en Californie. Cette hybridation des pratiques n'est pas sans poser des questions identitaires (Dagiral & Parasie 2011). Issus du monde de l'informatique, ces hybrides se caractérisent par une forme de militantisme pour le logiciel libre, et ils sont souvent attachés aux valeurs de la culture "hacker" qui "*valorise surtout la liberté de l'information et de la connaissance et l'accessibilité universelle à la technologie*" (Parasie 2011). En 2010, l'Américain Aron Pilhofer s'interrogeait sur la manière adéquate de les catégoriser : "*Nous ne sommes pas juste des ingénieurs (...) À un moment donné je pensais que nous serions peut-être des journalistes (...) De toutes les options que j'ai envisagées, 'développeur d'applications d'informations' est probablement (le plus approprié)*"⁷⁵.

⁷⁵ "Programmer-Journalist? Hacker-Journalist? Our Identity Crisis", Aron Pilhofer, Mediashift, 22/04/2010, consulté le 14/10/2017, URL : <http://mediashift.org/2010/04/programmer-journalist-hacker-journalist-our-identity-crisis107/>

Les journalistes doivent-ils apprendre à programmer? La question, régulièrement posée, ne rencontre pas de réponse tranchée. Mais comment ne pas coder dès lors que l'on souhaite maîtriser son interface pour prototyper ses propres projets éditoriaux, et comment ne pas programmer dès lors que l'on souhaite explorer des jeux de données de manière approfondie? Les techniques d'extraction et de nettoyage de données, par exemple, supposent une certaine maîtrise des outils informatiques. Toutefois, combler le fossé des compétences entre les journalistes et les technologues, ou aider les journalistes à développer de telles compétences en matière de données et en programmation informatique, n'est ni facile ou largement institutionnalisé (Lewis 2015). Dans l'optique où un système d'information est conçu comme un appui au travail journalistique, ces compétences ne se justifieraient plus puisque celui-ci fournit des synthèses et/ou des analyses livrées clé en main. Il s'agit donc d'une manière de pallier des compétences qui requièrent temps et pratique pour les maîtriser, d'autant que tous les journalistes ne sont pas nécessairement prêts à se plonger sous le capot de la machine. En comprendre les processus peut néanmoins faciliter le dialogue avec les professionnels de la technologie. S'il s'agit de professionnaliser la maîtrise du code informatique et de la manipulation de données, en admettant que les pratiques du journalisme computationnel font partie des possibles spécialisations du métier, au même titre que le journalisme audiovisuel ou que le photojournalisme par exemple, cela suppose que des formations spécifiques se mettent en place. Dans l'espace anglo-saxon, plusieurs programmes de ce type ont vu le jour dans la formation initiale des journalistes (Cardiff University, Columbia University, Georgia Institute of Technology, Stanford University, Syracuse University, Northwestern University, ...).

Ces considérations ne doivent pas faire oublier que le propre du journaliste est de donner du sens (Derville 1999). John Burn-Murdoch, datajournaliste au *Guardian* puis au *Financial Times*, souligne qu'au-delà des éléments tangibles que sont les données et les outils, la pertinence journalistique reste l'élément-clé : "*le journalisme de données n'est pas différent de toute autre forme de journalisme (...) l'essentiel est qu'il y ait une ou plusieurs personnes qui comprennent vraiment ce qu'est un bonne histoire basée sur les données*"⁷⁶. Pour Casswell et Dörr (2017), les journalistes devraient surtout apprendre à coder pour penser au journalisme de la même manière que les développeurs. De plus, il n'est pas garanti que les journalistes s'engageant dans l'apprentissage du code trouveront un emploi (Narin 2018). S'ils en trouvent un, le risque, selon Pilopher, est qu'ils ne soient jamais considérés comme des journalistes par leurs pairs mais bien comme des développeurs, et qu'ils restent cantonnés dans un service IT où leurs compétences ne seront pas exploitées autrement que dans le seul développement informatique, sans réelle perspective de carrière⁷⁷. De plus, souligne Pilopher, ces professionnels sont moins bien rémunérés que s'ils travaillaient dans le secteur privé. Aussi, les rédactions auraient-elles du mal à les retenir. La double compétence "journalisme-développement" relèverait donc davantage du handicap plutôt que d'un atout dont les rédactions tireraient parti.

⁷⁶ "Squirrel Talk with Data Journalist John Burn-Murdoch", Journocode, non daté (2017), consulté le 24/12/2017, URL : <http://www.advent17.journocode.com/door/23/>

⁷⁷ Aron Pilopher, thread publié sur Twitter le 27/01/2019, consulté le 28/01/2019, URL : <https://twitter.com/pilhofer/status/1089507217588609029>

Développer une pensée computationnelle

De nombreux chercheurs et professionnels engagés dans une vision prospective (Gynnild 2013, Coddington 2015, Broussard 2015, Karlsen & Stavelin 2014, Young & Hermida 2015, Lewis 2015, Rodgers 2015, Linden 2017) défendent l'idée selon laquelle, pour s'adapter aux innovations technologiques, les journalistes devraient développer un mode de pensée computationnelle. Cette prise de position s'inscrit dans un mouvement plus général qui revendique la transparence des procédures algorithmiques, laquelle ne pourrait être réalisée qu'à la seule condition de pouvoir le comprendre... et donc d'apprendre le langage de la machine. Gynnild (2010) soutient que le journalisme computationnel ne dépend pas d'outils ou de compétences mais du développement de la pensée computationnelle, qui se réfère à la manière de résoudre un problème, de concevoir des systèmes et de comprendre le comportement humain qui s'appuie sur des concepts fondamentaux de l'informatique comme les modèles de calcul et la solution algorithmique (Denning 2010). Cela implique d'intégrer des dimensions logiques, algorithmiques et scientifiques, et cela présuppose de l'ouverture et de la curiosité.

Bradshaw définit la pensée computationnelle comme la capacité à décomposer des problèmes logiques en séquences. Elle recouvre des compétences en abstraction et modélisation numérique, ainsi qu'en algorithmique⁷⁸. Pour Berret & Phillips (2016), la pratique avancée du journalisme de données nécessite un minimum de compréhension du fonctionnement de la programmation informatique, et elle peut être considérée comme une entrée en matière en termes de pensée computationnelle. Celle-ci concernerait moins les tâches effectuées par les ordinateurs que le processus cognitif que cela implique (Coddington 2015). Il ne s'agirait donc plus de penser au journalisme en tant que "produit" mais en tant que "processus" (Robinson 2011, cité par Lewis 2012). Nous pouvons toutefois considérer que les journalistes pratiquent déjà un mode de pensée similaire dans le "séquençage" de leurs routines. En prendre conscience et être capable de traduire ces séquences routinières en pseudo-code peut permettre de les associer à la conception d'outils liés à leurs pratiques. Considérant que le code informatique renferme des décisions éditoriales, il apparaît logique que ce soit aux journalistes de les poser. C'est aussi en ce sens que se justifie le développement d'une pensée computationnelle, utile pour les journalistes amenés à travailler au sein d'équipes multidisciplinaires.

"Penser" comme un ordinateur dès lors que l'on se retrouve confronté à de larges jeux de données induit de réfléchir en termes mathématiques et moins littéraires, ce qui contraste avec une tradition latine du journalisme davantage orientée vers le récit (Cornu 2009). Les bénéfices d'intégrer une démarche s'appuyant sur le calcul sont donc aussi ceux de pouvoir mieux appréhender des données chiffrées. Ici, on parlera moins de pensée informatique que de pensée mathématique, une qualité appréciable dans la mesure où derrière les chiffres, peuvent se trouver des récits journalistiques pertinents. Mais les journalistes témoignent souvent de réticences face aux mathématiques. Lorsque des étudiants en journalisme sont interrogés, ils dé-

⁷⁸ Paul Bradshaw, Medium, 11/09/2017, consulté le 12/09/2017, URL : <https://medium.com/thoughts-on-journalism/computational-thinking-and-the-next-wave-of-data-journalism-a2d801786bea>

clarent qu'ils ont choisi d'apprendre ce métier parce qu'ils veulent écrire et pas calculer (Curtin & Maier 2001, Schmitz-Weiss *et al.* 2018). Pour autant, les journalistes sont déjà habitués à travailler avec des chiffres : les votes s'expriment en nombre de voix, les budgets en montants, les résultats sportifs en points, le chômage en chiffres... (Maier 2003). Par ailleurs, la plus grande accessibilité à des programmes de formation en ligne fait souvent oublier la complexité de la programmation informatique, et occulte le fait qu'il s'agisse aussi d'un métier (McCosker & Milne 2014), au même titre que le journalisme en est un. Son apprentissage nécessite un temps qui n'est pas toujours disponible dans une profession régie par les lois de l'immédiateté. Cela étant, comprendre la manière dont fonctionne un programme informatique permet de mieux dialoguer avec le médiateur du code. Mais si la technologie peut faciliter l'exécution de certaines tâches répétitives, elle n'enlève rien au caractère indispensable du facteur humain.

Développer une pensée journalistique

Considérant que les journalistes sont amenés à travailler de plus en plus souvent avec des informaticiens, il ne faudrait pas que les pas s'opèrent à sens unique. La mise en œuvre d'artefacts d'automatisation de la production d'informations est d'abord le fait de professionnels issus d'un monde social qui, à bien des égards, va se distancier du monde du journalisme. Cela est d'autant plus vrai lorsqu'ils sont employés par des start-ups technologiques, dont la particularité est de ne pas se revendiquer comme un média, pas plus que l'on y considère faire acte de journalisme. Il s'agit là d'un autre enjeu de la production automatisée d'informations : comment intégrer ces professionnels dans le monde journalisme ? En premier lieu, cela suppose que ceux-ci admettent qu'ils participent à un processus éditorial, ce qui implique l'exercice d'une forme de responsabilité sociale régie par une série de droits et de devoirs qui devraient être communément partagés⁷⁹.

En second lieu, il s'agit de développer un mode de pensée journalistique, qui leur permette de mieux comprendre les spécificités et contraintes d'un processus éditorial. Dès lors que de nouveaux acteurs participent à la chaîne de production de l'information, il n'est pas possible de les envisager sous le seul prisme du monde de la technologie. Les choix techniques qu'ils posent ne sont pas dénués de neutralité, et cela d'autant plus qu'ils s'inscrivent dans le cadre d'un processus de traduction d'intentions humaines (Gillespie 2014). Ces choix sont intrinsèquement liés à un savoir et à un savoir-faire qui relèvent du domaine du journalisme, même si cette activité de traduction implique également un savoir et un savoir-faire qui relèvent du domaine de la technique. Il s'agit donc d'encourager une compréhension mutuelle en vue de favoriser des interactions fructueuses. Certains vont plus loin, suggérant que les pratiques des programmeurs qui font partie du "nouveau monde technologique du journalisme" soient balisées par un code de déontologie spécifique (Monti 2019). Ces aspects sont développés dans le deuxième chapitre, consacré aux dimensions de la culture professionnelle du journalisme⁸⁰, en relation avec la production automatisée d'informations.

⁷⁹ Voir "Dimension normative", p.122

⁸⁰ Voir infra, p.96

1.4 Imaginaires du "robot journalisme"

Le terme "robot" est régulièrement utilisé pour décrire des activités habituellement associées à des activités humaines traditionnelles (Latar 2015). Dans l'espace de l'information numérique, la métaphore du robot désigne des activités effectuées de manière automatique par des programmes informatiques, à l'instar du service d'agrégation Google News. Ce vocable a été utilisé pour la première fois en 1924 par l'écrivain Karel Tchapek pour désigner un androïde capable d'exécuter des travaux à la place de l'homme. En langue tchèque, "roboto" signifie "travail pénible, corvée" (Chavand 2017). Dans le contexte de la production automatisée d'informations, il s'agit d'une métaphore qui est régulièrement utilisée (Clerwall 2014).

Le recours à la symbolique du robot est un raccourci tentant car elle facilite la formation d'images mentales, mais elle est inappropriée car elle ne correspond en rien à la réalité (Linden 2017) : d'une part, parce qu'elle désigne un processus informatisé; d'autre part, parce qu'elle ne peut être considérée comme une activité journalistique à part entière. S'il s'agit bien d'automatiser des routines éditoriales, il ne s'agit jamais que de simuler l'activité d'écriture d'un journaliste humain en produisant, la plupart du temps, des textes courts et standardisés. Cette imitation sera toujours incomplète : aucun système de production automatisée d'informations ne peut, du moins jusqu'à présent, mener des interviews, expliquer des phénomènes, fournir une analyse approfondie ou participer à la formation des opinions publiques (Graefe 2016). Le magazine américain *Wired*, lorsqu'il souhaite rendre hommage à Marvin Minsky – considéré comme l'un des pères de l'intelligence artificielle – en automatisant la production du récit médiatique se heurte à cet écueil : le modèle du texte n'est pas entièrement automatisé en raison de données difficiles à extraire, et il n'intègre pas la dimension "touchante" d'une nécrologie rédigée par un humain⁸¹. Adrienne Lafrance, lorsqu'elle se met en tête de savoir si une machine est capable de rédiger comme elle, conclut que les machines devraient probablement s'en tenir au traitement de résultats sportifs ou à la rédaction de bulletins météorologiques. Le corpus qui avait nourri le système s'appuyant sur des algorithmes de *machine learning* n'était pas suffisant pour donner lieu à des résultats probants, malgré un volume de 725.000 mots correspondant à l'ensemble des articles rédigés par la journaliste américaine⁸².

La symbolique portée par la métaphore du robot place erronément la machine sur le même pied d'égalité que le journaliste. Voire, elle lui confère un caractère humain qui entraîne dans son sillage la représentation d'une machine qui prendrait le pas sur l'homme (Linden & Dieckx 2019). Elle interroge aussi l'autorité professionnelle et l'identité du journaliste (Neveu 2010). Ce faisant, elle véhicule un caractère éminemment anxiogène. Mais elle peut aussi produire l'effet inverse : la perception d'un agent physique plutôt que celle d'un processus logiciel peut tout aussi bien être la source d'un "syndrome du robot hollywoodien", selon lequel

⁸¹ "We Asked a Robot to Write an Obit for AI Pioneer Marvin Minsky", Adam Rogers, *Wired*, 28/01/2016, consulté le 07/07/2018, URL :

<https://www.wired.com/2016/01/we-asked-a-robot-to-write-an-obit-for-ai-pioneer-marvin-minsky/>

⁸² "À Computer Tried (and Failed) to Write This Article", Adrienne Lafrance, *The Atlantic*, 08/06/2016, consulté le 10/07/2018, URL : <https://www.theatlantic.com/technology/archive/2016/06/story-by-a-human/485984/>

le souvenir de films populaires tend à obscurcir cette anxiété, les robots apparaissant alors comme des artefacts sympathiques et amicaux (Waddell 2018). L'usage de cette métaphore s'inscrit dans une dynamique où les discours présentent différents niveaux d'abstraction. Ils constituent des "*manières de représenter des aspects du monde - les processus, les relations et les structures du monde matériel, le 'monde mental' des pensées, des sentiments, des croyances, etc.*" (Fairclough 2003 : 126). En tant que ressource disponible pour produire des représentations distinctes du monde, les métaphores mobilisent des représentations et des imaginaires (Fairclough 2003 : 131-132). Elles participent à la construction de la réalité et peuvent être utilisés consciemment, "*par une catégorisation impliquant la sélection de certaines caractéristiques aussi critiques que d'autres non critiques*" (Goatly's 1997 : 155, cité par Soler 2008). Les métaphores ont des caractéristiques cognitives, esthétiques et persuasives (Charaudeau 2002) : leur usage ne peut donc pas être considéré comme formellement neutre. Mais elles sont indispensables à la fois à la pensée et au langage (Soler 2008) : la majeure partie de notre système conceptuel ordinaire est de nature métaphorique, et celle-ci est omniprésente dans notre vie quotidienne (Lakoff & Johnson 1980).

Sur le continent asiatique, où la mythologie du robot est ancrée dans la culture populaire en tant qu'auxiliaire aimable et bienveillant (Munier 2014), cette métaphore n'en est plus une. Les premiers robots androïdes et interactifs sont expérimentés sur des plateaux de télévision, pour présenter des journaux télévisés ou pour mener des interviews. Sur le terrain sud-coréen, caractérisé par une société technologiquement très ouverte, il est intéressant de relever l'étude de Daewon et Seongcheol (2017) relative à l'attitude des journalistes face aux systèmes de production automatisée d'informations. Il en ressort une attitude partagée : entre une vision positive relevant d'une idée d'élitisme dans le journalisme, et une vision négative, voire dystopique, mettant en avant le "complexe de Frankenstein" où les moyens ne font aucun cas de la fin. Toutefois, les journalistes disaient ne pas croire que les "robots journalistes" allaient les remplacer, en raison de leurs spécificités humaines et des limites des systèmes techniques.

Les technologies d'automatisation de la production d'informations mobilisent des imaginaires susceptibles d'être influencés par l'usage répandu de la métaphore du "robot journaliste". Celle-ci participe à forger un ou des imaginaires construits à partir de représentations sociales collectives et individuelles, desquelles peuvent dépendre la formation des usages (Massit-Folléa 2012, Flichy 2001). Quelles sont celles proposées par les journalistes lorsqu'ils sont amenés à traiter ce phénomène dans le cadre de leurs activités professionnelles? Prennent-ils position ou, au contraire, se placent-ils en retrait de l'objet de leur discours? Ces questions se trouvent au centre de cette dernière section, consacrée à une analyse du discours médiatique des journalistes.

1.4.1 Méthode

Cette recherche combine une approche linguistique, où un corpus constitue à la fois la donnée de base et le point de départ, et une analyse critique du discours, où est étudiée la relation entre le langage et l'idéologie (Cheng 2013). En linguistique, un corpus implique "*un*

nombre fini de textes considérés comme représentatifs d'une variété particulière du langage à un moment spécifique" (McEnery & Wilson 2001, cités par Soler 2008). Les intérêts de cette approche résident non seulement dans ses aspects quantitatifs (Fairclough 2003 : 6), mais aussi dans sa capacité à identifier le sens contextuel de mots ou d'unités lexicales, et dans celle de fournir une analyse qualitative des niveaux syntaxiques et sémantiques (Soler 2008). Sa faiblesse se trouve dans le manque potentiel de représentativité et le risque d'échec dans les généralisations (Cheng 2013). Malgré sa valeur limitée, l'approche par corpus est porteuse d'enseignements. Cette méthode a été employée par Soler (2008) dans une recherche portant sur l'usage des métaphores dans les titres de presse. Il y a souligné qu'une approche par corpus rend possible l'identification du sens contextuel de mots ou d'unités lexicales, des structures syntaxiques et des métaphores employées. Elle permet également d'entreprendre une analyse qualitative des niveaux syntaxiques et sémantiques, le chercheur soulignant l'intérêt de la méthode lorsque le corpus est multilingue (ici, en anglais et en espagnol).

Dans cette analyse du discours médiatique des journalistes, deux services ont été utilisés pour collecter le corpus de recherche, constitué de titres d'articles publiés par la presse en ligne générale ou spécialisée, en anglais et en français : Google News, via des requêtes portant sur des termes identifiés comme caractérisant le domaine d'application ; et Twitter, utilisé comme un outil de veille et de recherche (fonctionnalités de recherche avancée). Les termes de recherche sont exposés en français, mais ils ont aussi été exploités dans leur déclinaison en anglais : "robot journalisme", "génération automatique de textes", "journalisme automatique", "journalisme automatisé", "newsbots", "journalisme algorithme", "journalisme intelligence artificielle". La recherche a également porté sur le nom des principales entreprises de génération automatique de textes actives dans le secteur des médias d'information (Syllabs, AEXEA, Automated Insights, Narrative Science, Retresco). Les articles émanant de ces sociétés et de publications d'organisations d'éditeurs de journaux, de même que ceux publiés dans un contexte de communication, de promotion et de marketing ont été écartés. S'il est préconisé de travailler sur de plus grands corpus en matière d'analyse du discours (Charaudeau *et al.* 2002), la limite est posée par la thématique spécifique, qui ne fait pas l'objet d'un traitement abondant et/ou systématique dans la presse en ligne. Toutefois, une approche basée sur un corpus peut être considérée comme bénéfique, y compris lorsque le corpus est petit (Soler 2008).

1.4.1.1 Approche internationale

Le corpus comprend 175 échantillons rédigés en anglais et 125 échantillons rédigés en français. L'évolution dans le temps de la quantité d'échantillons collectés (Figure 1.3) peut s'expliquer par l'évolution du phénomène, qui progresse lentement au début de la décennie, pour s'accélérer à partir de 2014, année à partir de laquelle il a franchi les frontières des États-Unis. Cette approche internationale est fondée sur le fait que le journalisme intègre des valeurs ou idéaux communs (Deuze 2005)⁸³, malgré la diversité potentielle des contextes de production de l'information et les différences significatives pouvant être constatées quant à la manière

⁸³ Cet aspect est plus particulièrement étudié dans le chapitre 2, "Entre deux mondes : les dimensions culturelles de l'automatisation", p.96

dont les journalistes abordent leurs pratiques (Hermans & Vergeer 2009, cités par Spyridou *et al.* 2013). Par exemple, si l'on compare les États-Unis et la France, les structures médiatiques et les cultures journalistiques sont souvent décrites comme des contraires polaires en termes de forces du marché, de rôle de l'État et de forme de l'information (Christin 2016).

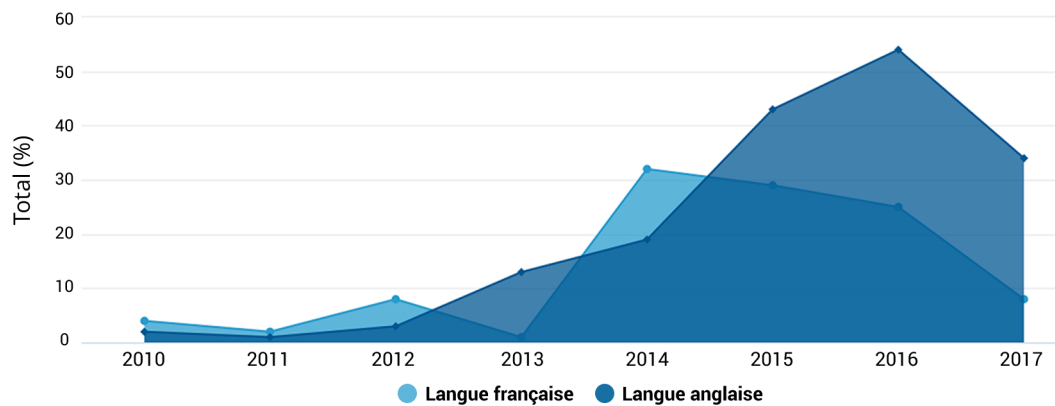


FIGURE 1.3 – Corpus de recherche

En tant qu'idéologie professionnelle, le journalisme véhicule "un système de croyances caractéristique d'un groupe particulier, incluant, sans toutefois s'y limiter, le processus général de production de significations et d'idées (au sein de ce groupe)" (Deuze 2005). L'objectif était donc de comprendre en quoi les journalistes peuvent partager des représentations communes à propos du phénomène de la production automatisée d'informations. Toutefois, des comparaisons ont été effectuées entre les pays de publication (Figure 1.4) ou par sous-corpus en fonction de la langue utilisée (français ou anglais).

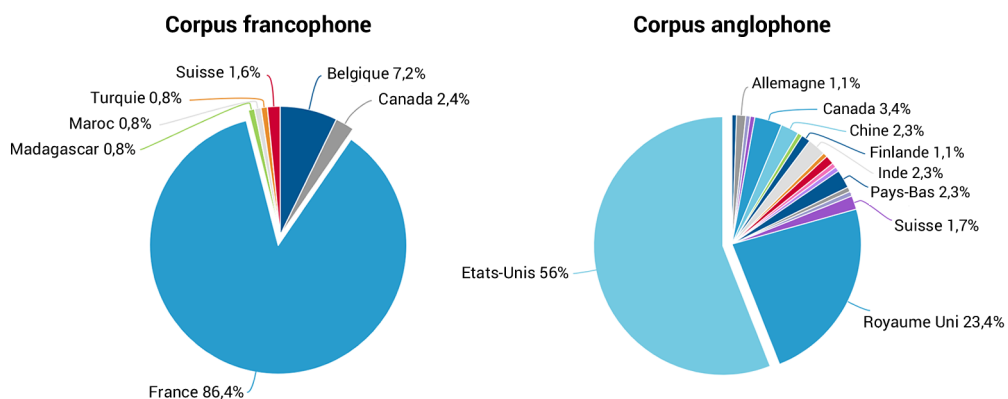


FIGURE 1.4 – Répartition du corpus par langue et par pays

L'analyse comparative "est précieuse dans la recherche sociale car elle nous sensibilise à la variation et à la similitude et cela peut contribuer de manière puissante à la formation de concept et à l'affinement de notre appareil conceptuel" (Hallin & Mancini 2004 :4). Aussi, la comparaison est-elle envisagée comme un moyen de comprendre, "d'éviter d'essentialiser des dynamiques

qui ne sont propres qu'à un seul cas plutôt qu'à un phénomène tout entier", et comme "un instrument heuristique — théorique et méthodologique — pour découvrir et expliquer des rapports et — éventuellement — des liens de causalité autour de phénomènes, quelle que soit leur nature" (Anciaux et al. 2017).

1.4.1.2 Focus sur les titres

Le corpus comprend 300 titres d'articles publiés en ligne entre 2010 et 2017. Ce parti pris se justifie par la force d'attraction exercée par cette unité indépendante et sémantique d'une production journalistique (Kronrod & Engel 2001, Ho-Dac et al. 2013), dont l'objectif est de stimuler la lecture. C'est également la partie la plus visible d'un article lorsqu'il apparaît à la suite d'une requête sur les moteurs de recherche ou sous forme de lien partagé sur les réseaux sociaux. Les titres seraient également la seule partie d'un article que les lecteurs liraient lors du partage d'un lien (Gabiolkov et al. 2016). Ces différents contextes d'usage accordent au titre un statut autonome. Le titre est donc ici considéré comme un segment unique, "*qui forme de ce fait un discours particulier et peut donc être analysé pour lui-même*", à la fois unité indépendante du texte qu'il chapeaute et sous-genre journalistique (Kronrod & Engel 2001). Le titre contient habituellement des informations concises, à valeur éditoriale ajoutée. La nécessaire densité de cette microstructure sémantique a pour résultat une hétérogénéité des dispositifs discursifs pouvant être répartis en deux catégories : les titres informatifs, sémantiquement neutres en s'en tenant à une description des faits; et les titres incitatifs, dont la rhétorique exacerbe leur sens (usage de métaphores, d'exagérations, de formules provocantes...) (Herrero Cecilia 2007). Rebeyrolle et al. (2009) considèrent qu'il existe deux autres types de titres : référentiels, relatifs au sujet du discours⁸⁴; et thématiques, centrés sur l'objet du discours. Le fonctionnement discursif des titres "*contribue à la construction d'un modèle interprétatif du discours*". Trois fonctions peuvent y être associées : dénomminative (ou désignative), méta-discursive (le titre donne au lecteur une indication sur la nature ou le contenu du texte) et séductrice (qui donne envie de lire le texte).

Charaudeau (cité par Soler 2008) souligne que le titre de presse joue trois rôles distincts : "phatique", de prise de contact avec le lecteur; "épiphanique", d'annonce de la nouvelle; et "synoptique", de guidage du parcours visuel du lecteur dans l'espace informatif du journal. Ho-Dac et al. (2013) identifient trois méta-fonctions inhérentes au titre : interpersonnelle (lorsque le titre est incitatif), idéationnelle (c'est la manière dont le titre présente l'expérience du monde), et textuelle. La méta-fonction textuelle concerne le titre en tant que texte et la manière dont il est organisé. Il est à appréhender comme "*une unité, un tout cohésif, comme quelque chose qui construit son propre monde*"; et il est défini comme procurant au locuteur un ensemble de stratégies de mise en texte pour guider l'interprétation. Pour de Villers et al. (2010), les titres de presse sont à la fois des reflets et des modèles de société. Leur étude permet de suivre l'évolution de la langue, de dégager certaines tendances et de "*circonscrire des traits définitoires de la norme linguistique d'un groupe culturel donné*". Dor (2003) suggère que le titre d'un article

⁸⁴ Ces types de titres ont toutefois moins pour objet de spécifier un référent que de poser un cadre permettant de spécifier le contexte du discours.

apporte aux lecteurs un contexte de représentation et que, pour ce faire, il fait l'objet d'une "optimisation pertinente". Un "bon titre" aurait ainsi pour vocation de fournir un maximum d'effets contextuels pour un minimum d'efforts cognitifs, alors que le processus de rédaction des titres implique de jongler en permanence avec de nombreux impératifs de communication différents et parfois contradictoires. Si le titre dirige la compréhension du lecteur, son impact s'explique par ses caractéristiques linguistiques qui le rend mémorable et efficace, encapsulant non seulement le contenu de l'article, mais également son orientation. Les titres offrent une perception globale et synthétique d'un événement sous des caractéristiques considérées comme essentielles pour les lecteurs.

En raison de leur nature condensée, les titres sont susceptibles de laisser des nuances ou d'autres aspects dans l'ombre, ce que le lecteur ne pourra connaître qu'en lisant le corps de l'article (Herrero Cecilia 2007). Le choix de ne se concentrer que sur les titres peut donc être considéré comme un biais, dès lors que le contenu d'un article peut nuancer l'objet de son titre, voire ne pas lui correspondre sur le plan sémantique. Cela étant, ce choix est approprié dans le cadre d'une approche internationale, dès lors que les structures de titre semblent être très régulières dans toutes les langues, malgré les différences pouvant être observées entre les représentations culturelles (Develotte & Rechniewski 2001).

1.4.1.3 Diversité des médias et des auteurs

Les aspects relatifs à la diversité du corpus ne sont pas seulement liés à son caractère international, mais également à celui des publications – toutes diffusées en ligne – prises en compte⁸⁵ (Figure 1.5).

- Médias audiovisuels (7,67%) : e.a. CNN (États-Unis), Sverige Radio (Suède), BBC (Royaume-Uni), France Télévisions (France), Radio Canada (Canada), RTBF (Belgique).
- Presse quotidienne, y compris les agences de presse (19%) : e.a. *The Guardian* (Royaume-Uni), *The New York Times* (États-Unis), *Le Monde* (France), *The Vancouver Sun* (Canada), *Le Soir* (Belgique).
- Magazines (13,33%) : e.a. *Wired* (Royaume-Uni et États-Unis), *Mashable* (États-Unis), *Le Vif / L'Express* (Belgique), *Le Nouvel Observateur* (France), *Paris Match* (France).
- Pure players (28%) : e.a. *The Huffington Post* (États-Unis), *Slate* (France), *Digiday* (Royaume-Uni), *Quartz* (États-Unis), *Politico* (États-Unis), *Rue 89* (France).
- Blogs (32%) : e.a. blogs hébergés notamment sur Medium.com et Theconversation.com, Poynter (États-Unis), Médiacadémie (France).

La diversité du corpus tient également dans celle de ses auteurs, dès lors qu'un article de presse peut être rédigé non seulement par des journalistes professionnels, mais aussi par des correspondants ou des collaborateurs occasionnels (Moirand 2007). De plus, l'auteur d'un article n'est pas toujours celui du titre : les éditeurs finaux et les correcteurs jouent également un rôle

⁸⁵ Une base de données reprenant l'ensemble des publications du corpus est consultable en ligne : <https://www.ohmybox.info/these/discours/corpus.html>; tandis que l'ensemble des résultats, présentés sous la forme de graphiques, sont accessibles via <https://www.ohmybox.info/these/discours/> (utilisateur + mot de passe : t2020ulb)

actif dans le processus éditorial, dès lors qu'ils peuvent être amenés à corriger ou à modifier un titre : "*Leurs interventions contribuent à maintenir la crédibilité et la qualité d'un journal et améliorent l'expérience de lecture*" (Vandendaele 2018). Dans ce contexte, ils sont susceptibles d'être amenés à modifier ou réécrire les titres. Cependant, en tant que maillon de la chaîne éditoriale, nous considérons qu'ils appartiennent au monde social de l'information d'actualité dans sa plus large acceptation. Par ailleurs, en intégrant des blogs, le corpus reflète l'extension du champ du journalisme observée depuis plusieurs années (Le Cam 2006). Les discours journalistiques ainsi que les discours méta-journalistiques, faisant référence à des expressions publiques du journalisme, ont donc été pris en compte dans cette étude (de Maeyer & Malcorps 2015).

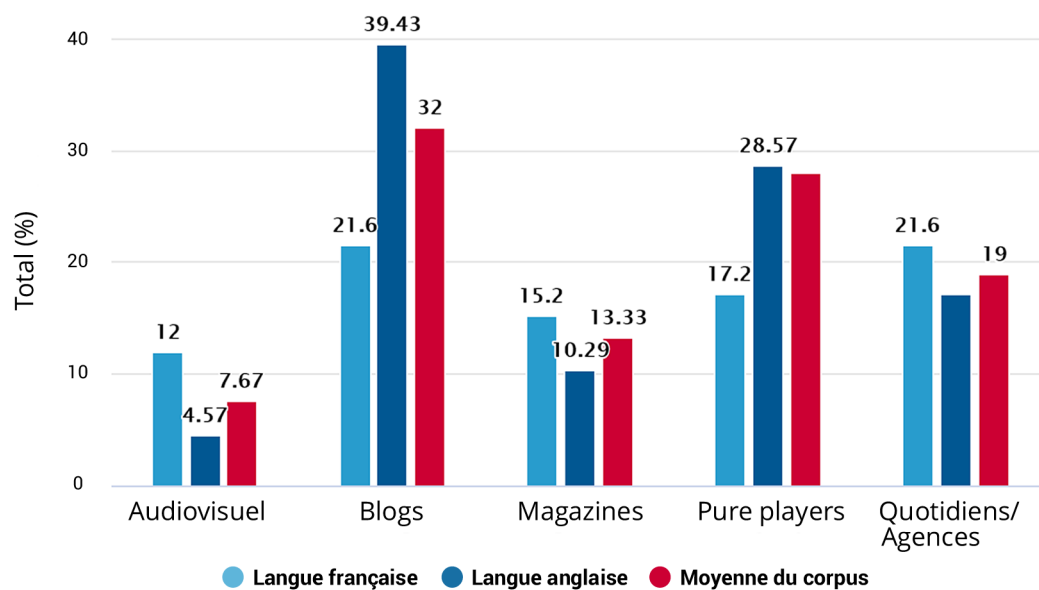


FIGURE 1.5 – Répartition des échantillons du corpus par type de média en ligne.

Ce corpus, qui ne prétend pas être exhaustif, peut donc être considéré comme significatif en raison de la diversité de ses sources (articles issus de la presse en ligne quotidienne et magazine, articles publiés sur des sites ou blogs spécialisés, articles rédigés en langue française ou anglaise), de ses auteurs représentant le monde "élargi" du journalisme, et des pays qui y sont représentés.

1.4.1.4 Modèle d'analyse du discours

Une analyse du discours fournit des renseignements sur la structure d'un texte et le rôle de chaque élément au sein de cette structure, contrairement à la linguistique descriptive qui se borne à décrire le rôle de chaque élément dans la structure de la phrase qui le contient. Elle établit donc un rapport entre la culture et la langue (Harris & Dubois-Charlier 1969). Dès lors, une analyse qui ne s'appuierait que sur les seuls niveaux lexical et syntaxique, se limiterait à une analyse formelle sans tenir compte du sens, cette "couleur" donnée au discours. Y adjoindre le

niveau pragmatique suppose que ce sens soit déterminé par l'action, même si celle-ci n'est pas explicitement induite. C'est pourquoi le modèle d'analyse proposé tient compte des quatre niveaux d'analyse du discours (lexical, syntaxique, sémantique et pragmatique), dans l'esprit de les mettre en relation avec l'objet de la recherche (Moeschler 1985).

Charaudeau (1995) parle de sémiolinguistique pour désigner le phénomène de construction psycho-langagière réalisée par l'intervention d'un sujet. C'est pourquoi, du point de vue des sciences du langage, une analyse du discours ne peut qu'être empirico-déductive. Il considère également une seconde option, correspondant à une perspective textuelle, qui tente de décrire *"la façon la plus exhaustive possible, les traits qui le caractérisent"*. Charaudeau distingue l'analyse du discours de l'analyse textuelle : la première approche, qui consiste à analyser un texte de manière linéaire, est *"le résultat d'une combinaison de certaines conditions de production avec des opérations de mises en discours"* ; tandis que la seconde porte sur un corpus de textes *"rassemblés au nom d'un type de situation (contrat) qui les sur-détermine, et dont on étudie les constantes (pour définir un genre), et les variantes (pour définir une typologie des stratégies possibles)"*. Il souligne que le linguiste du discours *"doit traquer le sens au-delà de l'emploi des mots et des constructions (.. et aller..) voir derrière le masque de l'effacement énonciatif, celui du positionnement discursif"*.

Les énoncés journalistiques correspondent *"à la façon dont l'énonciateur met en scène le discours d'information à l'adresse d'un destinataire imposé en partie par le dispositif et en plus imaginé et construit par lui"* (Charaudeau 2006). Il s'agit d'un contrat d'énonciation qui n'est pas exempt de partialité : chacun des choix posés dans la chaîne de production journalistique ne peut pleinement satisfaire les principes de neutralité et de distance. Le positionnement du journaliste dépend d'un ensemble de procédés discursifs et *"d'un ensemble de mots dont le sémantisme est révélateur de son positionnement au regard de certaines valeurs, le tout en rapport avec les conditions situationnelles de production"* (Charaudeau 2006). Van Dijck (1985) souligne l'influence des contraintes sociales, culturelles et cognitives sur les propriétés organisationnelles des messages multimédias, supposant qu'il existe une relation systématique entre texte et contexte. En d'autres mots, les formes structurelles et les significations générales d'un texte d'information résultent de routines sociales et professionnelles. Elles sont également liées au format du média ainsi qu'à des contraintes éditoriales ou organisationnelles.

En linguistique, trois niveaux d'analyse du discours cohabitent traditionnellement : lexical, caractérisé par l'usage d'un vocabulaire/lexique particulier ; syntaxique, relatif à la relation entre les signes (structure de la phrase) ; sémantique, désignant le sens de la phrase dans sa relation entre les signes et leurs référents. La tradition sémiotique, indique Moeschler (1985), retient les niveaux d'analyse syntaxique et sémantique et y adjoint un niveau d'analyse pragmatique, lequel étudie *"la relation entre les signes et les usagers ou, de façon plus précise, l'emploi du système par les utilisateurs (...) C'est l'étude du sens des énoncés en contexte"*. Moeschler précise que la sémantique est attachée à une fonction représentationnelle, alors que la pragmatique est attachée à une fonction instrumentale.

À. Niveau lexical : relatif à l'usage d'un terme de référence lié au domaine. Quelle unité lexicale ou locution est le plus fréquemment utilisée pour désigner l'automatisation de la production d'informations? Existe-t-il des caractéristiques communes, malgré l'hétérogénéité du corpus? La locution métaphorique de "robot journaliste" est-elle passée dans le langage journalistique ou un vocabulaire technique est-il privilégié?

B. Niveau syntaxique : permet d'identifier un premier type de titre, en relation avec la structure de la phrase et ses règles de combinaison (affirmative, exclamative, interrogative). La variable verbale a été retenue, dans la mesure où cette forme syntaxique n'est pas sans incidences sur le niveau sémantique du titre, dès lors qu'elle répond à une activité de conceptualisation et de concision. Elle est également susceptible d'influer sur les niveaux sémantique (en renforçant un caractère affirmatif) et pragmatique (induisant une relation entre le texte et son usager dans le cadre, par exemple, d'une injonction performative). Le niveau syntaxique est donc ici envisagé comme pouvant se rapporter à des fonctions sémantiques différentes (Garric 2013). De même que le niveau lexical, il est considéré comme un support pour l'analyse des deux niveaux suivants.

C. Niveau sémantique : ce niveau couvre le sens de la phrase et son rattachement à une fonction de représentation (Moeschler 1985). Il est relatif à l'engagement actif de l'auteur dans le titre (Englebretson 2007 : 3), considérant que la position du journaliste dépend d'un ensemble de processus discursifs où la sémantique d'un ensemble de mots est "*indicative de son positionnement avec le respect de certaines valeurs*" (Charaudeau 2006). L'étude du niveau sémantique se déroule en trois temps : (1) identification du type de titre (informatif ou incitatif), (2) identification de la nature du titre (référentielle ou thématique), (3) identification de la connotation du titre (neutre, positive, négative).

La connotation d'un titre est considérée comme neutre lorsque le titre est uniquement lié à des faits avec un retrait énonciatif. Le concept d'objectivité n'est pas ici considéré, car sa définition peut être interprétée de différentes manières (Donsbach & Klett 1993). Bien qu'il apparaisse comme un élément clé de la perception professionnelle de l'idéologie du journalisme, le concept d'objectivité est contesté par les chercheurs depuis des décennies. Il s'inscrit dans la mouvance de courants de pensées qui tendent à le légitimer, tels que le positivisme, le journalisme scientifique ou le journalisme de précision (Wien 2005). L'impossibilité de la neutralité est régulièrement soulignée non seulement par les chercheurs (Watine 2004, Cornu 2009, Munoz-Torres 2012), mais aussi par les journalistes eux-mêmes (Deuze 2005). Comme tout acteur social, le journaliste ne peut se débarrasser de ses propres référents ou représentations socioculturelles (Tuchman 1978). De plus, il est impossible de rapporter des faits sans les interpréter (Cornu 2009 : 323-396). Dans cette perspective, le choix du sujet et la manière de le présenter ne peuvent être neutres, malgré la distance que les journalistes peuvent établir entre eux et leur sujet. De plus, aucun choix éditorial ne peut être séparé de la ligne éditoriale des médias, qui est définie en fonction des intérêts supposés d'un public cible (Ruellan 2005). C'est

aussi pour cette raison que le discours d'actualités n'est pas seulement un texte : c'est aussi un texte en (con)texte (Catanuccio *et al.* 2011).

Malgré une dimension positiviste, également contestée (Wien 2005), une approche axée sur le caractère factuel du titre a été privilégiée pour évaluer le positionnement "neutre" du titre. Si la syntaxe est interrogative, aucune réponse engagée de l'auteur ne doit être induite par la sémantique. Un titre dont la connotation est considérée comme positive s'appuie sur un usage de termes et de locutions dont la connotation est plutôt positive. Cela comprend notamment les unités lexicales et locutions suivantes (ainsi que leurs équivalents en langue anglaise) : bon, rentable, gagner, surprendre, délivrer/délivrance, parier, nouveau, pas de menace (négation du négatif), mariage, avenir, futur, sauver l'emploi, bienvenue, ça marche, bénéficiaire/bénéfices, bonne nouvelle, plus. Un titre dont la connotation est considérée comme négative s'appuie sur un usage de termes et de locutions dont la connotation est plutôt négative. Cela comprend notamment les unités lexicales et locutions suivantes (ainsi que leurs équivalents en langue anglaise) : remplacer des humains, contre, opposition, guerre, attaque, maltraitance, sans journalistes, survivre, menace, apocalypse, frontière, fin, perdre l'emploi, perdre/perdant, au revoir, adieu, peur, pas/ne pas. Cette approche lexicale est, toutefois, insuffisante pour évaluer la connotation générale d'un titre : c'est pourquoi le sens global induit par les niveaux syntaxique et sémantique a été pris en compte.

D. Niveau pragmatique : fait référence à la relation entre les signes et entre les signes et leurs référents. Il est donc rattaché à la fonction instrumentale du titre (Moeschler 1985). Ce niveau est examiné sous l'angle des liens entre les titres et les enjeux professionnels liés à l'automatisation de l'information : éthique, avenir ou progrès, relation homme-machine, menace sur l'emploi ou sur l'identité professionnelle. Ces enjeux ne peuvent être abordés que de manière explicite ou de manière induite par le sens de l'énoncé. Par exemple, le concept d'identité professionnelle, de par sa nature complexe, s'est essentiellement limité aux expressions assimilées à "Cet article n'a pas été écrit par un robot". Bien que cela apparaisse comme une affirmation identitaire forte, elle est également liée la dualité de la relation homme-machine.

Les résultats de cette analyse ont été comparés à ceux d'une analyse des similitudes dans chaque corpus linguistique. Ce principe, issu de la théorie des graphes, se traduit dans la création graphique d'arbres maximaux. Un arbre maximum consiste en une simplification extrême des similitudes, seules celles présentant des maximums locaux sont conservées pour donner lieu à la représentation la plus simplifiée possible tout en conservant la connexité. Une chaîne d'un arbre est maximale si elle n'est contenue dans aucune autre chaîne (Degenne & Vergès 1973). Les résultats ont également été comparés à ceux d'une analyse automatique de sentiments. Ces dernières étapes avaient pour objet d'affiner les résultats et de pondérer un travail d'interprétation qui, comme toute activité d'analyse, comporte une part de subjectivité (Espeland & Stevens 2008).

1.4.2 Résultats

Au niveau lexical (Figure 1.6), les journalistes ont plutôt tendance à faire usage de la métaphore du robot, que ce soit en anglais ou en français (76,49% en moyenne), et à délaisser des termes spécifiques plus appropriés (logiciel, algorithme). Cela témoigne du passage de la métaphore du robot dans la langue. Cette métaphore est davantage utilisée dans le corpus en langue française (83,2%). Le recours à cette métaphore participe aux activités de vulgarisation qui caractérisent les écrits journalistiques. La reformulation s'explique en raison d'un vocabulaire scientifique ou technique qui n'est pas nécessairement compris par tous les publics auxquels s'adressent les médias (Moirand 2007). La métaphore permet également de représenter un processus logiciel abstrait, en témoignent les illustrations accompagnant ces articles : du bras mécanique pianotant les touches d'un clavier, au robot "cartoonisé" muni d'une carte de presse.

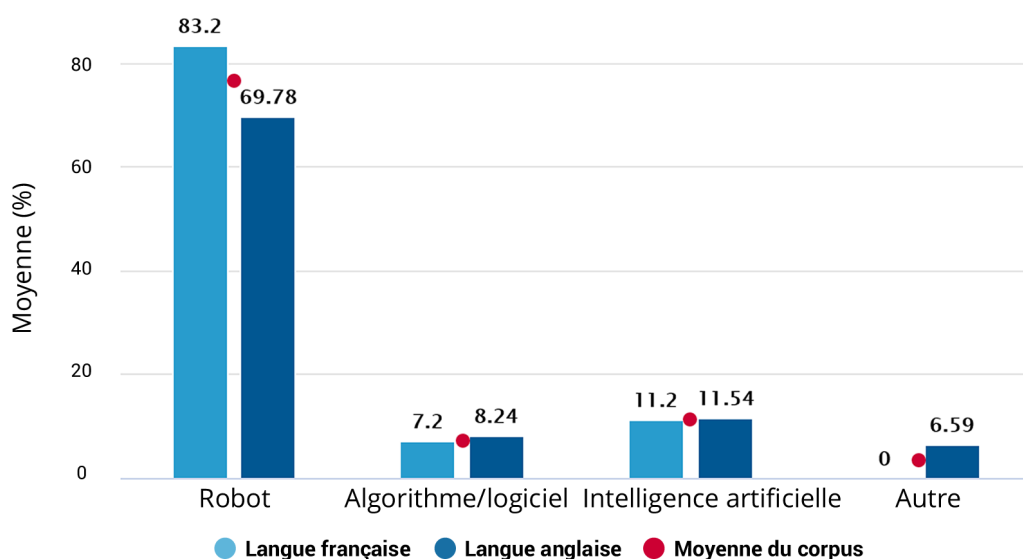


FIGURE 1.6 – Analyse du discours : termes employés

Un glissement lexical (et sémantique) vers la locution "intelligence artificielle" est observé depuis 2016. Dans le monde anglophone, la tendance est à utiliser davantage la locution "intelligence artificielle" (11,53%). Cette utilisation peut laisser supposer un effet de mode, si l'on considère également l'augmentation constante des requêtes effectuées, au cours de ces six dernières années via le moteur de recherche Google (Figure 1.7). Au niveau syntaxique, la majorité des titres (65,33%) consiste dans des affirmations. Ceci n'influe en rien sur le caractère "neutre" du titre : dans 56,33% des cas, une prise de position est constatée et celle-ci peut s'exprimer de diverses manières. Par exemple, "Tremblement de terre de Los Angeles : un bon exemple d'article écrit par un robot" (*Slate*, France) indique un caractère positif; tandis que "À robot took my job" (*Sparksheet*, Canada), dont le mode est également affirmatif, renvoie à l'anxiété professionnelle induite par le phénomène d'automatisation. Les titres informatifs au sein du corpus en langue française (61,6%) sont plutôt neutres (55,26%). Cette proportion passe à 71,77% pour les titres affirmatifs détectés dans le corpus en langue anglaise.

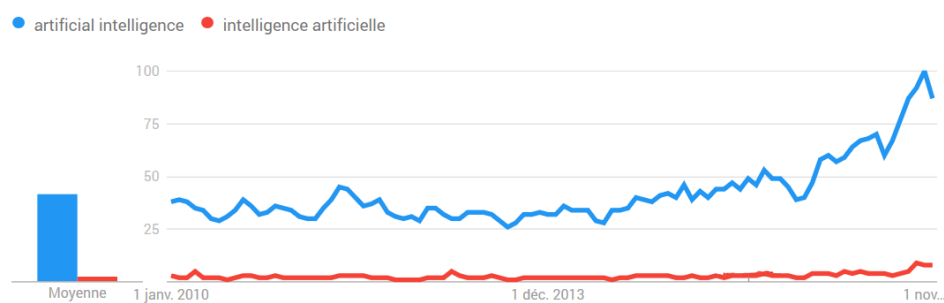


FIGURE 1.7 – Évolution de la recherche des locutions "Intelligence artificielle" et "Artificial intelligence" dans Google Trends

Les modes exclamatifs et impératifs sont très marginaux, tandis que le mode interrogatif est utilisé dans 20,8% des cas. Les questions posées reflètent des préoccupations essentiellement liées à une course contre la machine : "Un algorithme peut-il mieux écrire qu'un humain?", "Les robots vont-ils remplacer les journalistes?", "Fin de parcours pour les journalistes?", "Journalistes de demain : tous robots?", "Un robot a-t-il écrit cet article?", "Écrit par un robot : l'algorithme va-t-il tuer le journalisme?". Toutefois, sur le plan sémantique, un titre interrogatif ne conduira pas nécessairement à des représentations négatives : cela a été observé, en moyenne pour l'ensemble du corpus, dans 28,33% des cas. Dans le corpus anglophone, les titres averbaux (13,71% des cas) concernent des titres thématiques (87,5%) et informatifs (66,67%). Au sein du corpus francophone, les titres averbaux (31,2% des cas) concernent les titres thématiques (66,67%) et informatifs (56,41%). Les connotations négatives ont été trouvées principalement dans les titres thématiques (74,78%), qui couvrent 57% du corpus.

Les différences entre les titres en langue anglaise et ceux en langue française sont marginales en ce qui concerne leur nature informative ou incitative. Les titres informatifs sont majoritairement privilégiés, avec une moyenne de 72%. Dans une large majorité, ils ceux-ci ne sont porteurs d'aucune connotation (63,49%). La connotation négative est davantage marquée dans les titres publiés en français : 25,6% contre 12,57% en anglais. Le monde anglophone aborde aussi le sujet de manière sensiblement plus positive : 19,43% contre 16,8%. Les titres de publications émanant du Royaume-Uni observent une attitude plus neutre dans le corpus anglophone (77,6% contre une moyenne, pour l'ensemble du corpus en langue anglaise, de 68,57%).

Les journalistes abordent davantage le domaine d'application dans ses aspects thématiques (57%), plutôt que spécifiques (par exemple, lors du lancement d'un nouveau projet d'automatisation dans une rédaction). Peu de différences sont observées entre les deux sous-corpus. Cela signifie qu'une connotation positive ou négative consiste en la représentation d'un phénomène davantage traité dans sa globalité que dans sa singularité référentielle. Le sujet est souvent abordé comme une nouvelle forme de journalisme qui induit l'idée d'une compétition ("Are companies finally hiring robot writers to automate content creation?", "Human journalists hate robot journalists, says new report", "Innovation : avec les robots, la fin des journa-

listes?", "Robots vs journalistes : 1-0") ou de complémentarité entre l'homme et la machine ("How algorithms and human journalists will need to work together", "AI is going to be helpful for personalizing news — but watch out for journalism turning into marketing", "South Korea developing AI system to keep journalism alive", "Quand les journalistes et les robots travaillent main dans la main", "Le journalisme est sauvé, merci aux algorithmes"). Cet échantillon n'est donc pas dénué de force représentationnelle, la technologie étant abordée tantôt comme une force destructrice, tantôt comme une force créatrice.

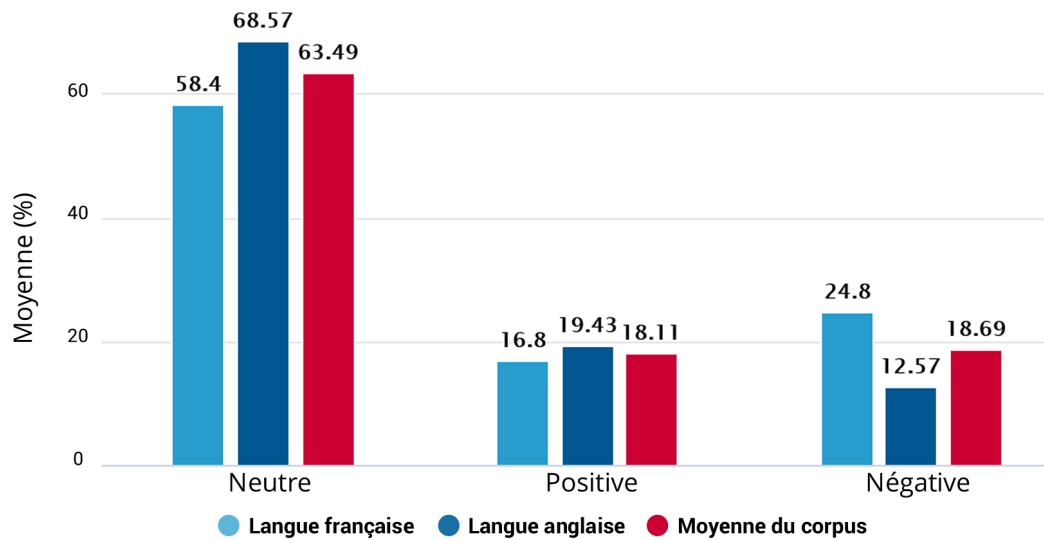


FIGURE 1.8 – Analyse du discours : connotation des titres

La combinaison des niveaux syntaxique et pragmatique montre que les titres impératifs, reflétant une injonction performative liée à "*ce qui se fait avec des mots*" (Fish 1976), sont absents du corpus en langue française et concernent 2,86% des titres du corpus en langue anglaise. Les injonctions sont celles d'apprendre à ne pas avoir peur ("Learn to stop worrying and love robot journalist"), d'imaginer ("Imagining a newsroom powered by artificial intelligence"), de s'habituer ("Get used to automation in newsroom") et d'avoir peur ("Be afraid... robotic journalism" – sic). Aucune incidence particulière de l'usage de la métaphore du robot n'est constatée sur la connotation générale des titres (Figure 1.8). Toutefois, en examinant de plus près chaque sous-corpus, il apparaît que la connotation négative tend à augmenter dans le corpus en langue anglaise lorsque cette métaphore est utilisée (de 12,57% à 14,96%).

Le phénomène inverse est observé au sein du corpus francophone, où le taux de connotation positive augmente légèrement (de 16,8% à 18,27%) alors que la connotation négative diminue (de 24,8% à 24,04%). Cela signifie que si le corpus en langue française présente une connotation plus négative, l'utilisation de la métaphore contribue à donner une représentation plus positive du phénomène. À l'inverse, l'utilisation de la métaphore a un impact négatif sur le corpus en langue anglaise. Si la métaphore semble affecter la perception des rédacteurs de chaque sous-corpus, elle joue donc un rôle opposé en déforçant la tendance négative ou positive, selon la langue.

Si l'on examine le poids de la métaphore du robot selon le type de publication (Figure 1.9), il apparaît que la tendance est un peu plus négative si l'on extrait les blogs du corpus (de 19,05 à 20%) et qu'elle l'est encore un peu plus lorsque l'on extrait les blogs et les pures players, nés dans le giron d'internet (la moyenne s'élève alors à 24,18%). Le sous-corpus francophone influe davantage sur cette tendance, avec une moyenne de 25,93% sans les blogs et de 28,3% sans les blogs et les pures players. Les blogs francophones présentent une moyenne de connotation négative plus faible (17,39%) et positive plus forte (21,74%). Ce phénomène n'est pas observé dans le corpus en langue anglaise où la connotation positive est sensiblement plus élevée sans les blogs et sans les pures players (21,95%), de même que la moyenne relative à la connotation négative (21,95%). Cela signifie que les blogs et les pures players en langue française donnent une représentation légèrement plus positive du phénomène de l'automatisation de la production d'informations. Dans le monde anglophone, ce sont les médias traditionnels qui contribuent à en donner une image plus positive.

Type	Neutre	Positif	Négatif
FR (uniquement usage de la métaphore)	57.69	18.27	24.04
FR (global)	58.40	16.80	24.80
Blogs FR	60.87	21.74	17.39
Sans les blogs FR	56.79	17.28	25.93
Sans les blogs et pure players FR	62.26	9.43	28.30
EN (uniquement usage de la métaphore)	66.14	19.69	14.96
EN (global)	68.57	19.43	12.57
Blogs EN	66.67	18.75	14.58
Sans les blogs EN	65.82	20.25	15.19
Sans les blogs et pure players EN	58.54	21.95	21.95

FIGURE 1.9 – Influence de l'usage de la métaphore du robot par type de publication

Au niveau pragmatique (Figure 1.10), les titres relatifs aux enjeux liés à la production automatisée d'informations (relation homme-machine, progrès/avenir, identité, emploi et éthique) sont le plus souvent liés à une approche thématique plutôt que référentielle. Cela signifie que les auteurs des titres s'impliquent davantage dans leurs discours dans un contexte général. Dans le corpus en langue française, les moyennes les plus hautes sont observées pour les enjeux relatifs à la relation homme-machine (35,2%) et à ceux de l'emploi et de l'identité professionnelle (16% dans les deux cas). Dans le corpus en langue anglaise, les moyennes les plus élevées sont observées en ce qui concerne les enjeux relatifs à l'avenir et à la relation homme-machine (respectivement, 21,14% et 18,86%). Dans le corpus en langue française, l'enjeu de l'identité professionnelle véhicule une connotation plutôt négative (80%). Il est suivi, dans ce registre, par ceux relatifs à l'emploi (60%) et à la relation homme-machine (40,91%). L'enjeu véhiculant la connotation positive la plus importante est celui de la relation homme-machine (38,1%). Il est suivi par

celui du futur, qui induit les notions d'avenir et de progrès (31,25%). Dans le corpus en langue anglaise, l'enjeu de l'emploi revêt la connotation la plus négative (47,37%). Il est suivi par celui relatif à l'identité professionnelle (42,86%). Toutefois, les enjeux abordés dans ce sous-corpus présentent plutôt un caractère dénué de prise de position, à l'exception de celui se rapportant au futur, revêtant un caractère plutôt positif (21,62%). Les résultats observés dans le corpus en langue anglaise doivent être ici nuancés. L'enjeu portant sur la relation homme-machine est davantage abordé au Royaume-Uni et il y est perçu de manière plus positive qu'aux États-Unis (20% contre 7,14%). Celui de l'emploi est moins prégnant au Royaume-Uni (7,1% contre une moyenne générale de 13% et une moyenne de 10,2% aux États-Unis, témoignant ainsi d'une préoccupation sensiblement plus forte en dehors de ces deux pays). Enfin, celui portant sur les questions éthiques est plus important aux États-Unis (7,07%) qu'au Royaume-Uni où il n'est pas du tout abordé. Cela pourrait s'expliquer par le fait que les technologies d'automatisation de la production d'informations s'y développent depuis plus longtemps qu'ailleurs : les questions éthiques y auraient donc émergé plus rapidement. Il s'agit là des différences les plus remarquables observées au sein du corpus en langue anglaise.

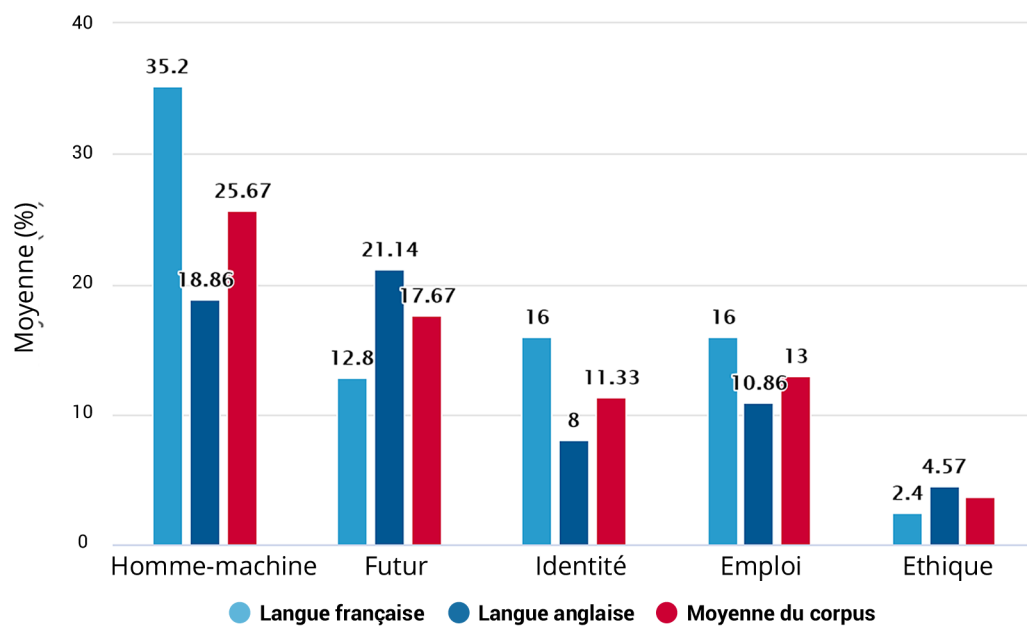


FIGURE 1.10 – Analyse du discours : enjeux socioprofessionnels (niveau pragmatique)

En considérant l'ensemble des résultats, des préoccupations communes apparaissent, celles de l'avenir de l'emploi journalistique et de l'identité professionnelle des journalistes. À ce propos, quatre auteurs du corpus en langue française indiquent que leur article n'a pas été rédigé par une machine ou un robot. *A contrario*, aucun auteur n'affirme que son article a bien été rédigé par un être humain. L'effet rhétorique est ici considéré comme plus impactant en raison de l'utilisation d'une négation. Cela peut également être interprété comme un moyen, pour ces auteurs, de se mettre en scène dans la construction du sens. Ce faisant, le journalisme est considéré – une fois pour toutes – comme une activité humaine. Il s'agit donc là d'un enga-

21). Il en va de même pour le volet anglophone (20 contre 17). Bien que cette dernière analyse témoigne d'attitudes partagées, elle confirme certaines observations précédentes. Toutefois, elle détecte une dominante de sentiments négatifs dans le corpus en langue anglaise, à l'inverse de ce qui avait été constaté lors de l'analyse linguistique mais, ici, la méthode est axée sur le niveau lexical et ne tient donc pas compte des niveaux syntaxiques et pragmatiques qui peuvent également influencer sur la sémantique.

1.4.3 Mise en perspective

Les résultats de cette analyse du discours médiatique des journalistes montrent que les prises de position, en tant que formes sociales articulées linguistiquement dont le sens doit être interprété dans le cadre plus large du langage (Du Bois 2007), ne dominent pas la manière dont les journalistes abordent un sujet qui leur est proche. Néanmoins, l'usage répandu de la métaphore du robot ne permet pas de conclure que les journalistes restent formellement neutres vis-à-vis du phénomène, étant donné sa force représentative. La locution "robot journalisme" (ou "robot journaliste") est passée dans la langue pour désigner les technologies d'automatisation de la production d'informations. Mais si les journalistes utilisent moins les termes spécifiques relatifs au domaine d'application, cela peut également s'expliquer parce qu'ils s'adressent à un public large et/ou à de non-experts, induisant l'usage d'un langage vulgarisé. Cela pourrait être tout autant lié à la nécessité de "rendre concret" un processus abstrait. Lorsque des enjeux relatifs au phénomène sont abordés, et ils le sont essentiellement dans un contexte général (thématique). Le corpus en langue anglaise place l'accent sur la question de l'avenir du journalisme, tandis que le corpus en langue française fait émerger des inquiétudes relatives à l'emploi et à l'identité professionnelle des journalistes.

Les différences entre le corpus en langue française et celui en langue anglaise peuvent s'expliquer par les polarités opposées observées au sein des structures médiatiques en France et aux États-Unis (Christin 2016). Elles peuvent également être liées à deux traditions particulières du journalisme : contrairement aux journalistes francophones, les journalistes anglophones sont plus éduqués à tracer une ligne nette entre les faits et les commentaires (Chalaby 1999). En ce sens, ils sont moins susceptibles de remettre en cause la notion d'objectivité, pourtant considérée comme une norme, mais qui reste difficile à définir (Schudson 2001, Wien 2005, Cornu 2009). Au-delà de ces différences sociales et culturelles, les incertitudes observées au sein du corpus en langue française pourraient être liées au fait que le phénomène y soit plus récent. Les auteurs francophones semblent avoir le besoin de proclamer leur identité professionnelle et leur peur de perdre leur emploi, dans un contexte économique particulièrement difficile. Les médias traversent une crise profonde, qui n'est pas sans conséquences pour les journalistes (Leteinturier 2016, Deuze 2016). Bien que cette situation ne se limite pas au seul espace de la francophonie, on pourrait considérer qu'elle y soit plus durement ressentie. Le phénomène de l'automatisation de la production d'information ajouterait une couche supplémentaire aux préoccupations liées à l'avenir de l'emploi dans un secteur fragile. De plus, si le corpus en langue française présente une certaine homogénéité dans les résultats, cela peut s'expliquer par le fait que la France y soit le pays le plus représenté. Si des connotations po-

sitives sont plus souvent observées dans le corpus en langue anglaise, cela pourrait être – en partie – lié au fait que la question d'un avenir associé à la notion de progrès y soit plus dominante, en particulier aux États-Unis, terrain fertile pour les innovations technologiques. La question de l'éthique journalistique est davantage abordée dans le corpus en langue anglaise, notamment dans les publications américaines où les technologies de l'information automatisées sont développées depuis plus longtemps qu'ailleurs. Les différences observées au sein du corpus en langue anglaise peuvent également être liées au fait que les principaux échantillons de titres proviennent des États-Unis et du Royaume-Uni, où la taille des scènes médiatiques ne peut être comparée, et où les contextes socioprofessionnels disposent de caractéristiques qui leur sont propres, malgré des schémas communs (Deuze 2002).

Du point de vue de la dimension cognitive de la culture du journalisme (Singer 2003), les journalistes sont amenés à reconsidérer leurs compétences en raison d'une machine perçue comme une concurrente. Vue sous sa dimension normative, des considérations éthiques apparaissent lorsque l'information n'a pas été produite par un être humain. Abordée sous l'angle de sa dimension évaluative, l'automatisation remet en cause l'identité des journalistes car le journalisme est bien plus qu'un ensemble de tâches ou routines. Le concept d'idéologie professionnelle (Deuze 2005) est également contesté lorsqu'il est considéré que les systèmes automatisés pourraient remplacer les journalistes. Bien que ces préoccupations soient moins présentes dans le corpus en langue anglaise, elles ne sont pas totalement occultées. Dans le même temps, dans la plupart des cas, les auteurs des titres marquent généralement une distance entre leurs énoncés et l'objet de leurs discours. Ils sont donc davantage liés aux faits plutôt qu'à leur propre engagement.

Cette analyse du discours médiatique des journalistes a permis de mieux comprendre la manière dont ceux-ci prennent position lorsqu'ils sont amenés à traiter un sujet qui leur est proche, ainsi qu'à définir une représentation plurielle caractérisée par autant de résistance que d'enthousiasme. À la lumière de ces résultats, nous ne pouvons pas parler d'un imaginaire mais bien d'imaginaires collectifs, oscillant entre des sentiments de peur (de perte d'emploi ou d'identité professionnelle) et de confiance (vis-à-vis du futur et du progrès). La métaphore du robot contribue à polariser les discours. Dans les médias traditionnels en langue française, son usage accentue le caractère négatif du phénomène, alors que les blogs et les pure players tirent la tendance positive vers le haut. À *contrario*, la métaphore est perçue de manière moins positive dans les blogs en langue anglaise, tandis qu'elle influe de la même manière sur les connotations positives et négatives des médias traditionnels en langue anglaise. Les analyses des similitudes (arbres maximaux) et de sentiment corréleront généralement avec ces observations. La métaphore du robot influe donc sur les représentations d'un phénomène perçu de manière ambivalente, à la fois source d'opportunités et de menaces. Bien qu'elle soit insuffisante pour couvrir l'éventail des nuances caractérisant l'imaginaire technologique des journalistes, cette analyse met également en évidence des préoccupations communes. Aussi, peut-il être admis qu'une représentation collective va émerger : celle d'une relation homme-machine placée sous tension. Ceci illustre le rapport dual qui caractérise la relation que les journalistes entretiennent avec les technologies en général, et qui est souligné dans la recherche depuis quatre décennies.

2 | Entre deux mondes : les dimensions culturelles de l'automatisation

Il n'y a aucune raison de supposer que les machines ont des limites qui ne sont pas partagées par l'homme.

Marvin Minsky (1969)

La culture professionnelle du journalisme s'exprime à travers trois dimensions complémentaires : cognitive, qui se rapporte à la connaissance des techniques utilisées par les professionnels et à la formation pour les maîtriser ; normative, relative aux standards et règles appliquées dans le cadre de l'activité journalistique ; et évaluative, correspondant à la manière dont les journalistes perçoivent leurs rôle et identité (Singer 2003). Ces représentations socio-culturelles sont, toutefois, difficiles à aborder : le monde du journalisme ne forme pas un tout qui serait homogène. Aussi, des différences importantes peuvent-elles être observées entre les rédactions et les agents sociaux qui les composent. Abordées sous l'angle du contexte du développement d'innovations technologiques, celles-ci peuvent être considérées quant à la manière dont les journalistes envisagent leurs pratiques quotidiennes. .

Ce deuxième chapitre se fonde sur la seconde hypothèse de recherche, selon laquelle les usages journalistiques vont dépendre de l'adéquation des productions automatisées avec les savoir-faire et exigences du journalisme. Pour ce faire, il prend appui sur la définition des dimensions de la culture professionnelle du journalisme proposée par Singer, considérant que l'ensemble des acteurs participant à la chaîne de production de l'information participent à une mise en sens du monde social. L'accent est donc ici placé sur le double challenge journalistique et technologique qui sous-tend le phénomène de la production automatisée d'informations. Celui-ci va mobiliser des savoirs et savoir-faire spécifiques qui, s'ils présentent une filiation avec les pratiques du journalisme de données, sont également susceptibles de mettre en jeu des agents sociaux issus du monde de la technique.

La dimension cognitive de la culture du journalisme est abordée en se plaçant dans la perspective d'une approche par données devant rencontrer des exigences qui sont autant liées à la qualité des données, à la manière de les traiter et au sens journalistique qui sera apporté.

Plusieurs exemples de projets d'automatisation menés dans les médias d'informations vont mettre en avant des savoir-faire qui correspondent à la manière dont les journalistes travaillent habituellement, tout en soulignant les savoir-faire spécifiques mis en œuvre. Il s'agit également de comprendre en quoi la recherche traitant des bases de données informatisées peut nourrir utilement les pratiques journalistiques se fondant sur une approche par données. Dans une visée opérationnelle, cette première section propose un modèle conceptuel d'évaluation qui vise à rencontrer les challenges technique et journalistique, tant sur le plan de la qualité des données que sur celui de la pertinence de leur exploitation. Ce modèle, qui peut être envisagé en amont de la conception "sur mesure" d'un système d'automatisation de la production d'informations, a été éprouvé dans la première étude de cas présentée dans la deuxième partie de cette thèse¹.

L'examen de la dimension normative se rapporte à la manière dont le journaliste exerce sa responsabilité sociale envers les audiences auprès desquelles il est redevable. Considérant que les agents sociaux impliqués dans l'ensemble de ce processus font acte de journalisme, dès lors qu'ils participent à un processus journalistique, il s'agit également de s'intéresser à la manière dont les agents sociaux du monde de la technique envisagent l'exercice de leur responsabilité sociale. L'objet est d'analyser les points de convergence, de même que les points de divergence qui devraient être négociés en vue de fonder une culture commune. À la lumière de ces enseignements, cette deuxième section propose dix recommandations qui visent à encourager les bonnes pratiques lors de l'implémentation d'un système d'automatisation de la production d'informations dans une rédaction. Celles-ci sont à envisager de manière inclusive, dans le respect des standards journalistiques qui fondent la responsabilité sociale des médias d'information.

Dans une troisième et dernière section, l'examen de la dimension évaluative de la culture du journalisme s'attache à définir la manière dont les journalistes perçoivent les artefacts d'automatisation de la production d'informations, mais aussi celle dont les audiences appréhendent des contenus automatisés, dont elles ne connaissent pas toujours la nature "non humaine" de leurs auteurs.

Aussi, ce chapitre souligne-t-il la forte implication culturelle d'un processus d'automatisation de la production d'informations, qui n'est pas seulement journalistique mais aussi technique, dès lors que des agents non-humains sont embarqués dans un processus éditorial (Primo & Zago 2015). Mais il considère aussi qu'aucune technologie ne naît et ne se déploie par elle-même. Le code informatique que renferme un logiciel n'est rien de plus que la représentation d'une réalité construite à partir d'intentions humaines (Berry 2011). Celles-ci résultent également de pratiques culturelles (Geiger 2014) qui "*transportent l'espace social dans les réseaux logiciels*" (MacKenzie 2006). Avant d'être un objet technique, la technologie est donc, d'abord et avant tout, sociale (Flanagin *et al.* 2010). Dans cette perspective, il ne peut y avoir de séparation radicale entre la construction de l'objet technique et celle de sa construction sociale, qui

¹ Voir infra, p.159

est collective (Flichy 1991, cité par Pélissier & Romain 1998) et qui résulte d'un jeu d'associations entre agents sociaux (Bijker, Hugues & Pinch 1987, Flichy 1995).

Cette approche caractéristique des études des logiciels (*software studies*) fait écho au tournant matériel qui s'est engagé dans les études du journalisme dès le début des années 2000, où technique et social ont été étudiés sous l'angle de leurs influences réciproques (Boczkowski, 2004). Dans le contexte de la production automatisée d'informations, seuls les points d'entrée du système vont faire l'objet d'interventions humaines (Ekbia & Nardi 2014). Ces dernières s'inscrivent dans le cadre d'un processus qui va mettre en relief deux ressorts : celui des intentions, qui définiront le "quoi dire" ; et celui des médiations, qui définiront le "comment le dire"². Comme dans une réaction en chaîne, chaque étape donnera lieu à une nouvelle version de la représentation du réel. La représentation finale de l'information résultera de ces itérations où les choix posés seront relatifs (1) aux données qui vont nourrir le système en entrée, et qui fournissent une première représentation symbolique de l'information (Manovich 1999, Fuller *et al.* 2008) ; (2) aux opérations de quantification qui s'appuieront sur des interprétations humaines (Desrosières 2008, Flew *et al.* 2010, Espenland & Stevens 2008) et donneront un sens à l'information ; (3) aux résultats attendus du processus d'automatisation.

S'il peut être mis en parallèle avec un processus éditorial traditionnel, lui aussi caractérisé par une succession de choix (Gillespie 2014), ce processus peut également être envisagé comme le fruit d'un assemblage sociotechnique dont les composants peuvent être compris à travers les valeurs ou enjeux qu'ils véhiculent (Kitchin & Dodge 2011, Mackenzie 2006, Geiger 2014, McCarthy & Wright 2007). À la manière des poupées russes, la "boîte noire" des technologies d'automatisation de la production d'informations implique donc d'y imbriquer une "boîte noire" journalistique. Si des préoccupations communes peuvent émerger entre agents sociaux du monde du journalisme et de celui de la technique, en raison de normes et valeurs professionnelles distinctes, les points de vue pourront être différents à propos de la nature du journalisme et de ses processus (Lewis & Usher 2013). Rencontrer les savoir-faire et exigences du journalisme suppose donc un changement de paradigme, les nouveaux acteurs du monde de l'information devant désormais admettre qu'ils font acte de journalisme lorsqu'ils participent à la chaîne de production d'informations. Si cela peut commencer par le développement d'une pensée journalistique³, cela implique également de s'ouvrir davantage au monde du journalisme, en vue de souscrire à un cadre qui soit communément partagé : l'acte d'informer n'est pas anodin, et cela d'autant plus qu'il engage la responsabilité sociale (et morale) de l'ensemble de ses acteurs.

² Voir le processus théorique "en pipeline" de génération automatique de textes, p.68

³ Voir dans le chapitre précédent "Développer une pensée journalistique", p.76

2.1 Dimension cognitive

La dimension cognitive du journalisme est liée à l'ensemble de connaissances et des techniques que les journalistes professionnels mobilisent dans le cadre de leur travail (Chung & Nah 2014). Elle implique une formation professionnelle, en vue de maîtriser ces compétences (Singer 2003). Dans le contexte de la production automatisée d'informations, ces dernières peuvent relever de deux logiques : relatives aux données, la "cause matérielle" des contenus générés (Karlsen & Stavelin 2014), et relatives aux processus informatisés, supposant des compétences en matière de traitement automatique de la langue (TAL). Dans tous les cas de figure, cela va mobiliser des savoir-faire spécifiques qui sont autant liés à des pratiques techniques qu'à des pratiques journalistiques. Les questions relatives au choix et à l'évaluation des données seront constitutives d'un processus journalistique, tandis que celles relatives à la validation et à la normalisation des données appartiendront aux programmeurs (Hansen *et al.* 2017, Linden 2017). Si l'on considère que l'automatisation de la production d'informations participe au "*tournant quantitatif dans le journalisme*" (Coddington 2015), de même que si l'on tient compte de la nécessité de disposer de données fiables, exactes et précises pour générer des contenus fiables, exacts et précis, la dimension cognitive du journalisme va donc notamment recouvrir les compétences relatives à la collecte et au traitement des données.

Dans le contexte de la production automatisée d'informations, les données sont le matériau de base à partir duquel l'information est produite. Elles se trouvent au centre de processus créatifs où les contenus et leur interface sont séparés par leurs codes respectifs qui les mettent en forme. En tant que telles, les données sont passives : ce sont les algorithmes qui des rendent actives (Manovich 1999). La réalité que les données représentent ne participera à la construction du sens qu'à partir du moment où elles seront analysées et traitées. Ces activités peuvent être comprises comme un travail d'interprétation des phénomènes observés, lequel s'inscrit dans la perspective d'une sociologie de la quantification. Elles vont mobiliser des compétences spécifiques⁴ qui, sur un plan opérationnel, vont poser les jalons d'un processus éditorial remixé. Cela questionne le concept d'objectivité qui fait débat, de longue date, dans le monde du journalisme. Nous verrons également en quoi les théories de la science des données peuvent soutenir une démarche journalistique s'inscrivant dans le champ d'une approche par données. Les apports de ce domaine de connaissances scientifiques vont permettre d'élaborer un modèle conceptuel d'évaluation qui vise à relever le double challenge technique et journalistique. Celui-ci consiste en un appui à l'identification des problèmes susceptibles de se poser en amont d'un processus d'automatisation, en vue de les traiter et de les résoudre, sur un plan opérationnel.

⁴ Voir "Comprendre le fonctionnement de la machine", p.70

2.1.1 Les données, une cause matérielle

Aujourd'hui, les bases de données informatisées jouent un rôle structurant dans le contexte journalistique, où il peut être considéré que les faits sont "vrais" dès lors qu'ils existent dans une base de données (Anderson 2018 : 31-32). Bien qu'elles ne donnent pas "à voir" de manière complète une réalité donnée – leur représentation du réel observable sera donc toujours imparfaite (Boydens 1999) – les bases de données informatisées constituent une forme symbolique de cette réalité. En ce sens, elles sont "*une manière de structurer notre expérience sur nous-même et sur le monde*" (Manovich 1999). Aussi, une approche par données dans le journalisme ne peut-elle être abordée autrement que sous l'angle de la compréhension professionnelle de ce qu'est une base de données et de la manière dont celle-ci peut se déployer dans un contexte journalistique. Cela suppose une expertise qui dépasse largement des compétences journalistiques traditionnelles (Anderson 2018).

Les données sont devenues la matière première de l'information, que ce soit dans le cadre du traitement de l'actualité, d'un travail d'investigation ou du développement d'outils numériques dédiés à la mise en récit journalistique. Cette évolution des pratiques a entraîné dans son sillage de nouvelles collaborations entre des journalistes et des programmeurs, voire l'émergence de profils professionnels hybrides⁵. Cet essor, dans le monde du journalisme, est donc à la fois lié à des transformations techniques et sociales, attestant du fait que les bases de données "*échappent de plus en plus aux intérêts des informaticiens. Celles-ci sont aujourd'hui au cœur des pratiques et des représentations d'individus qui rencontrent cette forme sociotechnique, alors même qu'ils s'inscrivent dans les mondes sociaux les plus variés*" (Flichy & Parasie 2013).

Dans toute approche par données dans le journalisme, qu'elle soit automatisée ou non, la nécessité de disposer de données structurées sera une pré-condition pour une utilisation journalistique. Cela suppose de traiter la source des données avec scepticisme (Bradshaw 2013), et de pouvoir bien en identifier l'origine (Dörr & Hollnbuchner 2016). Si l'on ne peut faire confiance au producteur ou au diffuseur des données, le risque est que l'on ne puisse pas davantage faire confiance aux données et donc à l'information. Cet aspect est d'autant plus important que l'un des plus grands défis actuels, pour les médias, est précisément de restaurer un rapport de confiance avec leurs audiences (Fink 2019). Comme n'importe quelle autre source d'informations, les données doivent pouvoir être vérifiées : il s'agit là de satisfaire le principe journalistique de recherche de la vérité, lequel suppose des activités de vérification (Cornu 2009 :78).

Une information fiable et précise ne pourra s'appuyer que sur des données fiables et précises (Bradshaw 2015). C'est le principe du "garbage in, garbage out" largement reconnu dans le monde de l'informatique. La qualité de précision constituera "*probablement la considération éthique la plus fondamentale dans le journalisme de données. Elle consiste dans la nécessité d'être exact et dans celle de donner un contexte approprié aux histoires que nous racontons. Cela peut influencer la façon dont nous analysons les données et les récits s'appuyant sur des don-*

⁵ Voir "Comprendre le fonctionnement de la machine", p.70

nées"⁶. Dans le contexte de l'automatisation de la production d'informations, des données qui manquent de fiabilité et de précision pourront donner lieu à des contenus erronés (Dörr & Hollnbuchner 2016).

Il convient aussi de s'interroger sur la manière dont les données ont été récoltées ainsi que sur leur mise en contexte (Bradshaw 2015). Les métadonnées attachées à un jeu de données donneront, à ce propos, des indications importantes pour aider les utilisateurs à facilement évaluer et comprendre les données (Shanks 1999). Toutefois, celles-ci ne seront pas nécessairement attachées à un ensemble de données, ce qui signifie que celui-ci ne pourra pas toujours être bien compris et que certaines ambiguïtés liées à la nature des variables ou de leurs attributs sont susceptibles d'apparaître. De plus, la mise en disponibilité de données ne signifiera pas que celles-ci seront les plus pertinentes pour un usage journalistique, pas plus qu'elle ne garantira qu'elles soient exemptes d'anomalies dans leurs valeurs (Casswell & Dörr 2017). Pour toutes ces raisons, les données devraient d'abord faire l'objet d'une vérification humaine, et cela est d'autant plus important lorsque celles-ci sont destinées à nourrir un processus automatisé (Bradshaw 2015).

Il s'agit de bonnes pratiques qui témoignent de l'enjeu fondamental de la qualité de l'information numérique (Stray 2016). Cependant, celui-ci sera moins crucial lorsque les données proviendront de fournisseurs monnayant leurs services via des licences d'exploitation, qui en garantissent normalement la qualité et mobilisent la confiance de leurs utilisateurs. Par exemple, dans le secteur économique et financier, certains prestataires pratiquent une tarification progressive à mesure que l'on se rapprochera d'une fourniture de données en temps réel⁷. D'autres prestataires proposeront des tarifs progressifs, selon que leurs services incluent ou non un historique des données⁸. La valeur d'échange des données, souvent fournies à leurs utilisateurs finaux via des interfaces de programmation applicatives (API), explique que les professionnels de la vente de données n'affichent pas leurs tarifs de manière systématique sur leurs pages web : le client sera le plus souvent invité à solliciter un devis.

L'enjeu de la qualité des données reviendra au premier plan lorsque les données seront extraites de pages web, caractérisées par leur format potentiellement mouvant. Sont-elles à jour, y a-t-il une gestion des enregistrements et des versions? (Batini *et al.* 2009). Dans ce contexte, la question de la crédibilité de la source d'informations devra aussi être posée, en tant que norme professionnelle constitutive du processus de sélection d'informations (Hermans *et al.* 2009). Cet enjeu sera aussi prégnant dans le contexte des données publiques ouvertes, que l'on désignera ici sous le vocable *open data*. Si leur objet est de servir la transparence de l'action publique, dans de nombreux cas, elles se limiteront à des données qui ne comportent pas de

⁶ "Ethics in data journalism : accuracy", Paul Bradshaw, Online Journalism Blog, 13/09/2013, consulté le 08/07/2017, URL : <https://onlinejournalismblog.com/2013/09/13/ethics-in-data-journalism-accuracy/>

⁷ Ce cas de figure sera constaté dans le quatrième chapitre, relatif à l'automatisation de l'information boursière pour les journaux du groupe Mediafin. Voir *infra* p.208

⁸ Par exemple, le site Polygon propose une fourchette d'abonnement oscillant entre 49 et 399 dollars par mois. Source : <https://polygon.io/pricing>

dimension politiquement sensible : la transparence de l'action publique ne serait donc pas nécessairement l'effet recherché par la libération des données publiques (Goëta & Mabi 2014). De plus, leur qualité devrait être considérée comme potentiellement douteuse, "*dans la mesure où celles-ci prolifèrent dans l'environnement ouvert et non contrôlé du web*" (Boydens 2014).

Considéré comme une opportunité pour les journalistes, le mouvement de l'*open data* n'a pas rencontré ses promesses dans le monde du journalisme de données – où le phénomène a été surestimé (Gray 2014) – et cela essentiellement pour trois raisons : (1) la non-prise en compte des journalistes en tant qu'utilisateurs finaux dans les politiques d'ouverture de données (Stoneman 2015, Goëta & Mabi 2014); (2) le manque de pertinence pour conduire un projet journalistique (Linden 2017b); (3) la nature des données, souvent obsolètes ou incomplètes (Stoneman 2015). De plus, la dépendance aux sources de données publiques influencera les sujets susceptibles d'être couverts. En conséquence, certains domaines sociaux pour lesquels les données ne sont pas produites ou accessibles régulièrement pourraient être négligés (Loosen 2018).

Sur le fond, il convient de garder à l'esprit que le mouvement de l'*open data* ne s'inscrit pas dans la perspective d'un outil de preuve relevant de l'argument statistique (Desrosières 2008), mais dans celle d'un dispositif de diffusion des connaissances (Courmont 2015). Considérant qu'une donnée est le résultat d'un réseau sociotechnique, il est dès lors "*impossible de dissocier la donnée de son usage, de détacher l'instrument technique de l'environnement social dans lequel il s'inscrit*" (Courmont 2015). Cette assertion renvoie aux modèles de gouvernance des données et de leur corollaire, des politiques pertinentes et adaptées en matière d'*open data*. Si la figure du journaliste est peu – voire pas du tout – mise en avant dans les politiques d'ouverture des données, la figure du développeur, elle, est omniprésente : c'est vers lui que sont destinés les encouragements à la réutilisation des données, via le développement d'applications mobiles ou en ligne. L'un des principaux défis serait donc, aujourd'hui, de proposer aux journalistes des outils simples, de manière à leur permettre d'étudier les données (Colpaert *et al.* 2013).

Dans le contexte de la production automatisée d'informations, le manque de qualité des données accessibles en *open data* ne permet pas toujours de réaliser des objectifs de précision et d'actualité (Graefe 2016). S'il s'agit d'un frein à leur usage, cela ne signifie pas, pour autant, que les données publiques ouvertes soient totalement exclues du champ des possibles : plusieurs expériences journalistiques en attestent. Dans la majorité des cas, elles prennent la forme d'applications automatisées, pouvant être considérées comme "*des fenêtres sur les données d'une histoire*" (Stray *et al.* 2013). Leur objet est de faire gagner du temps aux journalistes dans le traitement de l'information, d'élargir la zone de couverture médiatique du média ou de proposer un service supplémentaire aux lecteurs. Dans certains cas, elles porteront sur une collaboration active entre les journalistes et le système d'automatisation. Les sept projets exposés ci-après illustrent ces possibilités.

- **L.A. Crime Maps** – Développée en 2007 pour le *Los Angeles Times*, il s'agit d'une carte interactive qui présente une information instantanée sur les homicides commis dans la ville. Les données publiques fournissent un contenu de base qui est susceptible d'être enrichi par les journalistes. Ce dispositif les alerte également à chaque nouvel homicide commis (Hermida & Young 2015). Chaque nouvelle entrée est rédigée de manière automatique avant d'être publiée en ligne. Elle mentionne la date, l'heure, l'emplacement de l'incident et l'origine ethnique de la victime. Pour la rédaction, il s'agit d'un outil permettant de renforcer son rôle. Pour Ken Schwenke, son auteur, il s'agit d'un outil simple dont l'impact est disproportionné par rapport à la quantité de temps consacré au projet (Hermida & Young 2015). Avant cette automatisation, le *L.A. Times* couvrait seulement 10% des crimes commis dans la région (Graefe 2016). Ce projet a inspiré le *Chicago Tribune* pour un projet similaire dans la ville de Chicago⁹.

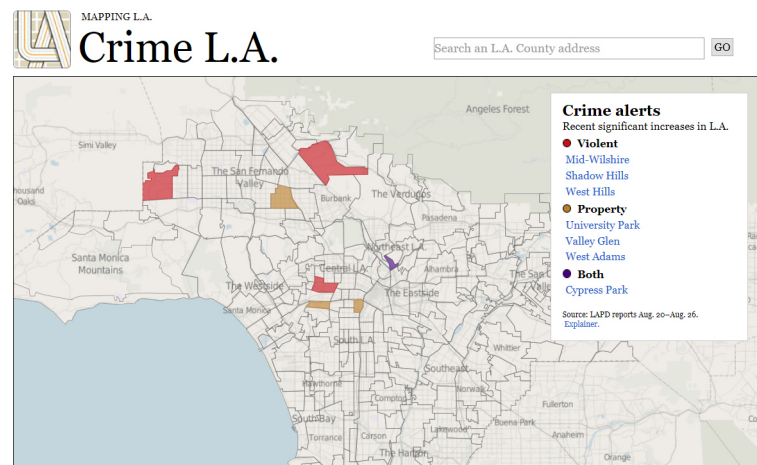


FIGURE 2.1 – L'application "Crime L.A." du *L.A. Times*

- **Quakebot** – Cette application consiste en un programme informatique développé, en 2014, par le journaliste-programmeur Ken Schwenke. S'appuyant sur les données géologiques du US Geological Geological Survey, elle alerte en temps réel sur les secousses sismiques enregistrées dans la région. Chaque texte généré de manière automatique fait l'objet d'une validation humaine avant sa diffusion sur le site web du *Los Angeles Times* (Antheaume 2016). En donnant l'alerte en temps réel, ce système permet d'accélérer le processus de publication de l'information en ligne, dans un format standardisé. Il a été construit à partir du même algorithme que celui de la carte des homicides à Los Angeles (*lire ci-dessus*). Graefe (2016) indique que "le travail de Schwenke marque une étape importante dans l'ère du journalisme automatisé, démontrant combien les solutions internes simples peuvent aider à accroître à la fois la rapidité et la zone de couverture des informations".

⁹ "This application is based on Homicide Report, created for the Los Angeles Times by Ken Schwencke. It was adapted for use in Chicago", consulté le 14/09/2017, URL : <http://homicides.redeyechicago.com/neighborhood/bridgeport/>

MARCH 27, 2015, 9:44 AM

A shallow magnitude 3.6 earthquake was reported Friday morning 14 miles south of Twentynine Palms in Joshua Tree National Park, according to the U.S. Geological Survey. The temblor occurred at 8:36 a.m. Pacific time at a depth of 1.2 mile.

According to the USGS, the epicenter was 17 miles northeast of Indio, 19 miles north-northeast of Coachella and 24 miles northeast of Palm Desert.

In the last 10 days, there have been no earthquakes of magnitude 3.0 or greater centered nearby.

This information comes from the USGS Earthquake Notification Service and this post was created by an algorithm written by the author.

FIGURE 2.2 – L'application "Quakebot" du *L.A. Times*

- **Die Feinstaub-Monitor** – Le "moniteur des particules fines" du *Berliner Morgenpost* traite de la qualité de l'air berlinois. Développé en interne, ce système automatisé est capable de comparer des données actuelles avec des données antérieures (jusqu'à 2008) en vue d'adapter les titre, texte et visualisations interactives¹⁰. Les données publiées sont celles de la veille, et elles sont fournies par le ministère du développement urbain et de l'environnement. Cet objet du journalisme s'inscrit dans la logique d'une information servicielle qui se fonde sur l'impact sanitaire de ce type de polluant¹¹.

The screenshot shows the Berliner Morgenpost website interface. At the top, there is a navigation bar with 'Anmelden', the logo 'Berliner Morgenpost', and 'Suche'. Below the navigation bar, there are links for 'Home', 'Berlin', 'Bezirke', 'Interaktiv', 'Politik', 'Wirtschaft', 'Sport', 'Panorama', 'Kultur', 'Wissen', 'Reise', 'Lifestyle', 'Abo', 'Specials', and 'Service'. A news ticker at the bottom of the navigation bar reads 'In den Nachrichten: Bundestagswahl | Air Berlin | Flughafen Tegel | Schulserie | Unter 20 Euro | Alle Themen'. Below the navigation bar, there are social media sharing buttons for Twitter, Facebook, and LinkedIn. The main content area features the article title 'Alle Berliner Messstationen unter EU-Jahresgrenzwert' with a sub-headline '10.09.2017 Feinstaub-Monitor'. The article text states: 'Berlin zählt zu den deutschen Städten, in denen die höchsten Feinstaub-Konzentrationen gemessen werden. 2016 wurde der von der EU vorgegebene Jahres-Grenzwert in der Hauptstadt aber noch an keinem Ort überschritten.' To the right of the text is a map of Berlin with several black dots indicating measurement stations. A legend indicates that dots represent 'Messstation unter Tageslimit' and red dots represent 'über Tageslimit'. At the bottom of the article, there is a timestamp: '+++ Letztes automatisches Update: 10.09.2017, 14:00 Uhr +++'.

FIGURE 2.3 – L'application "Fenstaub Monitor" du *Berliner Morgenpost*

- **Données du Monde** – Cette plateforme interactive est le fruit d'un partenariat entre le quotidien *Le Monde* et la start-up française Syllabs, spécialisée dans les technologies sémantiques. Elle s'inscrit dans le prolongement de l'automatisation des résultats

¹⁰ "Le journalisme : vers une intelligence artificielle?", Sophie Roche, Future.Arte.tv, 03/03/2017, consulté le 20/06/2017, URL : <http://future.arte.tv/fr/le-journalisme-vers-une-intelligence-artificielle>

¹¹ "Alle Berliner Messstationen unter EU-Jahresgrenzwert", *Berliner Morgenpost*, consulté le 20/06/2017, URL : <https://www.morgenpost.de/berlin/article127420411/Wo-die-Feinstaub-Belastung-in-Berlin-am-hoehsten-ist.html>

des élections régionales de 2015¹², bien que la génération automatique de textes n'en constitue que l'un des aspects : l'application propose également des cartes et des graphiques. Pour ce faire, elle s'appuie sur des données électorales fournies par le ministère français de l'intérieur, et des données démographiques et économiques provenant de l'Institut national de la statistique et des études économiques (INSEE)¹³.



FIGURE 2.4 – La plateforme en ligne "Données du Monde"

— **JSA Figures** – Ce dispositif génère, chaque mois, du texte et des graphiques relatifs au taux de chômage sur l'île de Wight (Royaume-Uni) pour le *pure player* *On The Wight*, à partir des de la caisse d'allocations de chômage (*Jobseekers' Allowance*). L'objet de ce système est de décharger les journalistes d'un travail formel et, partant, de se consacrer à d'autres types d'informations hyperlocales. L'initiative émane des journalistes, qui ont sollicité un développeur. Les contenus générés ne sont pas publiés tels quels : ils passent d'abord entre les mains d'un éditeur humain qui en vérifie la correction et la complétude, avant de procéder à la mise en ligne. Le seul bémol résiderait dans le temps de développement du générateur, qui s'est avéré "*bien plus long que la rédaction d'un premier article*"¹⁴. Pour le développeur du programme, Tony Hirst, cette approche *data-to-text* permet de "*donner un sens humain aux données*"¹⁵.

¹² Il s'agit de la première expérience d'automatisation de la production d'informations menée, à grande échelle dans un média français : en l'espace d'une nuit, plus d'un million d'articles, mises à jour comprises, avaient été générés couvrant les résultats des votes dans plus de 30.000 communes et 2.000 cantons français. Pour le directeur des rédactions, Luc Bronner, il s'agissait "*d'expérimenter de nouveaux outils susceptibles d'apporter un nouveau service à nos lecteurs*", mais aussi de rendre ces textes "*plus repérables par les moteurs de recherche*". Source : "Des robots au Monde pendant les élections départementales? Oui... et non", in "Back-Office, les coulisses du Monde.fr", 23/03/2015, consulté le 20/04/2019, URL : <http://makingof.blog.lemonde.fr/2015/03/23/des-robots-au-monde-pendant-les-elections-departementales-oui-et-non/>

¹³ "Le Monde pousse sa data éditoriale", Thierry Wojciak, 14/10/2015, consulté le 20/05/2017, URL : <http://www.cbnews.fr/nl-Media/le-monde-et-la-data-editoriale-a1023121>

¹⁴ "Isle of Wight innovates in a new area of Journalism", Simon Perry, *On The Wight*, 07/09/2015, consulté le 20/06/2017, URL : <https://onthewight.com/isle-of-wight-innovates-in-new-area-of-journalism/>

¹⁵ "Data Textualisation – Making Human Readable Sense of Data", 18/11/2013, consulté le 20/06/2017, URL : <https://blog.ouseful.info/2013/11/18/data-textualisation-making-human-readable-sense-of-data/>

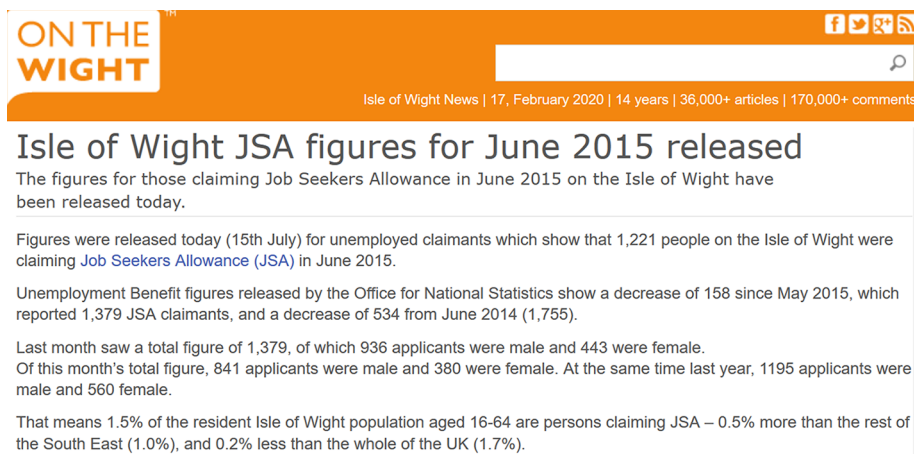


FIGURE 2.5 – "JSA Figures", la rubrique automatisée de *On The Wight*

- **RADAR (Reporters and Data and Robots)** – Le projet RADAR consiste en un partenariat entre l'agence de presse britannique Press Association et la start-up technologique Urbs Media. Lancé dans une version bêta en juin 2018, il vise à produire de manière automatique jusqu'à 30.000 textes chaque mois, à une échelle locale, en utilisant des données publiques ouvertes. Les sujets traités sont susceptibles de couvrir une large palette de domaines : du taux de natalité au taux d'obésité, en passant par le taux de criminalité. Son approche consiste à associer des journalistes aux processus automatisés.



FIGURE 2.6 – Premières productions de "RADAR" publiées dans la presse britannique

À ses débuts, l'équipe éditoriale était composée de six journalistes. Ce projet a été financé par le Google Digital News Initiative, un fonds destiné à soutenir l'innovation dans les médias européens, à hauteur de 700.000 euros¹⁶. Il ne s'agit pas du premier

¹⁶ Sources : "Robots can save local journalism, but will they make it more biased?", Alexander Fanta, DataJournalism.com, 16/03/2018, consulté le 13/07/2019, URL : <https://datajournalism.com/read/longreads/robots-can-save-local-journalism-but-will-they-make-it-more-biased/>; "PA's localised data journalism service Radar secures first paid customers", John McCarthy, *The Drum*, consulté le 13/07/2019, URL : <https://www.thedrum.com/news/2019/04/09/pa-s-localised-data-journalism-service-radar-secures-first-paid-customers>; et fiche de présentation du projet sur le site

projet d'automatisation s'appuyant sur de l'*open data* soutenu par le fonds Google : en 2016, il soutenait "Marple", un projet lancé par la start-up Journalism++¹⁷. En 2019, RADAR est devenue une agence de presse locale, accessible par abonnement, s'adressant aux éditeurs de presse locale (Linden *et al.* 2019).

- **Tobi** – S'il porte le nom d'un robot dont le graphisme affiche une mine sympathique, ce projet a été réalisé avec le logiciel de génération automatique WordSmith, commercialisé par la société américaine Automated Insights. Tobi est transversal au groupe de presse Tamedia, qui édite plus de trente quotidiens en Suisse (dont *20 minutes* et la *Tribune de Genève*). Il génère, de manière automatique, les résultats des votations fédérales (sortes de consultations populaires) en fonction de la commune de résidence de l'internaute – la Suisse en compte 22.000. Lors de sa première mise en service, le 25 novembre 2018, il a généré environ 40.000 textes personnalisés en moins de cinq minutes. Le logiciel a été configuré par une équipe de cinq journalistes politiques expérimentés, ce qui représente deux à trois jours de travail par personne. Un journaliste de données y a apporté une couche d'analyse supplémentaire, en vue d'apporter des informations de contexte (Plattner & Orel 2019). Le système, multilingue, est nourri par des données publiques.



Tobi a écrit cet article pour vous:

Locarno veut davantage de vaches à cornes et de surveillance des assurés

La ville refuse l'initiative populaire de l'UDC sur les juges étrangers. Avec 57% de non, une majorité souhaite une Suisse ouverte. Balayé au niveau national, le texte des démocrates du centre visait à instaurer la primauté du droit suisse sur le droit international. Les Locarnais disent non comme le Tessin, qui refuse l'initiative avec 54% de non.

La commune demande une base légale claire pour la surveillance des bénéficiaires des assurances sociales. Sur l'ensemble des 3575 citoyens ayant pris part au vote, 2383 soutiennent ce texte. Ces nouvelles mesures sont donc approuvées à 67% (70% de oui dans le canton). Les Suisses disent oui à 64,7%.



FIGURE 2.7 – Premières productions de "Tobi" publiées sur les sites du groupe Tamedia

Ces expériences témoignent de la mobilisation d'un savoir-faire correspondant à la manière dont les journalistes travaillent habituellement : choisir un sujet, choisir un angle, rechercher de sources fiables et vérifiables, choisir la manière dont sera traitée l'information : ces choix éditoriaux existent depuis que la presse existe (Gillespie 2013). Bien qu'elle fasse l'objet de controverses, la définition du travail journalistique présuppose une division particulière et une

consacré aux Data Journalism Awards, URL : <https://datajournalismawards.org/projects/radar-data-journalism-and-automation-to-produce-local-news-at-scale/>

¹⁷ "Journalism++ Stockholm gets Google funding to develop automated news service", 06/04/2016, consulté le 13/07/2019, URL : <http://jplusplus.org/sv/blog/journalism-stockholm-gets-google-funding-to-develop-automated-news-service/>

structuration du travail journalistique – incluant des activités de collecte et de recherche d'informations et des activités de sélection et de présentation de l'information – (Örnebring 2008) qui n'apparaît pas abordée de manière très différente à l'examen des différents projets présentés ci-dessus.

Les aspects relatifs à la qualité des données ont été très peu abordés dans la recherche en journalisme, mais ils l'ont été davantage au milieu des années 2000, où il a été notamment souligné le manque d'expertise des journalistes pour eux reconnaître des données de qualité médiocre et en améliorer la qualité (Lowrey 2019). De nos jours, les journalistes de données n'ont pas plus tendance s'intéresser à ces aspects plus "profonds" : ils exprimeront davantage leurs critiques à propos de l'*open data* et de la mise à disposition de données gouvernementales, et auront souvent tendance à considérer que des données émanant d'institutions publiques sont fiables. Cela suggère que les professionnels de l'information développent des routines plus strictes, dès lors qu'ils s'inscrivent dans le cadre d'une approche par données (Lowrey 2019).

Les processus de production automatisée d'informations, qui restent d'abord et avant tout des processus éditoriaux, impliquent des agents humains et donc des jugements de valeurs (Kraemer *et al.* 2011). Cela étant, des exigences seront posées en matière de qualité des données, supposant que des compétences spécifiques soient mobilisées pour traiter cet aspect bien plus complexe qu'il n'y paraît. Le champ de la science des données va permettre de mieux comprendre les enjeux liés à ce concept. Son examen va également mettre en relief, sur un plan théorique, des préoccupations partagées avec le monde du journalisme. Considérant que les journalistes sont de plus en plus confrontés aux données dans le cadre de leurs pratiques professionnelles, la compréhension de ce concept permet d'être mieux armé pour identifier les problèmes éventuels, et cela en vue de les résoudre¹⁸.

2.1.2 Gérer la qualité des données

Le concept de qualité est difficile à définir. Dans l'esprit de la norme ISO 9000, relative au management de la qualité, la qualité des données peut être comprise comme se rapportant à "*des données qui s'adaptent aux usages des consommateurs de données*" (Wang & Strong 1996). Aussi, la qualité d'une donnée ne peut-elle être envisagée sous le seul prisme formel du triplet "entité, attribut, valeur", issu du vocabulaire des bases de données informatiques (Redman 1996), selon lequel une donnée de qualité serait liée à une entité (par exemple, une voiture), un attribut (par exemple, la couleur) et une valeur (par exemple, rouge).

Pour Fürber et Hepp (2010 :144), la qualité des données peut être déterminée par la comparaison de l'état actuel des données à son état désiré. *A contrario*, une mauvaise qualité de données caractérise la non-conformité entre la représentation des données et le réel observable (Wand & Wang 1996). Juran (1999) distingue le concept de qualité de celui du symptôme de non-qualité : "*la qualité signifie l'absence de déficience (...) Un défaut est un état d'inaptitude à*

¹⁸ "The Quartz guide to bad data", Quartz, 2016, consulté le 12/01/2017, URL : <https://github.com/Quartz/bad-data-guide>

l'utilisation (...) Un symptôme est la preuve extérieure d'un défaut (...) Une théorie est une affirmation non prouvée quant aux raisons de l'existence de défauts et de symptômes".

La littérature scientifique propose également plusieurs définitions pour caractériser ce qu'est une donnée. Elle peut être considérée comme le matériau de base (*input*) à partir duquel l'information est développée (*output*) (Dorn 1981, cité par Fox *et al.* 1994). Elle est aussi un jeu de faits, un état des choses, qui résulte d'observations (Yovits 1981 & Dorn 1981, cités par Fox *et al.* 1994). Si une base de données informatique constitue une modélisation du monde, les données en présentent une vue abstraite. Cependant, toutes les données ne sont pas le résultat d'observations mesurées, certaines – comme le nom ou le numéro de sécurité sociale – sont assignées (Fox *et al.* 1994). Redman (1996) propose de considérer une donnée comme un symbole, un substitut aux objets. L'information consiste en un processus de données donnant lieu à n'importe quel type de connaissance : c'est la partie non-redondante d'un message, le résumé pertinent des données (Redman 1996). À l'inverse de l'entropie – principe défini dans la théorie de l'information de Shannon, qui consiste en une formule mathématique pour déterminer la quantité d'informations contenues ou délivrées dans une source d'information –, "*c'est la partie du message qui informe*". Une donnée, c'est aussi un objet-frontière entre des programmeurs, des designers, des journalistes et, de manière plus générale, entre tous les acteurs qui participent au processus de production de l'information (Sandoval-Martín & La-Rosa 2018).

Le concept de qualité des données peut être abordé dans ses aspects multidimensionnels, lesquels désignent un ensemble d'attributs qui le caractérisent (Wang & Strong 1996). Dans la recherche, ces dimensions ont pu être interprétées de diverses manières.

- **Brodie (1980)** – La qualité mesure ce en quoi la base de données représente de manière précise les propriétés essentielles et attendues de l'application. La qualité des données comporte trois composantes : la fiabilité des données, l'intégrité sémantique et l'intégrité physique.
- **Yu *et al.* (1990)** – La qualité des données revêt trois dimensions intrinsèques : (1) l'exactitude, qui correspond à la correction des données (à ne pas confondre avec la précision, qui est une mesure du degré de reproductibilité) ; (2) la complétude, principe selon lequel un jeu de données est complet à partir du moment où il contient toutes les données pertinentes ; (3) la consistance, définie par la représentation des données dans un format cohérent.
- **Fox *et al.* (1994)** – Le concept de qualité des données recouvre quatre catégories ou dimensions : l'exactitude, l'actualité, la complétude et la consistance. La dimension d'exactitude regroupe les caractéristiques de précision et de fiabilité ; la dimension d'actualité se rapporte à la date de mise à jour de la base de données ; la dimension de complétude indique le degré auquel les attributs d'un jeu de données disposent de la valeur qu'ils sont supposés avoir ; la dimension de la consistance est souvent assimilée aux

dimensions d'exactitude et de correction, mais elle concerne davantage le respect des contraintes imposées par la base de données.

- **Wand & Wang (1996)** – Plusieurs dimensions caractérisant la qualité des données sont identifiées par les chercheurs : exactitude et précision (*a contrario*, l'inexactitude implique que le système d'information représente un état du monde réel différent de celui qui aurait dû être représenté), fiabilité (dimension liée à la probabilité de prévenir des erreurs ou des échecs, ainsi qu'à la cohérence et à la fiabilité des informations en *output*), rapidité et actualité (dimension liée à la fréquence de mise à jour du système au regard du caractère volatile du monde réel), exhaustivité (toutes les valeurs sont enregistrées dans le système), cohérence (à la fois dans la représentation des données et dans leurs valeurs). Les chercheurs identifient quatre dimensions intrinsèques qui correspondent à quatre types de problèmes de qualité : données incomplètes, données ambiguës, données vides de sens, données incorrectes.

- **Strong et al. (1997)** – Des données de haute qualité rencontrent quatre types de caractéristiques : intrinsèques (exactes, objectives, crédibles), accessibles, contextuelles (données pertinentes, à valeur ajoutée, actuelles et complètes), et représentationnelles (données interprétables, faciles à comprendre, représentation concise et consistante). Wang et al. (1996) précisent, dans un article antérieur, que la dimension "intrinsèque" fait également référence à l'objectivité et à la réputation. La dimension représentationnelle est liée au format des données ainsi qu'à la signification des données.

- **Shanks (1999)** – La notion de qualité est organisée en quatre niveaux sémiotiques : syntaxique (objectif de correction et de forme, porte sur la structure de la donnée), sémantique (objectif de complétude et de validité, concerne la signification de la donnée), pragmatique (objectif d'utilisabilité et d'utilité, a pour objet l'usage de la donnée), sociale (objectif de partage et de compréhension, c'est le savoir partagé via la donnée). Cet axe sémiotique a été étudié par Teejay et al. (2005). Il inclut un niveau empirique constitué des dimensions d'accessibilité et de temporalité. Le niveau syntaxique tient compte de critères tels que la précision, la consistance et l'intégrité; le sémantique comporte des dimensions de non-redondance, de crédibilité, de fiabilité et de non-ambiguïté; le pragmatique comprend notamment les dimensions de pertinence et de complétude.

- **Haug et al. (2009)** – Ces chercheurs retiennent trois dimensions, dans le contexte des ERP (*Enterprise resource planning*) : intrinsèques (données exhaustives, non ambiguës, exactes et qui font sens), accessibles (en termes de droits d'accès et de stockage des données), utiles (pertinence et valeur ajoutée des données).

Ces différentes approches témoignent de la complexité à définir ce qu'est une donnée de qualité. Des considérations d'ordre formel (syntaxe, format, structure) mais aussi empirique doivent être prises en compte, en fonction du domaine d'application des données (Eckerson 2002).

Elles démontrent également l'impossibilité d'une approche axée sur la qualité totale : d'une part, il n'existe pas de référentiel absolu pour vérifier la correction de l'information répertoriée dans une base de données (Boydens & van Hooland 2011) et, d'autre part, la qualité des données va aussi dépendre de la qualité des processus (Moody *et al.* 2003, Wang *et al.* 2006, Batini *et al.* 2009).

La définition de la qualité des données peut être envisagée selon deux types d'approches aux fondements opposés : positiviste, qui part du principe qu'une base de données représente le réel observable; et constructiviste, qui argue que les données se construisent dans le temps (Boydens 2012). L'approche positiviste repose sur "*l'hypothèse de l'existence d'un isomorphisme entre les bases de données et le réel observable que celles-ci représentent : une donnée est jugée correcte si elle correspond 'parfaitement' à la réalité observable représentée*" (Boydens 2012). Toutefois, cela ne tient pas compte du caractère évolutif des données empiriques, dans la mesure où la représentation d'une réalité mouvante implique un *work in progress* permanent. Une base de données fonctionne sur l'hypothèse d'un monde clos, en vertu de laquelle toute valeur non incluse dans son domaine de définition est considérée comme fautive (Boydens 2011, 2012). Les anomalies y sont moins le fait d'erreurs formelles que de l'évolution des données dans le temps : il n'y a donc pas d'isomorphisme entre une base de données et le réel observable, les anomalies étant souvent liées à l'évolution du réel par rapport aux valeurs attendues (Boydens 1999). Comme autant de photos instantanées, les valeurs contenues dans les bases de données empiriques ne reflètent jamais que le monde réel tel qu'il a été observé à un instant T.

L'association mondiale des éditeurs de journaux (WAN-IFRA) a commandé à des acteurs académiques un rapport destiné à dresser un état des lieux de l'automatisation de la production d'informations journalistiques (Linden *et al.* 2019 :35-37). Celui-ci aborde la question de la qualité des données sous l'angle des "5 V de la qualité des données" décrits par Shutaharan (2014) et par Wanba *et al.* (2015) : volume, vitesse, variété, valeur et véricité. Soulignant que "*la qualité de la production d'un système journalistique automatisé dépend fortement de la disponibilité et de la qualité des données qui y sont introduites*", il détaille ces différentes dimensions :

- **Volume et vitesse** – Se réfèrent à la quantité de données disponibles et à la rapidité à laquelle celle-ci devient disponible. Des données de faible vitesse seraient des données relatives à des résultats électoraux, tandis que des données à forte vitesse seraient des données relatives aux relevés des températures dans une ville donnée. Les données dont le volume et la vitesse sont faibles donneront lieu à des générations plus courtes.
- **Variété et valeur** – Se rapportent à la complexité d'un jeu de données et aux types de données. Plus la complexité est élevée, moins les données seront automatisables.
- **Véricité** – Tout projet d'automatisation de la production d'informations doit tenir compte de cette dimension relative à la fiabilité et l'exactitude des données.

Les causes de la non-qualité de données sont nombreuses et peuvent relever d'une variété de paramètres : de la modélisation du système d'information (déterminant majeur pour les coûts de développement, de la flexibilité du système et sa capacité à rencontrer les exigences des utilisateurs) (Moody & Shanks 2003), du respect des contraintes d'intégrité de la base de données relationnelle (Fox *et al.* 1994), ou encore de la sémantique des données, dont l'hétérogénéité peut donner lieu à des problèmes d'interprétation (Madnick & Zhu 2006). Les problèmes de qualité se produisent souvent en raison d'une modification des usages, d'un changement de perception ou de violations des contraintes d'intégrité de la base de données, lesquelles ont pour fonction de préserver la consistance des données et leur précision (Chiang & Miller 2008). Une mauvaise qualité des données peut également découler d'un manque de routine dans les validations, ainsi que d'un pauvre design du système d'information et d'erreurs dans la conversion des données (Eckerson 2002). De plus, le choix de la représentation des données constitue l'une des tâches "*les plus cruciales dans le développement d'un système d'information*" (Madnick & Zhu 2006). Il est donc important de s'assurer que le modèle de données soit sans défauts¹⁹.

Les problèmes de qualité des données ne sont pas anodins, en particulier lorsque les données sont utilisées comme matière première pour les contenus d'actualité. Comme l'a souligné Essa (cité par Flew *et al.* 2010), journalisme et technologies de l'information sont concernés par la qualité et la fiabilité de l'information. De même que des données de mauvaise qualité ne donneront pas lieu à une information de qualité, si l'on ne peut faire confiance aux données, on ne peut pas davantage faire confiance à l'information délivrée (Haug *et al.* 2011).

Dans le monde du journalisme, la "qualité de l'information" est un concept qui renverra moins aux données qu'à la manière dont elles vont être traitées. Cela étant, il s'avère tout aussi difficile à définir. Selon Clerwall (2014), la qualité de l'information journalistique est un concept multifacette pouvant se rapporter "*à une notion générale, mais un peu vague, de 'ceci était un bon article'*". Elle peut également dépendre du contexte socioculturel du destinataire : Clerwall cite l'exemple du professeur de littérature qui peut apprécier la structure d'un texte journalistique, tout en négligeant l'utilisation de sources médiocres. Pour Sundar (1998), la qualité de l'information journalistique peut être comprise comme le degré ou niveau d'excellence dans l'information. Diakopoulos (2019 : 24) indique que la qualité peut être entendue à travers les dimensions de précision, de compréhensibilité, d'actualité, de fiabilité et de validité²⁰. Il souligne son importance dès lors que l'information et la connaissance produites par les journalistes sont utiles à la société.

Dans le contexte de l'automatisation de la production d'informations, il est unanimement reconnu que la qualité ne peut être réalisée qu'à partir du moment où les données alimentant l'information générée présentent une qualité suffisante. Toutefois, ce seul aspect ne peut être pris en considération : la qualité des processus via lesquels s'opèrent l'analyse et la mise en

¹⁹ Un modèle de données consiste en une définition des entités (classes), de leurs attributs (propriétés) et de leurs relations (Haug *et al.* 2009).

²⁰ Voir aussi "Dimension évaluative", p.145

forme est elle aussi importante, de même que la qualité du récit médiatique généré de manière automatique. Une information journalistique s'appuyant sur une mauvaise qualité de données peut s'avérer critique. Cela est particulièrement vrai dans le domaine de l'information économique et financière, où une information est susceptible d'exercer une influence sur marchés (McCallum 2012).

Lecompte (2015) cite le cas du traitement automatisé du deuxième rapport trimestriel de Netflix, publié en juillet 2015, qui affirmait que l'entreprise n'avait pas rencontré ses attentes. Cette information erronée avait eu pour conséquence une chute du cours de l'action Netflix... alors que c'était l'inverse qui s'était produit dans les faits. Le logiciel WordSmith, utilisé pour générer la dépêche erronée de l'Associated Press, n'avait pas scindé deux données, d'une valeur respective de "7" et de "1", aboutissant au signalement d'une chute de 71% du cours de l'action. Lecompte souligne que le principal enseignement de cette histoire est celui de la nécessité d'une forme de contrôle éditorial pour détecter les valeurs aberrantes. Des données de mauvaise qualité dans un projet de journalisme automatisé ont provoqué un mouvement de panique sur les réseaux sociaux. Le 22 juin 2017, "Quakebot"²¹ annonçait un tremblement de terre d'une magnitude de 6,8 sur l'échelle de Richter dans la région de Los Angeles. Or ce séisme ne s'était pas produit ce jour-là... mais en 1925. L'erreur provenait d'un bug dans la mise à jour dans la base de données de l'institut géologique américain (UGS), qui avait révisé ce tremblement de terre en l'encodant non pas en 1925 mais en 2017. Pour Ken Schwenke, l'auteur de "Quakebot", cela signifie que le facteur humain reste important dans un projet d'information automatisée. Cela démontre également les limites d'un tel projet journalistique : les données doivent être vérifiées régulièrement, de même que les générations produites, et cela d'autant plus lorsque celles-ci présentent une valeur d'alerte²². Ces exemples illustrent également le fait que lorsqu'un récit journalistique se base sur des données de mauvaise qualité, il sera mauvais lui aussi. Cela implique un nécessaire contrôle éditorial des contenus automatisés. Linden (2017, 2017b) souligne que *"l'intégrité, la qualité et la fiabilité des données disponibles sont essentielles aux informations automatisées tout comme à d'autres formes de journalisme. Par conséquent, le choix et l'évaluation des données devraient être un processus journalistique alors que la validation, la standardisation et la normalisation sont normalement transmises aux programmeurs"*.

La standardisation fait partie des bonnes pratiques (Boydens 20014) pour assurer une interopérabilité optimale en vue de la réutilisation des données (Batini 2009, Janssens 2012, Boydens). Standardiser, c'est appliquer la "commune mesure" (Boydens 2012), c'est disposer de conventions univoques pour une représentation correcte des données. Cela porte non seulement sur les formats (ouverts et donc accessibles de manière indépendante d'un logiciel) mais aussi sur les valeurs des données et leur étiquetage uniforme. La standardisation participe au cycle d'amélioration des données, donnant lieu à des activités de normalisation qui consistent dans la correction ou le nettoyage d'un jeu de données. Cela étant, elle peut s'avérer infruc-

²¹ Voir supra p.103

²² "Revenge of Y2K? A software bug might have caused false alert for big (and very old) earthquake", Rong Gong Li II, *Los Angeles Times*, 27/06/2017, consulté le 28/06/2017, URL : <http://www.latimes.com/local/lanow/la-me-earthquakes-earthquake-68-quake-strikes-near-isla-vista-calif-jyhwh-htmstory.htm>

tueuse pour différentes raisons : une hétérogénéité représentationnelle (à propos des attributs), une hétérogénéité ontologique (à propos des entités), ou une sémantique susceptible d'évoluer dans le temps (Madnik & Zhu 2006).

2.1.3 Modèle conceptuel d'évaluation de la qualité des données

Toute démarche s'appuyant sur des données dans le journalisme porte en elle plusieurs grands défis : celui de la collecte des données, qui suppose que les journalistes soient suffisamment formés, et cela d'autant plus lorsqu'ils décident de constituer eux-mêmes leur propre base de données et que ceux-ci sont responsables de la compréhension des données; celui du respect de la vie privée; celui de donner un contexte approprié aux données; et celui d'interroger correctement les données, ce qui sous-tend la nécessité de comprendre comment le jeu de données a été collecté, dans quels buts et de s'interroger sur les valeurs manquantes dans un jeu de données (McBride 2016). Toutefois, dans un projet d'automatisation de la production d'informations, les challenges ne sont pas seulement journalistiques : ils sont aussi techniques, dès lors qu'ils sont également relatifs à la manière dont les données ont été encodées. Prévenir les problèmes avant qu'ils ne se posent suppose de pouvoir identifier ceux-ci en amont du processus d'automatisation. Ces aspects seront importants dans le contexte de données extraites de pages web ou accessibles en *open data*, en raison de leur qualité potentiellement douteuse.

Comment rencontrer les challenges technique et journalistique inhérents à tout projet d'automatisation de la production d'informations? Si la qualité des données est un concept multidimensionnel, l'objet de cette section est d'établir comment répondre aux besoins des projets journalistiques reposant sur une automatisation du traitement des données. Le challenge technique comprend des indicateurs de qualité formels dont l'évaluation permet de prévenir les problèmes en amont du processus, tandis que le challenge journalistique comporte des indicateurs de qualité liés au caractère empirique des données. Toutefois, au-delà de la problématique liée à la qualité données et à celle de leur traitement, la valeur journalistique d'un projet automatisé dépendra du sens qui lui sera donné.

Le modèle conceptuel d'évaluation de la qualité des données développé ci-dessous a été testé et éprouvé dans le contexte de la première étude de cas consacrée au traitement automatisé de données relatives à la qualité de l'air en région bruxelloise²³. Il trouve son origine dans le cadre d'une recherche MARSOUIN (Université de Rennes 1) consacrée à l'étude de la qualité du portail *open data* de Rennes Métropole²⁴ (Trédan & Dierickx 2017) (Figure 2.8). Il a été adapté au contexte spécifique de la production automatisée d'informations, en s'appuyant sur les concepts développés dans les deux premières sections de ce chapitre, de manière à joindre les enjeux technologiques (liés à la qualité formelle des données) et les enjeux journalistiques

²³ Voir le chapitre 3, "L'artefact 'Bxl'air bot' dans les usages de la rédaction d'Alter Echos", p.159.

²⁴ Bien que ce portail soit considéré comme un modèle du genre, cette recherche posait la question générale de sa qualité, tout en la confrontant aux discours qui ont accompagné son développement. Cette recherche était articulée en deux temps, celui de l'analyse quantitative de la qualité du portail – et de ses 154 jeux de données – compte tenu de quatre types de publics identifiés (administrations, collectivités, développeurs/entrepreneurs et journalistes), et celui de l'analyse qualitative du discours des acteurs de l'*open data* Rennais.

(liés aux dimensions de la qualité des données et aux exigences journalistiques communes à toute démarche relevant d'une approche par données dans le journalisme) (Dierickx 2017).

Qualité formelle des données		Dimensions de la qualité des données			
Axe documentaire	Mention de la licence, identifiant unique, présence de métadonnées, conformité aux métadonnées.	Intrinsèques	Contextuelles		
Axe d'encodage	Pas de problème d'encodage, pas de surcharge avec du code HTML, pas de doublon(s), pas de colonne(s) ou ligne(s) vide(s).			Provenance (source authentique)	
Axe normatif	Syntaxe e-mail normalisée, syntaxe URL, adressage, dates et chiffres normalisés, respect des normes géographiques, application de la norme Dublin Core pour les métadonnées.			Exactitude (correction syntaxique)	Nombre de données approprié (colonnes ou lignes vides)
Axe sémiotique	Pas d'incohérence(s) orthographique(s), étiquetage des colonnes explicite, étiquetage des colonnes non-ambigu, pas de valeur(s) manquante(s).			Précision des valeurs (anomalies)	Pertinence
		Actualité (date et fréquence de mise à jour)	Complétude (valeurs manquantes)		
		Représentationnelles			
		Interopérabilité (formats ouverts, lisibles par un humain et par une machine)			

FIGURE 2.8 – Modèle réalisé pour l'évaluation du portail *open data* de Rennes Métropole

Les limites de ce modèle tiennent aux faits qu'une donnée de mauvaise qualité peut cohabiter avec une donnée correcte sans que cela génère d'erreurs et que, malgré une identification claire des problèmes, ceux-ci ne vont pas pouvoir toujours être résolus. Il s'agit donc de l'envisager comme un appui aux prises de décisions, dans le cadre d'un processus de conception d'un artefact d'automatisation. Une autre limite concerne l'évolution dans le temps des connaissances liées à un domaine d'application : par exemple, une norme de référence peut très bien être révisée à la hausse ou à la baisse. Par ailleurs, il est possible que ce modèle, appliqué à d'autres contextes d'utilisation de données, ne donne pas lieu à des résultats probants : l'évaluation de données numériques et toujours dépendante de son contexte d'usage.

Relever le challenge technique

Les indicateurs de qualité des données relatifs au défi technique ont pour objet une réutilisation optimale des données (Tableau 2.1). Leur objet est de pouvoir évaluer la qualité formelle d'un jeu de données en s'appuyant sur quatre axes complémentaires : (1) l'axe documentaire, lié à la compréhensibilité des ensembles de données, (2) l'axe d'encodage, relatif à l'aspect du codage technique des données, (3) l'axe normatif, considérant que le respect de standards favorise l'interopérabilité et donc la réutilisation des données, (4) l'axe sémiotique, qui concerne la cohérence syntaxique et sémantique des données. L'évaluation de ces différentes variables s'appuie sur un mode booléen "vrai" ou "faux".

Le nombre total d'indicateurs utilisés pour une évaluation peut varier d'un ensemble de données à un autre, car chaque domaine d'application dispose de ses propres particularités. Il convient également de rester prudent quant à la manière d'aborder ce modèle, dans la mesure où des anomalies formelles peuvent faire l'objet d'interprétations différentes. Ainsi en va-t-il de la valeur NULL, qui peut être interprétée de plusieurs manières : l'information existe

mais n'est pas connue, l'information n'est pas pertinente pour l'entité, l'information est pertinente mais n'existe pas pour l'entité, la valeur de l'attribut est égale à zéro (Hainaut 2012). De plus, des données de mauvaise qualité peuvent coexister avec des données correctes sans générer d'erreurs (Wang *et al.* 1995). Une source peut également ne pas contenir d'erreurs mais les données n'auront pas la signification attendue par l'utilisateur (Madnick & Zhu 2016). Cela implique une bonne connaissance du domaine d'application. Dans certains cas, la sollicitation d'experts s'avèrera donc indispensable. La résolution des problèmes rencontrés pourra être effectuée de plusieurs manières : dans le cas d'un identifiant manquant, une variable supplémentaire pourra être ajoutée au jeu de données si les métadonnées sont manquantes ou non conformes, le producteur des données pourra être contacté afin d'apporter des éclaircissements ; des problèmes d'encodage pourront être résolus en changeant l'encodage du document en UTF-8, qui tient compte des caractères accentués ; des surcharges du code HTML et des données dupliquées donneront lieu à un nettoyage du jeu de données.

Axe	Variables
Axe documentaire	Mention de la licence d'utilisation Identifiant unique Présence de métadonnée Conformité aux métadonnées
Axe d'encodage	Pas de problème d'encodage Pas de surcharge du code HTML Pas de données dupliquées
Axe normatif	Application des standards (adressage, géolocalisation, unités de mesure, URL normalisés, métadonnées, ...)
Axe sémiotique	Pas d'incohérences orthographiques Etiquetage des colonnes explicite et non ambigu Pas de valeurs manquantes

Tableau 2.1 – Indicateurs relatifs à la qualité formelle des données

Relever le challenge journalistique

Les enjeux liés à une approche par données dans le journalisme²⁵ et les standards relatifs aux pratiques professionnelles²⁶ ont été mis en parallèle avec les dimensions relatives à la qualité des données. Celles-ci s'intéressent à ses caractères intrinsèques (les valeurs des données sont exactes, précises et actuelles)²⁷, représentationnels (les données sont interopérables, un préalable dès lors que celles-ci sont utilisées dans des projets d'automatisation) et contextuels (indications sur la provenance des données, complétude du jeu de données et pertinence des données). Les indicateurs de qualité qui y sont associés reposent sur deux dimensions en par-

²⁵ Voir "Enjeux et défis professionnels", p.57

²⁶ Voir "Dimension normative", p.122

²⁷ Voir "Dimension normative", p.122

ticulier, car elles sont davantage en adéquation avec le domaine d'application : les dimensions intrinsèque et contextuelle, qui répondent aux préoccupations journalistiques de précision, d'actualité et d'exhaustivité.

Dimension	Variabes
Intrinsèque	Exactitude (correction syntaxique) Précision des valeurs (pas d'anomalies) Actualité (date et fréquence de mise à jour)
Contextuelle	Provenance (source authentique) Nombre de données approprié (pas de lignes vides) Complétude (pas de valeurs manquantes) Pertinence

Tableau 2.2 – Indicateurs relatifs aux dimensions de la qualité des données

Ces différentes variables sont évaluées sur un mode booléen (vrai ou faux). La dimension de la "complétude" pourrait être plus difficile à évaluer, en raison du problème de la valeur NULL, déjà identifié dans le volet relatif à l'évaluation du challenge technique. L'indicateur lié à la quantité appropriée de données pourrait également être difficile à évaluer, surtout si l'ensemble de données comporte des centaines d'enregistrements. La résolution des problèmes rencontrés prendront différentes formes : la compréhension du domaine d'application permet de détecter des valeurs anormales et d'expliquer des valeurs manquantes à l'examen d'un jeu de données, et la fiabilité d'une source de données requiert des activités de vérification (*fact-checking*). Un expert du domaine d'application peut permettre une meilleure compréhension des problèmes observés. Par ailleurs, une source primaire, qui désigne le responsable du contrôle et du suivi dans le temps de la qualité de l'information produite, ne garantit pas nécessairement la qualité d'un ensemble de données (Boydens 2014).

Relever les challenges technique et journalistique

Si des indicateurs empiriques relèvent de l'interprétation, ils ne peuvent être évités dans un contexte journalistique. Il appartient aux journalistes, ou à ceux qui sont impliqués dans un projet axé sur les données, de rester critiques vis-à-vis d'un jeu de données, y compris lorsqu'il provient d'autorités publiques ou lorsqu'il est signalé comme provenant d'une source authentique. Ce constat a donné lieu à ce modèle général d'évaluation, qui propose de relever, de concert, les challenges technique et journalistique. Ce modèle s'articule en cinq axes complémentaires. Ils visent à s'assurer (1) de la fiabilité de la source des données, (2) de disposer des accès nécessaires à l'utilisation données, (3) de disposer des clés de lecture nécessaires à la compréhension des données et de leur contexte de production, (4) de disposer de données dont les caractéristiques permettent une automatisation, (5) de présenter des caractéristiques pertinentes pour un usage journalistique. Il consiste en une série de questions dont le niveau critique peut nécessiter une part d'investigation.

- **Source** : Le diffuseur des données en est-il le producteur et/ou la source authentique? Dans le cas où le diffuseur des données n'en est pas le producteur original et/ou la source authentique, quelle est la nature de sa relation avec le producteur original des données et/ou la source authentique? Le diffuseur, le producteur et la source authentique sont-ils dignes de confiance?
- **Accès** : Les données sont-elles librement accessibles? Font-elles l'objet d'une licence permettant leur libre réutilisation? Sont-elles disponibles dans un format structuré?
- **Documentation** : Les données sont-elles documentées par des métadonnées ou par tout autre type d'information permettant de comprendre la structure de la base de données, de lever les éventuelles ambiguïtés dans l'étiquetage des données, d'apporter un expertise pour comprendre à quoi correspondent les valeurs des données? Des éléments de contexte sont-ils apportés?
- **Automatisation** : Les données sont-elles fournies dans un format libre et exploitable? Les valeurs des données respectent-elles les standards? Les valeurs des données sont-elles exactes? Le jeu de données est-il complet et à jour?
- **Pertinence journalistique** : Les données présentent-elles une valeur ajoutée en termes journalistiques? En quoi le traitement des données fait-il sens?

Ce modèle d'évaluation conceptuel peut s'avérer insuffisant dès lors que les données sont délivrées en temps réel. L'identification du cycle de vie des données (Figure 2.9), un concept issu du data management, peut permettre de lever, partiellement ou totalement, certaines ambiguïtés qui pourraient subsister quant à la manière dont les flux de données sont organisés. En d'autres mots, comment récupérer la "bonne donnée" au "bon moment"? Cette identification a pour objet de baliser les processus d'acquisition, de curation et de préservation des données (Fox 2014). La modélisation du cycle de vie des données vise à optimiser la gestion des flux au sein d'un système d'information, de la création des données à leur archivage et destruction (Reid 2007). Les conditions de production des données – et les prises de décision y afférant – peuvent également éclairer sur la manière dont sont envisagés les flux de données. Elles peuvent mettre en évidence des problèmes sur le plan de leur gestion et de leur maintenance dans le temps, renvoyant aux questions de politique, de stratégie et de moyens humains et matériels mobilisés pour produire ces données.

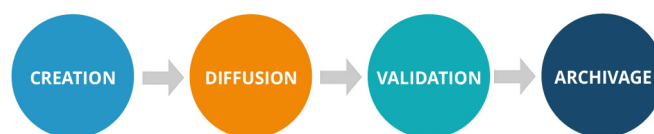


FIGURE 2.9 – Modèle conceptuel du cycle de vie des données

Ce modèle conceptuel d'évaluation de la qualité des données témoigne de préoccupations partagées, sur un plan théorique, entre les mondes du journalisme et de la technique, ainsi que de l'apport de la littérature relative aux bases de données informatisées. Sa mise en œuvre peut aider à prendre les décisions les plus appropriées pour la conception d'un système d'automatisation de la production d'informations, en vue de prévenir des erreurs dans la récupération des données ou des anomalies dans les valeurs des données. Les ambiguïtés n'ayant pu être traitées avec ce modèle pourront être levées, partiellement ou totalement, en s'intéressant au cycle de vie des données. Dès lors que des métadonnées, de même d'autres types d'informations contextuelles, ne sont pas toujours fournies pour documenter les données, un expert du domaine d'application sera requis. Sur le plan opérationnel, la principale limite de ce modèle est liée à des interprétations humaines subjectives. Il convient de garder à l'esprit que des problèmes de qualité des données peuvent survenir dans n'importe quel ensemble de données, qu'il provienne de sources privées ou publiques. Dans la perspective d'usages finaux, au mieux ces problèmes seront identifiés, au mieux ils pourront être traités.

2.1.4 Traitement des données : l'illusion de l'objectivité

Une approche par données dans le journalisme peut être définie comme un phénomène culturel, technologique et académique s'appuyant sur l'interrelation de trois facteurs : la technologie pour collecter et traiter des données, l'analyse des données, et le mythe selon lequel de larges ensembles de données vont permettre de réaliser des objectifs de vérité et d'objectivité (Sandoval-Martín & La-Rosa 2018). Si la quantification "*transforme le monde en données*", l'analyse dit ce qu'elles signifient. Cette activité implique des "*choix complexes de processus qui formalisent l'acte de dompter, de mesurer ou de catégoriser*" (Stray 2016). Aussi, l'analyse de données constitue-t-elle une autre forme de représentation du réel qui, malgré l'aura de "*vérité objective*" des données, peut donner lieu à des déformations ou à de "*mauvaises*" représentations (Leon 2018). Toute activité d'analyse de données renferme une part de subjectivité (Espeland & Stevens 2008). Y compris lorsqu'elle fait l'objet d'une automatisation, celle-ci va également dépendre de choix et de jugements éditoriaux où peuvent être autant mis en balance des critères "objectivables" que des critères plus subjectifs²⁸. Une mauvaise qualité de données ne sera donc pas la seule cause de l'échec des processus d'automatisation de la production d'information : celui-ci pourra également découler d'une mauvaise interprétation ou de l'introduction de biais dans les résultats (Daniel *et al.* 2010).

Le concept d'objectivité est difficile à définir : en journalisme, il désigne à la fois des normes, des attitudes, une idéologie, un état d'esprit (Anderson 2018). Il s'agirait également d'une "armure discursive" qui protégerait les journalistes pour se défendre des critiques ou de velléités de régulation qui seraient externes à la profession (Schudson 2001, Anderson & Schudson 20019). L'idéal d'objectivité peut donner lieu à une variété d'interprétations : du déni de la subjectivité personnelle à une représentation qui serait (subjectivement) estimée comme

²⁸ "Robot news : responses to data-driven news production. Making and measuring news : data and algorithms in journalism", Alison Powell, London School of Economics and Political Science, 2016, consulté le 28/03/2016, URL : <http://www.lse.ac.uk/media@lse/Polis/documents/Data-Delve-Report-Mar-2016.pdf>

impartiale ou équilibrée (Lesage & Hackett 2014). Sur le plan philosophique, *"les journalistes affirment l'existence d'une vérité objective ultime (...) qui ne devrait, en aucune manière, être créée, inventée ou modifiée"*. Une vision constructiviste objectera qu'il n'existe pas de réalité objective et que l'information, en tant que représentation du monde, est forcément sélective et nécessite une interprétation. Ce concept suppose que *"la neutralité soit rendue possible, du fait de la séparation de l'objet de l'étude du sujet qui étudie"* (Hanitzsch 2009). Pour les journalistes, la posture de l'objectivité permet de revendiquer une autorité sociale en présentant leur travail comme crédible, équilibré et "vrai" (Lewis 2012). De plus, les journalistes impliqués dans une approche par données ne considèrent pas cette impossible objectivité, tant ils *"sont devenus obsédés par l'exactitude et la certitude croissantes au détriment d'une compréhension plus humble et du doute"* (Anderson 2018). En agissant en tant que filtres, à commencer par la sélection des données qui seront traitées, ils posent des choix qu'ils ne reconnaissent donc pas comme le résultat d'une construction professionnelle qui va participer à la construction de la réalité sociale transmise par les médias d'information (Shoemaker 1996). Si l'objectivité n'existe pas – les codes de déontologie professionnelle ont d'ailleurs supprimé les références à l'objectivité au milieu des années 1990 (Hammond 2017) –, ne resterait que l'honnêteté (Cornu 2009 :323–396).

Une donnée en elle-même ne raconte pas d'histoire : seul son traitement lui donne du sens, et cela relève du domaine de l'interprétation subjective. Aussi, *"trouver une histoire"* dans un jeu de données *"constitue un véritable acte de création culturelle"* (Stray 2016). Dès lors qu'ils sont interprétés, les chiffres peuvent s'avérer aussi subjectifs que les mots : les données peuvent renfermer autant de vérités que des contre-vérités (Howard 2014). C'est pourquoi il s'agit de les envisager avec suspicion (Stray 2016). Les données peuvent aussi renfermer des biais liés à leur sélection, voire avoir fait l'objet d'une manipulation intentionnelle de la part du producteur d'informations. La question de la fiabilité de la source des données se pose ici en filigrane. *"Si vous ne comprenez pas les données, d'où elles viennent et ce qu'elles représentent, vos conclusions peuvent être porteuses de biais. Les données financières sont spécialement sensibles à ce problème, mais cela peut intervenir dans n'importe quel type de données"* (McCallum 2012 :127).

Le sens pouvant être accordé à des données présentant une valeur numérique passe par une analyse de ces valeurs, laquelle a pour objet de proposer différentes formes d'interprétation de cette quantification (statistique, par exemple). Toutes les formes de quantification transforment le monde *"par leur existence même par leur diffusion et leurs usages argumentatifs, y compris dans le champ journalistique"* (Desrosières 2008). De plus, *"comme pour les mots, le but et la signification des nombres changent souvent comme ils voyagent à travers le temps et l'espace social (...) Dans un monde saturé de chiffres, il est facile de prendre le travail de quantification pour acquis"* (Espeland & Stevens 2008). Dès lors qu'une activité d'analyse des données consiste en une représentation du monde réel, elle est susceptible de comporter une marge d'erreur quant à la manière dont elle va être menée. Des questions non pertinentes peuvent être posées, amenant des résultats biaisés servant moins l'objectif de découvrir la vérité que

celui de confirmer une hypothèse²⁹. Une représentation de données peut être trompeuse, en négligeant des aspects importants au profit d'éléments plus favorables à la démarche dans le cadre d'une analyse statistique. Hors de leur contexte ou prises de manières isolées, les données pourraient occulter une partie de la réalité ou signifier tout autre chose (Daniel *et al.* 2010). Privilégier une échelle de temps sur une autre peut conduire à une analyse divergente : si le taux de chômage n'a jamais été au plus bas en dix ans, celui-ci est peut-être bien plus élevé qu'il y a quinze ou vingt ans. L'utilisation d'indices peut donner lieu à des résultats faussés par la prise en compte de valeurs disparates. Si un algorithme d'exploration de données peut permettre de découvrir de nouvelles connexions entre de multiples variables présentant une très grande importance statistique en raison de l'énorme quantité de données analysées, les résultats peuvent être sans signification et n'ajouter aucune valeur à l'information (Latar 2015).

Lorsqu'une analyse de données repose sur des statistiques issues de données empiriques récoltées via des méthodes d'enquête ou de sondage, il convient également de s'interroger sur la méthode de collecte des données, la taille de l'échantillon, sa représentativité et la marge d'erreur des résultats obtenus : "*une enquête mal menée, un questionnaire mal rédigé ou un échantillon mal choisi peuvent aboutir à des résultats statistiques complètement erronés*" (Foucart 2001). Des données issues de statistiques ne peuvent être abordées que de manière prudente et critique, bien qu'elles permettent d'identifier des tendances et des schémas importants (Lowrey 2019). Établir une relation entre deux variables ne permet pas nécessairement d'établir un lien de cause à effets : "*La causalité demande donc une description de la réalité à laquelle on se limite. Les interprétations d'une relation statistique ne sont pas des vérités objectives, en ce sens que deux personnes peuvent en proposer des interprétations différentes puisqu'elle est interprétée nécessairement dans un contexte différent, ne serait-ce qu'à cause des personnalités différentes*" (Foucart 2001). De plus, les résultats seront souvent présentés en termes de probabilités. La fonction des statistiques est de formuler de "bons" arguments et d'expliquer des différences comparatives. C'est là le propre d'une sociologie de la quantification, dont la visée est pourtant celle d'une connaissance "objective et neutre" (Desrosières 2008 :27).

Le traitement de données ne donne donc pas lieu à des acquis immuables et cela est d'autant plus vrai dans un contexte journalistique, où la description de la réalité s'inscrit dans le cadre mouvant d'une histoire susceptible de se construire au fil du temps. De plus, les domaines d'expertise sont également susceptibles d'évoluer dans le temps, pouvant ainsi donner lieu à des interprétations différentes, voire divergentes. Il en va de même avec les valeurs des données qui, à l'instar d'une photographie instantanée, reflètent un état E à un moment T. En ce sens, l'autorité du nombre peut être considérée comme éphémère. La manière dont il est abordé dans un processus visant à automatiser la production d'informations ne peut donc être envisagée qu'à travers les choix posés : avant d'être technique, un processus de génération automatisée est culturel et social. Le comprendre, c'est comprendre les enjeux et valeurs qu'il véhicule (Fuller 2003, Flanagan *et al.* 2010). Si la qualité des processus de génération automatique de contenus journalistiques dépend de la qualité et de la fiabilité des données qui

²⁹ "The Quartz guide to bad data", idem.

les nourrissent, elle dépend également de la manière dont elle va traduire, par la médiation du code informatique, des intentions humaines. Derrière tout code informatique, il se trouve toujours un programmeur qui va en fixer les règles et les procédures (Manovich 2013). De la même manière, aucune base de données ne peut être considérée comme "neutre" car, elle aussi, résulte de décisions humaines relatives à la manière dont les données ont été collectées, triées ou analysées. Pas plus que son analyse, aucune donnée ne peut être considérée comme objective (Rahman & Wehrmeyer 2018) : en cela, le statut de la donnée ne garantit pas de rendre le journalisme plus "objectif" (Lesage & Hackett 2014). Si la "rationalité" de la technologie s'oppose à la subjectivité journalistique, elle apparaît donc comme un mythe.

2.2 Dimension normative

La culture professionnelle du journalisme s'exprime également dans le cadre d'une dimension normative, qui fait référence à l'éthique du journalisme et à son engagement dans un idéal de bien public (Singer 2003). Cette dimension encadre la responsabilité sociale des médias, et recouvre notamment les principes de recherche de la vérité, de minimisation des torts, et d'action indépendante et responsable (Chung & Nah 2014). Elle s'inscrit dans le cadre d'une autorégulation professionnelle qui consiste en une réponse à une généralisation de la critique de l'information (Grevisse 2003). Relevant du domaine des intentions, elle devrait être envisagée en amont de la conception d'artefacts d'automatisation où chaque décision adoptée façonnera, à son tour, les journalistes et/ou les audiences³⁰. La dimension normative de la culture du journalisme est aussi une affaire de perception quant à la manière dont les journalistes envisagent leur métier. Alors que, ces dernières années, le journalisme a davantage été caractérisé par un changement radical dans ses pratiques, la conception fondamentale de ce qu'il devrait être est restée stable (Singer 2019). Toutefois, les normes et pratiques journalistiques ne peuvent être considérées comme uniformes, dès lors qu'elles vont dépendre des contextes de production de l'information (Carson 2019).

Si un processus informatique incorpore les valeurs de ses concepteurs et développeurs, les journalistes ne sont pas en mesure d'y incorporer les leurs (Diakopoulos 2019b), à moins que celles-ci ne fassent l'objet d'une traduction sociotechnique... ou que les concepteurs et développeurs d'artefacts d'automatisation de la production d'informations alignent leurs valeurs sur celles des journalistes. Ceci rejoint l'un des principes posés par la théorie de la diffusion de l'innovation, qui bien que pouvant être considérée comme limitée en raison de la perspective déterministe qu'elle sous-tend, dispose que l'un des attributs permettant d'évaluer l'adoption d'une innovation consiste en sa compatibilité avec les valeurs et convictions socioculturelles de ses utilisateurs potentiels (Rogers 2003). La production d'informations pilotée par des algorithmes implique de nouvelles manières de faire du journalisme, induisant que la dimension éthique soit embarquée dans le code informatique (Diakopoulos 2019 :27).

³⁰ "10 principles for data journalism in its second decade", Paul Bradshaw, Online Journalism Blog, 07/08/2017, consulté le 20/09/2017, URL <https://onlinejournalismblog.com/2017/08/07/10-principles-for-data-journalism-in-its-second-decade/>

Pour les médias, l'intégration de certaines normes, routines et valeurs journalistiques sont essentielles lors de la conception d'artefacts d'automatisation, dès lors qu'ils vont refléter le savoir-faire de l'organisation médiatique (Linden *et al.* 2019 :39). L'arrivée de nouveaux acteurs dans le monde de l'information pose la question des valeurs ou standards éthiques communément partagés. Alors que les journalistes sont de plus en plus souvent invités à développer des compétences techniques, voire une forme de pensée computationnelle en vue de faciliter le dialogue avec ces nouveaux acteurs, une démarche à sens unique ne peut être considérée comme suffisante³¹. Cet alignement se justifie d'autant plus que les technologies d'automatisation peuvent reproduire, incarner ou modifier des normes professionnelles existantes (Anderson 2013).

Le terme norme est dérivé du nom latin "normat", qui signifie équerre, règle. Ce concept désigne " littéralement la formule abstraite ou le type concret de ce qui doit être". Une norme peut être considérée comme principe de comparaison et d'évaluation, et elle "renvoie à la notion de commune mesure" (Boydens 2012). Les normes consistent en des représentations du réel qui se présentent comme objectives (Berns 2011). Leur caractère est arbitraire : une norme a pour propriété d'être violée ou respectée (Moeschler 1985 :11). "*L'éthique et les normes existent pour des raisons rituelles, contribuant à la solidarité interne et à la cohésion à un groupe particulier; elles peuvent également représenter un moyen de définir un groupe par rapport à d'autres groupes*", indiquent Schudson et Anderson (2009), pour lesquels l'émergence de normes impliquent des mesures de contrôle hiérarchique sur les groupes sociaux. Déjà en 1890, Durkheim soutenait, dans une perspective fonctionnaliste, qu'il n'est pas possible pour une fonction sociale d'exister sans discipline morale, "*sinon rien d'autre ne reste que les appétits individuels, qui sont par nature dénués de liens et insatiables. S'il n'y a rien pour les contrôler, ils ne seront pas capables de se contrôler eux-mêmes*".

L'éthique professionnelle a pour objectif de baliser des pratiques responsables. Étymologiquement, le terme "éthique" est synonyme de morale : il renvoie au terme latin "ethicus", et au terme grec "êthikós". Ricoeur (1991) distingue néanmoins l'éthique de la morale : l'éthique est "*ce qui est estimé bon*", et la morale "*ce qui s'impose comme obligatoire*". La visée éthique, explique-t-il, consiste en une "*visée de la vie bonne avec et pour les autres, dans des institutions justes*". Par institutions, Ricoeur désigne "*les structures du vivre ensemble d'une communauté*". Dès lors, il estime nécessaire de soumettre cette visée à l'épreuve de la norme, la visée éthique étant "*suggérée par des conflits qui naissent de l'application même de ces normes à des situations concrètes*". Hanitzsch (2019) opère également cette distinction entre éthique et valeurs morales, bien que toutes les deux soient évaluatives : "*les valeurs morales sont spécifiques au contexte culturel dans lequel elles sont intégrées. Elles doivent donc être traitées comme des dimensions contextuelles de la culture du journalisme*" (Hanitzsch 2007).

L'éthique du devoir, à distinguer de l'éthique de la vertu (qui porte sur sa propre pratique) et de l'éthique téléologique (qui s'appuie sur ce qui crée le moins de torts), est relative aux codes

³¹ Voir "Développer une pensée journalistique", p.76

de déontologie dont la portée est universelle au sein d'une profession. À l'origine, la déontologie journalistique s'est affirmée comme constitutive de l'identité professionnelle (Cornu 2009 : 58). Mais les chartes et codes de déontologie journalistiques peuvent également être considérés comme un moyen d'assurer la responsabilité sociale des médias (Bertrand 1993), dès lors qu'ils ont pour vocation d'encadrer des pratiques responsables, dans le cadre d'un contrat moral passé avec la société. Non-contraignants, leurs objectifs ne sont pas purement fonctionnels : ils sont aussi sociaux. Pour Grevisse (2003), ils attestent des signes des faiblesses d'une profession, ils sont les témoins d'une recherche identitaire et ils constituent une réponse aux critiques publiques. Toutefois, tous les journalistes du monde entier ne partagent pas l'avis selon lequel la profession doit se doter d'un code de déontologie alors qu'ils partagent tous "*un sentiment d'être éthique*" (Deuze 2005). Dans l'univers anglo-saxon, les journalistes s'attacheront moins au respect de règles codifiées qu'à celui de règles personnelles ou intériorisées. Selon Paul Bradshaw, la déontologie journalistique ne doit pas être envisagée comme un code prescriptif gravé dans le marbre³². Il ajoute que la production automatisée d'informations soulève les mêmes questions éthiques que dans n'importe quel autre choix éditorial : dès lors, il importe de réfléchir davantage à la manière dont ces processus sont conçus.

Dans cet esprit, cette section pour objet (1) d'établir les points de convergence et de divergence entre la manière dont sont balisées les pratiques professionnelles dans le monde du journalisme et dans celui de la technique; (2) de définir un cadre commun pour les nouvelles pratiques professionnelles liées au développement des technologies d'automatisation de la production d'informations, au regard de leurs usages attendus : qu'elles soient envisagées en tant que produit final (c'est à dire, délivré en tant que tel aux audiences) ou en tant que produit intermédiaire (c'est à dire, faisant l'objet d'un retraitement journalistique).

2.2.1 Dans le monde du journalisme

"Le droit à l'information, à la libre expression et à la critique est une des libertés fondamentales de tout être humain. (...) La responsabilité des journalistes vis-à-vis du public prime toute autre responsabilité, en particulier à l'égard de leurs employeurs et des pouvoirs publics", peut-on lire dans le préambule de la Charte de Munich, adoptée le 24 et 25 novembre 1971 par les représentants de syndicats des journalistes de six pays membres de la Communauté Européenne, puis par la Fédération Internationale des Journalistes (FIJ) au Congrès d'Istanbul en 1972. Ce principe de responsabilité sociale des journalistes sera réaffirmé quarante-sept ans plus tard dans la "Charte éthique mondiale des journalistes" adoptée par la Fédération internationale des journalistes³³. Elle souligne notamment le respect et la fiabilité des faits, la distinction entre le commentaire et la critique, le libre accès aux sources d'information, la primauté de la vérification des faits sur l'urgence du traitement médiatique, et le respect de la vie privée et du secret des sources. La plupart des organisations professionnelles mettent ces principes

³² Communication personnelle, recueillie par e-mail, le 11/08/2014.

³³ Source : https://www.ifj.org/fileadmin/user_upload/CHARTE_D_ETHIQUE_MONDIALE_DES_JOURNALISTES_-_FR.pdf

en exergue dans leurs chartes et codes, explique Ricardo Gutiérrez³⁴, secrétaire général de la Fédération européenne des journalistes (FEJ), une organisation qui représente plus de 320.000 journalistes de 61 syndicats et associations dans 40 pays³⁵.

Si tous les médias n'ont pas pour vocation de concentrer leur ligne éditoriale sur les seules vertus démocratiques de l'information, ils ont en commun de s'adresser à des audiences sur lesquelles ils exercent potentiellement une influence en raison de leur rôle de caisse de résonance. C'est là un autre versant de leur responsabilité sociale. Selon Bardoel et d'Haenens (2004), il n'existe pas de théorie cohérente à propos de la responsabilité sociale des médias : ce concept est indistinctement employé pour désigner les contenus diffusés dans la presse, la fonction de la presse dans la société, la responsabilité du média envers la société, et la manière dont le média écoute et prend en considération le public. Les chercheurs indiquent également que la responsabilité des médias envers leurs publics n'est pas la seule variable à prendre en considération : par extension, ce principe s'applique à la responsabilité professionnelle via le respect de standards éthiques, à la responsabilité des médias envers le marché, et à la responsabilité des médias vis-à-vis des pouvoirs publics qui les contraignent à respecter les règles de la loi. McQuail (cité par Bardoel & d'Haenens 2004) propose une définition pratique du concept de responsabilité sociale des médias : celle-ci consiste dans leurs obligations à répondre aux attentes de la société. Il identifie quatre types de responsabilité : assignée (obligations établies par la loi), contractée (accord d'autorégulation avec la société et les politiques), assignée par soi-même (engagements professionnels volontaires pour le maintien de standards éthiques et d'objectifs publics), déniée (qui réfute les accusations d'irresponsabilité). Une autre définition proposée par McQuail (cité par Lauk & Kus 2012) désigne les "*processus volontaires ou involontaires par lesquels les médias répondent directement ou indirectement à la société pour la qualité et/ou les conséquences d'une publication*".

Un rapport réalisé dans le cadre du projet MediaAcT (*Media accountability and transparency in Europe - Responsabilité des médias et transparence en Europe, 2012*) relève que la responsabilité sociale des médias est avant tout une affaire de pratiques, lesquelles vont différer d'une culture médiatique à l'autre. Toutefois, il existe un consensus autour des concepts de "vérité", "exactitude" et "objectivité" (Hafez 2002). Dans le cadre de leurs pratiques professionnelles, affirme Ricardo Gutiérrez³⁶, les journalistes n'ont pas toujours conscience de cette dimension sociale ancrée dans leurs productions. "*L'expérience du Conseil de déontologie belge prouve que beaucoup de journalistes dérapent, à un moment ou un autre, par manque de temps — ils doivent travailler dans l'urgence et, de plus en plus, pour une variété de supports — ou parce qu'ils ont tendance à respecter de manière un peu trop aveugle les diktats des autorités. Les principes de la responsabilité sociale des journalistes sont influencés par l'exercice concret de leur travail*". Dans une situation de stress, regrette-t-il, les principes moraux et parfois légaux peuvent se perdre. Pour faire sens, l'exercice de la responsabilité sociale journalistique devrait impli-

³⁴ Communication personnelle, Bruxelles, le 11/01/2016.

³⁵ Site : www.europeanjournalist.org

³⁶ Communication personnelle, idem.

quer un exercice critique systématique constant. Cela suppose d'admettre que les journalistes peuvent commettre des erreurs et que celles-ci doivent être corrigées (Lauk & Kus 2012). Plaisance (2000) indique que la responsabilité sociale des médias devrait être plus largement comprise comme "*une dynamique d'interactions entre un milieu donné et la valeur des ensembles d'individus ou de groupes recevant des messages des médias*", dès lors que la responsabilité des médias est fonction de leur influence. De plus, le succès d'une communication dépend de sa crédibilité, laquelle serait définie tant par le contenu du message que par la nature de la transmission : à qui les médias doivent-ils rendre des comptes en ce qui concerne un public dont les membres peuvent avoir des interprétations différentes? Dans une approche libertaire de la presse, l'exigence de l'autonomie personnelle du journaliste, suggérant un concept de responsabilité limitée sera mis en avant ; tandis qu'une approche communautaire élargira ce concept aux conséquences sociales et culturelles de l'exercice de l'idéal journalistique (Plaisance 2000).

La recherche s'est souvent accordée sur l'idée que, malgré un "déluge" de nouveaux fournisseurs d'informations et des contextes de production différents, il existe des dénominateurs communs caractérisant les pratiques et valeurs des journalistes (Chung & Seungahn 2014). Il s'agit de l'une des manières d'aborder la culture journalistique qui, ici, transcende les frontières culturelles nationales et organisationnelles (Hanitzsch 2006). La littérature scientifique esquisse l'idéologie journalistique en cinq traits distincts (Linden 2017) : les journalistes fournissent un service public et se considèrent comme des chiens de garde (de la démocratie) au nom d'un public ; ils sont impartiaux, neutres, objectifs et équitables ; ils doivent être autonomes et libres de tout lien de subordination ; ils ont un sentiment d'immédiateté, d'actualité et de rapidité ; et ils disposent d'une solide base morale et éthique qui leur donne une raison légitime et valable d'agir au nom du public. Ce concept de "bien public", qui sous-tend la responsabilité sociale des médias, trouve un point d'équilibre entre les droits individuels et collectifs, et le droit de la presse à la liberté d'expression (White 2008). L'idéal du service au public peut également être considéré comme un élément puissant de l'idéologie du journalisme (Deuze 2005). Toutefois, une définition du journalisme en tant que culture et idéologie, se fondant sur un système de croyances et un ensemble de pratiques professionnelles, couvre seulement une partie de ce qu'est le journalisme (Kammer 2013).

L'idéal journalistique, affirme Bernier (2006), repose "*sur bon nombre de prescriptions, dont certaines concernent directement la substance des messages qui seront diffusés tandis que d'autres portent sur les relations des journalistes avec les autres acteurs sociaux*". Neveu (2010) souligne qu'il s'agit "*d'une vision enchantée du journalisme, de ses fonctions démocratiques, de ses pouvoirs (...) nombre de journalistes vivent leur métier comme une mission au service du public, à qui ils apportent des informations utiles*". Mais si le journalisme est un bien public, analyse McChesney (2012), il s'agit de "*quelque chose que la société requiert mais que le marché ne peut produire en qualité ou quantité suffisante (...) pour fournir les ressources à un journalisme démocratique*".

L'observation des pratiques journalistiques, donne lieu à "deux versions opposées qui alternent dans le discours traditionnel" : d'une part, un courant "à forte teneur subjectiviste", qui "exalte la responsabilité et l'indépendance des journalistes, quitte à stigmatiser les 'dérapages' et les 'dérives' de quelques 'brebis galeuses'; d'autre part, un courant "à forte coloration objectiviste, (qui) met l'accent (...) sur l'emprise croissante de la logique commerciale dans les entreprises de presse, celles-ci étant devenues pour la plupart la propriété de grands groupes industriels et financiers plus soucieux de 'parts de marché' que de la qualité de l'information ou des programmes" (Grevisse 1998). Un courant critique, qui met en doute l'existence d'un lien entre médias et démocratie, s'appuie sur ce dernier argument "dans un contexte capitaliste où la mission économique des entreprises de presse a pris une importance déterminante" (Bernier 2006). Aussi, l'idéal démocratique de la fonction d'un média d'information doit-il être pondéré par l'interdépendance des médias envers les pouvoirs politiques et économiques, étant entendu qu'un média est aussi une entreprise commerciale ayant pour vocation de se développer au sein d'un marché de consommateurs pour y assurer sa pérennité (McQuail 2003, Gingras 2009).

Le développement des pratiques journalistiques s'appuyant sur la collecte et le traitement de données a suscité une réflexion relative aux bonnes pratiques qu'il conviendrait d'adopter, tout en considérant que les principes éthiques qui le sous-tendent restent les mêmes que dans toutes les autres pratiques journalistiques. Une approche par données dans le journalisme accentue les questions relatives à la fiabilité des sources, à l'exactitude des données, au droit à collecter des données via des processus automatiques d'extraction, et à l'équilibre à trouver entre le droit du public à être informé et au respect de la vie privée. Il pose aussi celles de l'interprétation des données, dans le respect du principe fondamental du respect de la vérité. À partir d'une étude de cas portant sur une cartographie des permis de port d'armes aux États-Unis, Craig *et al.* (2017) proposent, pour baliser une approche par données dans le journalisme, une grille d'évaluation s'appuyant sur les principes fondamentaux de la déontologie journalistique. Leur proposition s'appuie sur les problématiques liées aux pratiques du datajournalisme : (1) la liberté d'informer versus la responsabilité et l'objectif journalistiques, (2) la confidentialité et la vérification des données, (3) les conséquences de la publication de données à caractère privé. Les chercheurs envisagent des alternatives, comme ne pas publier les noms et adresses de personnes identifiées dans des jeux de données, utiliser les données pour montrer des tendances générales plutôt que particulières, ou croiser des jeux de données sensibles avec d'autres jeux de données "pour servir un objectif journalistique plus large". Pour Loosen (2018), un journalisme de données responsable doit également être conscient du fait qu'il interpelle et questionne une société mise en données, tout comme il devrait réfléchir à sa dépendance vis-à-vis des données.

Aux États-Unis, l'Online News Association (ONA) a engagé une réflexion sur des recommandations spécifiques dans le contexte de l'automatisation de la production d'informations³⁷. Actif au sein de cette organisation, Tom Kent, éditeur à l'agence de presse américaine Associated Press (AP) qui utilise, depuis 2014, des technologies de génération automatique de textes,

³⁷ Source : <http://ethics.journalists.org/topics/robot-journalism/>, consulté le 10/01/2016

a établi une *check-list* éthique dans laquelle il propose notamment de s'interroger sur le degré de précision des données, la manière dont l'automatisation organise les données, le contrôle du système d'information ainsi que sa maintenance³⁸. À l'AP, des garde-fous ont été établis dans le cadre d'une collaboration avec Automated Insights³⁹ pour l'automatisation de données boursières et sportives. "*Les questions éthiques fondamentales dans le journalisme robotisé consistent à s'assurer que les données (...) sont correctes et que vous avez le droit de les utiliser; que vous divulguez que vous utilisez des processus automatisés; et que vous comprenez suffisamment l'automatisation pour pouvoir défendre la façon dont une histoire a été écrite. Certains systèmes automatisés créent des affichages vidéo ou photo pour accompagner les textes. Si c'est le cas, vous devez vous assurer que le système n'accède qu'aux images que vous avez légalement le droit d'utiliser, et qu'elles ne contiennent rien de satirique, haineux ou contraire à vos standards*". Le texte précise encore qu'un système automatisé doit faire l'objet de vérifications humaines en amont. Toutefois, dans la plupart des cas, la vérification des processus, en raison des compétences techniques qu'elle requiert, ne pourra pas être effectuée par un journaliste.

Dans un rapport publié par le Tow Center for Digital Journalism, Hansen *et al.* 2017 posent des exigences de transparence : que ce soit des technologues envers les journalistes ou des médias envers les audiences. Leurs recommandations préconisent notamment que les journalistes soient formés à l'utilisation des technologies, que ce soit sur le plan éthique ou sur celui des pratiques; qu'un cadre soit partagé par les agents du monde de la technique et ceux du monde du journalisme pour une utilisation éthique des données; et que des valeurs éditoriales soient intégrées lors de la conception d'artefacts d'automatisation. Elles soulignent également la nécessité d'informer les audiences sur les processus à l'œuvre. Pour être comprise, cette description devrait être traduite en termes non techniques, dans le souci d'informer sur la manière dont la technologie a été utilisée et comment les choix ont été opérés.

Montal et Reich (2017) se sont intéressés à la manière dont est attribuée la paternité d'un texte généré de manière automatique, en procédant à sept entretiens de "figures-clés" d'entreprises de presse pionnières, aux États-Unis, en matière de journalisme automatisé. Tous les répondants reconnaissent au logiciel un caractère humain lié au domaine de l'informatique ou du journalisme. S'ils soulignaient l'importance de la transparence envers le lecteur, une majorité ne se positionnait pas en faveur d'une divulgation complète de la nature logicielle de l'auteur. Pour les chercheurs, la divulgation – machine ou homme-machine – doit pourtant être effective pour atténuer l'apparence "biaisée" ou "non biaisée" d'une production journalistique générée de manière automatique. Cette exigence de transparence se justifie par le caractère d'intérêt public d'une information journalistique. Et, indiquent-ils, celle-ci ne concerne pas seulement l'auteur algorithmique mais aussi la mention de la source des données et de la méthodologie utilisée pour traiter ces données.

³⁸ "An ethical checklist for robot journalism", Tom Kent, Medium, 24/02/2015, consulté le 19/03/2017, URL : <https://medium.com/@tjrkent/an-ethical-checklist-for-robot-journalism-1f41dcbd7be2/>

³⁹ "Ethics of robot journalism : How Automated Insights poses issues for data collection and writing", Megan Doll, University of Wisconsin, Center for Journalism Ethics, 20/10/2015, consulté le 10/01/2016, URL : <https://ethics.journalism.wisc.edu/2015/10/20/ethics-of-robot-journalism-how-automatedinsights-poses-issues-for-data-collection-and-writing/>

Sur le plan opérationnel, la transparence n'est pas toujours réalisée. Si l'on prend l'exemple des sociétés spécialisées dans la génération automatique de textes, il est difficile d'établir clairement lesquelles, parmi elles, ont pour clients des médias d'information : certains contrats comportent, en effet, de strictes clauses de confidentialité. La confirmation de Frank Feulner, *Chief business development officer* au sein de la société allemande AEXEA, est sans équivoque : "des accords de confidentialité nous interdisent de nommer nos clients"⁴⁰. Cela signifie que les articles générés de manière automatique ne sont pas toujours présentés comme tels aux lecteurs, mais aussi que cette non-divulgaration relève de la seule responsabilité du management du média. Serait-ce parce que le média ne considère pas ses audiences prêtes à accepter le fait que des récits journalistiques puissent être générés de manière automatique, ou parce qu'il considère que cela pourrait engendrer une forme d'anxiété dans ses équipes rédactionnelles? Cela étant, une attribution claire de la nature non-humaine de l'auteur d'un récit journalistique ne va pas nécessairement supposer que le lecteur lui accordera davantage de crédibilité. Waddell (2018) a mis en exergue l'anthropomorphisme d'un lecteur qui aurait tendance à démontrer davantage d'affinités avec des auteurs humains et à se montrer sceptique sur les productions réalisées par des logiciels.

Ne pas divulguer la nature exacte de l'auteur d'un contenu journalistique ne participe pas à une démarche de transparence et d'explicabilité qui favorisent la confiance (Linden *et al.* 2019). Cette transparence devrait être réalisée à plusieurs niveaux : dans l'attribution de la source des données, dans des explications relatives à la manière dont les données ont été collectées, dans des prises de décision éditoriales qui reviennent aux journalistes et pas à des techniciens qui n'ont pas été formés au journalisme, dans l'autorégulation en respectant des règles relevant de l'éthique journalistique, et dans la prise en compte de la responsabilité sociale qu'exerce un média d'information, qui implique la correction de toute information erronée. Cette transparence devrait également être réalisée lorsque l'information qui apparaît au lecteur procède d'une logique de personnalisation (Linden *et al.* 2019 :41-43).

Le développement de stratégies d'automatisation de la personnalisation de l'information, dont l'objet est de fournir des contenus sur mesure aux audiences en fonction de leurs choix ou de leurs préférences, fait craindre une remise en question du droit du public à être informé de manière complète et pluraliste, ce qui constitue l'un des fondements de la responsabilité sociale des médias d'information. Le principal danger de ce phénomène est celui d'un enfermement culturel des audiences, par la création de "bulles de filtres" ("*filter bubbles*"), décrites par Pariser (2011) comme l'état dans lequel se trouve un internaute lorsque les informations affichées sur une page web ou une application mobile résultent d'une personnalisation basée sur la collecte de ses données personnelles⁴¹. En cloisonnant l'utilisateur dans ses propres univers et

⁴⁰ Communication personnelle, Stuttgart, le 12/03/2016.

⁴¹ La recherche récente s'est essentiellement concentrée sur ce phénomène dans le contexte des réseaux sociaux, sans pour autant s'accorder dans leurs conclusions en ce qui concerne le degré ou le danger d'enfermement induit par les algorithmes à l'œuvre (Sergeant & Tagg 2019, Sphor 2017, Zuiderveen *et al.* 2016, Nikolov 2015 *et al.*).

référents socioculturels, elles peuvent ainsi participer à la création de monocultures (Bozdag 2013). "Le souci est que la personnalisation amène les individus à consommer de plus en plus la même information. Par conséquent, il est peu probable que les gens consomment des informations qui défient leurs opinions ou contredisent leurs intérêts, ce qui peut comporter des risques pour la formation de l'opinion publique dans une société démocratique" (Graefe et al. 2015). "Une histoire a-t-elle été choisie pour sa prééminence statistique ou parce qu'un algorithme de personnalisation l'a choisi pour nous? Si la sélection était basée sur des statistiques, alors lesquelles – le nombre de mentions de l'histoire, l'autorité des médias qui le promeuvent, ou quelque chose d'autre?", demande Pasquale (2015).

Les journalistes ne sont pas nécessairement associés à la conception de ces systèmes personnalisés qui consistent "en une forme d'individualisation technique qui permet d'interroger des données et de les réassembler comme un ensemble aux caractéristiques différentes" (Fuller & Goffey 2013). Mais en opérant une segmentation du public, la personnalisation d'informations offre aux éditeurs un moyen d'offrir un service à valeur ajoutée pouvant potentiellement être monétisé (Lavie et al. 2010). Sa promesse est de réaliser de nouvelles expériences plus agréables, mais elle soulève des inquiétudes relatives à la perte de connaissances partagées (Lewis & Westlund 2015). Surtout, il s'agit d'une stratégie de captation des audiences : fournir un flux personnalisé d'informations est un moyen de réduire la dépendance des utilisateurs envers d'autres fournisseurs d'informations (Thurman & Schifferes 2012). "La personnalisation devrait être un moyen d'améliorer les décisions prises par les rédacteurs humains, les professionnels engagés dans un journalisme de qualité en tant qu'élément essentiel d'une société ouverte", souligne Powers⁴².

La personnalisation pourrait également être considérée comme une solution pour éviter les biais techniques existant dans les services en ligne, mais celle-ci induit de nouveaux biais sans pour autant éliminer tous les autres (Bozdag 2013). Toutefois, si un biais existe dans les données nourrissant des procédures automatisées, ces biais seront automatiquement répercutés dans le contenu personnalisé. Dans le contexte de la génération automatique de contenus journalistiques, la personnalisation peut se traduire par différentes versions d'une même histoire en fonction des publics ou des individus, ou par des rapports sur une thématique particulière pour des clients individuels (Carlson 2014, Graefe 2016). C'est ce que pratique la société américaine Narrative Science avec son logiciel Quill. Un article publié par *Wired* rapporte que les ingénieurs de cette société peuvent introduire des biais dans le logiciel : par exemple, dans le cadre d'une rencontre sportive mettant en jeu des enfants, le texte généré peut davantage être orienté sur l'héroïsme des joueurs plutôt que sur leurs performances. Le logiciel peut aussi "adoucir" une défaite ou accentuer les éléments positifs d'un texte⁴³.

⁴² Citée dans "The Power of Personalization", Eryn Carlson, Nieman Lab, 2017, consulté le 20/10/2017, URL : <https://niemanreports.org/articles/the-power-of-personalization/>

⁴³ "Can an algorithm write a better news story than a human reporter?", Steven Levy, *Wired*, 24/04/2012, consulté le 18/10/2017, URL : <https://www.wired.com/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter/>

Le principal défi de la personnalisation d'informations, dans le cadre de processus d'automatisation, serait donc de "*de trouver un équilibre entre la création de chambres d'écho et le maintien de la mission de service public du journalisme*" (Hansen *et al.* 2017), alors que l'une des tâches du journaliste "*consiste à suivre, rassembler, organiser et contextualiser les informations utiles pour le public cible. Il doit passer par une énorme quantité d'enregistrements contenant des informations dont le degré de granularité est très varié (...) pour raconter son histoire sous tous les angles significatifs, et, en même temps, il doit réduire le bruit des contenus non pertinents*" (Tran & Herder 2015). L'enjeu est donc de trouver l'équilibre entre ce qui sera éthiquement et socialement acceptable... et sur ce qui ne le sera pas.

Par ailleurs, le champ de la déontologie journalistique ne fait pas que poser des principes généraux communs à l'ensemble de la profession. Certains domaines font l'objet de recommandations particulières, tel que celui de la presse économique et financière. En Belgique francophone, où il est considéré qu'un instrument d'autorégulation constitue un rempart contre d'éventuelles velléités du monde politique à réglementer la profession, de telles recommandations ont été publiées en 2006 par l'Association des journalistes professionnels. Celles-ci figurent dans l'article 15 d'un code de déontologie remanié, en 2013, par le Conseil de déontologie journalistique (CDJ). Considérant les législations belge et européenne en matière de marchés financiers, elles précisent certaines règles "*de comportements et de principes de base relatifs aux opérations d'initiés, à la manipulation des marchés, aux recommandations d'investissements et aux conflits d'intérêts qui doivent être pris en compte par les journalistes*". Elles posent notamment les principes du respect d'un embargo, de précision des informations diffusées et d'interdiction de conflit d'intérêt. La production automatisée d'informations journalistiques serait plus particulièrement concernée par le principe de non-manipulation des marchés, qui consiste à ne pas diffuser d'informations fausses ou trompeuses, et à ne pas exercer d'influence artificielle ou anormale sur les marchés⁴⁴.

2.2.2 Dans le monde de la technique

Lorsque les agents sociaux du monde de la technique conçoivent des systèmes de production automatisée d'informations, ils le font dans le cadre d'une culture professionnelle qui leur est propre et qui ne va pas forcément rencontrer les principes de la culture professionnelle du journalisme. La leur sera connectée aux domaines de l'informatique ou de la science des données. Toutefois, dans les start-ups spécialisées dans la génération automatique de textes, il ne s'agit pas des seuls types d'agents sociaux impliqués dans la conception d'un logiciel ou d'une solution d'automatisation. Par exemple, en Allemagne, la société AEXEA, a été lancée par des journalistes et des communicants. Aux États-Unis, la société Narrative Science a imaginé un profil de méta-journaliste (ou de méta-écrivain), dont la mission est de définir les cadres du récit ainsi que le langage descriptif approprié. Il intervient donc en amont de la génération automatique avec le logiciel comme seul producteur final de l'information (Carlson 2014). Chez

⁴⁴ Source : Conseil de déontologie journalistique, URL : <https://lecdj.be/codededeontologiejournalistique/operations-dinities-manipulation-des-marches-recommandations-dinvestissement-et-conflits-dinterets/>

Syllabs, en France, un moteur de rédaction est paramétré par des linguistes qui définissent manuellement une structure de texte.

Le champ de l'éthique lié au domaine des technologies a pour objet de baliser des pratiques responsables, compte tenu de l'impact social des artefacts technologiques. Il a émergé dans les années 1950, dans la foulée des travaux du philosophe et mathématicien américain Norbert Wiener, considéré comme le père fondateur de la cybernétique (Martin 2014). Wiener prédisait une nouvelle révolution industrielle où la robotique allait remplacer des milliers d'emplois manuels et intellectuels (Van den Hieven & Weckert 2008). Il estimait nécessaire de définir des règles éthiques, de manière à encadrer les agents artificiels capables de prendre des décisions. Sa vision était pessimiste, augurant de "mauvais" usages de technologies qui allaient transformer le monde (Himma & Tavani 2008). Ses interrogations portaient essentiellement sur les conséquences d'une société automatisée, dans un état d'esprit moins utopique que pragmatique.

À la même époque, le britannique Alan Turing s'interrogeait sur l'interchangeabilité homme-machine : "Une machine peut-elle penser?", demandait-il dans "Computing Machinery and Intelligence" (1950). Cet article décrivait un jeu de l'imitation destiné à évaluer si une machine est capable de tromper un être humain en n'étant pas reconnaissable en tant que machine mais en tant qu'être humain. En soulevant cette question, Turing soulignait la résistance humaine dans l'hypothèse d'une réponse positive : "*Nous aimons croire que l'homme est, d'une certaine manière, supérieur au reste de la création*". Toutefois, Turing se concentrait sur les seules capacités d'un ordinateur, pas aux conditions morales dans lesquelles celles-ci se développeraient. Il concluait d'ailleurs son article en disant espérer qu'une machine puisse éventuellement entrer en compétition avec l'homme dans tous les domaines intellectuels.

La question éthique a été ravivée au cours des années 1960, tandis que les sciences de l'informatique enregistraient des progrès significatifs. S'il n'est pas encore question d'établir des normes destinées à encadrer les pratiques de l'informatique, la réflexion est amorcée en interrogeant non seulement l'aspect autonome de systèmes confrontés aux prises de décision, mais aussi la manière d'engager la responsabilité morale des ingénieurs. Aux États-Unis, les premiers programmes d'enseignement axés sur les aspects éthiques de la technologie voient le jour, mais le domaine de l'informatique est moins le moteur de cette démarche que celui de l'environnement, où des chercheurs dénoncent les effets nuisibles d'une industrie faisant usage de technologies en dehors de tout contrôle (Zandvoort *et al.* 2000). En 1967, Marvin Minsky, un autre pionnier du domaine de l'intelligence artificielle, publie "Computation : finite and infinite machines", dans lequel il admet qu'il n'y a aucune raison de supposer que les machines ont des limitations non partagées par l'homme. Il y rappelle la fonction première des machines en tant que modèles physiques de représentations abstraites : résoudre des problèmes. Minsky distingue les machines à états finis (les automates), des réseaux neuronaux qui renferment la notion de "boîtes noires", et les machines infinies, caractérisées par leurs successions de procédures algorithmiques. "*(Les machines) sont créées pour faire de nombreuses*

choses qui rencontrent les standards humains”, souligne Minsky, qui évacue ainsi la question des fins (la morale) car elle n’appartient qu’aux humains, en ne se concentrant que sur les moyens (la technique).

Minsky, pour qui il était important de comprendre les capacités et limites des ordinateurs, soulève alors une problématique aujourd’hui au centre des débats éthiques : celle des “boîtes noires”. L’important, selon lui, n’est pas tant de savoir ce que ces boîtes noires renferment (processus) que la compréhension de la manière dont elles se comportent (résultat). Quelques années plus tôt, en 1961, Wiener définissait en quoi elles se distinguent des “boîtes blanches” : dans les premières, nous n’avons pas de connaissance relative à la manière dont l’opération est réalisée; dans les secondes, la relation entre l’entrée (*input*) et la sortie (*output*) apparaît conformément à un plan structurel défini. Si le débat éthique s’est intensifié depuis le début des années 2010, c’est précisément parce qu’il est motivé par le manque de transparence de systèmes techniques, à propos desquels Latour (2000) estime que plus ils prolifèrent, plus ils deviennent opaques.

Les premières monographies dédiées à l’éthique dans le monde de la technique sont publiées dès le début des années 1970. La généralisation de l’usage de l’informatique puis d’internet contribuent à renforcer ces préoccupations, dans un monde devenu de plus en plus dépendant des technologies de l’information. En Europe, l’éthique de l’ingénierie et de l’informatique deviendront un centre d’attention systématique dès les années 1990 (Zandvoort *et al.* 2000). Harris *et al.* (1996) définissent l’éthique de l’ingénierie comme une éthique professionnelle qui s’oppose à la morale personnelle, en établissant des normes qui s’apprennent dans le cadre de l’enseignement ou dans celui de pratiques professionnelles. Par éthique professionnelle, ils désignent les normes de conduite moralement acceptables que, dans l’idéal, tous les membres d’un groupe professionnel devraient suivre en fonction de leur domaine d’application. Par extension, les normes éthiques relatives aux ingénieurs ne pourraient être les mêmes que celles des journalistes, compte tenu des spécificités de chacun de ces domaines.

L’un des premiers codes de déontologie relatifs à la profession d’ingénieur date de 1912. Il émane de l’American Institute of Electrical Engineers, et dispose notamment que “*l’ingénieur doit tenir compte de la protection des intérêts de son client ou de son employeur en tant que première obligation professionnelle, et il devrait éviter tout acte contraire à ce devoir*”. Zandvoort *et al.* (2000) indiquent que cette tension entre indépendance professionnelle et loyauté bureaucratique est restée prégnante. En 1974, l’Engineers’ Council of Professional Development (ECPD) adopte un texte intitulé “The canons of ethics for engineers” qui dispose, dès ses premières lignes, que “*les ingénieurs doivent tenir pour primordiaux la sécurité, la santé et le bien-être du public dans l’exercice de leurs fonctions professionnelles*”. Cette disposition continuera à caractériser la plupart des codes de déontologie relatifs à l’exercice de la profession d’ingénieur, pris en tenaille entre les exigences de son employeur et la responsabilité qui l’engage envers la société (Zandvoort *et al.* 2000). Il est désormais admis que l’éthique de l’ingénierie, par opposition à la morale personnelle de l’ingénieur, encadre des pratiques professionnelles

et peuvent être communément partagées (Harris *et al.* 1996). Toutefois, elles sont tendues tant par les exigences de l'employeur que par la responsabilité de l'ingénieur. De plus, l'ingénieur ne peut résoudre tous les dilemmes éthiques par lui-même, en raison de la nature multidimensionnelle d'une technologie qui est non seulement façonnée par eux, mais aussi par des agents politiques, économiques et sociaux. L'autorité morale d'un code de conduite éthique peut être délicate à considérer, celui-ci étant susceptible d'entrer en contradiction avec la morale personnelle. De plus, il pourrait être reconnu au sein d'une profession, sans pour autant exercer une forme d'autorité (Davis 1987). Cependant, il peut être considéré qu'un code de conduite constitue une forme de contrat de confiance passé avec les acteurs de la société.

Dans la polysémie des concepts recouverts par le terme "responsabilité", Davis distingue la responsabilité causale performative, de celle liée à l'exercice des compétences professionnelles, de celle relative à l'exercice de pouvoir, et de celle en vertu de laquelle il faut répondre de ses actes. Toutefois, il n'y a aucune raison de sous-estimer l'impact des ingénieurs sur le monde social, et celui-ci peut être autant positif que négatif (Vander Vorst 1998), même si plusieurs voix prétendent qu'un ingénieur ne saurait être tenu pour responsable des mauvaises utilisations de la technologie. Les ingénieurs ont toujours de bonnes raisons de ne pas reconnaître qu'ils ont des responsabilités qui découlent de leurs actes (Davis 2012). En premier lieu, une technologie donnée est rarement le fait d'une seule personne. Ensuite, une conséquence unique peut avoir plusieurs causes. De plus, les personnes impliquées dans un projet sont interchangeables et les contraintes institutionnelles doivent également être prises en compte. *Last but not least*, les ingénieurs sont des personnes sans pouvoir dans une organisation. Mais lorsqu'ils revendiquent la neutralité de la technologie, il est ici opposé qu'aucune technologie ne peut être neutre, dès lors qu'elle est façonnée par des humains pour être utilisée par d'autres humains. Avant d'être technique, la technologie est sociale (Pinch *et al.* 1987).

Lorsque l'objectif d'une technologie est d'être autonome, d'autres questions éthiques vont émerger. Ce domaine particulier est couvert par l'éthique de l'informatique, qui consiste à mettre en œuvre des décisions morales sur des ordinateurs ou des robots (Allen & Wallach 2006). L'éthique de l'informatique résulte de l'analyse de l'impact social des technologies informatiques et concerne les utilisations éthiques de ces technologies (Moor 1985). Selon O'Reilly (1985), un système de contrôle algorithmique présente quatre caractéristiques : (1) une compréhension approfondie du résultat souhaité, (2) une évaluation en temps réel pour vérifier que le résultat est bien atteint, (3) des ajustements algorithmiques basés sur les nouvelles données d'entrée du système, (4) des analyses approfondies effectuées périodiquement pour vérifier que les algorithmes fonctionnent comme prévu. Considérant que les développeurs sont inconscients des effets de leurs réalisations, Palm et Hansson (2006) proposent un modèle d'évaluation qui tient compte de la diffusion et l'utilisation de l'information, des concepts de contrôle et d'influence et de pouvoir, de l'impact social de la technologie et de son impact sur les valeurs humaines. Malgré la malléabilité des programmes informatiques, tout ne peut être contrôlé (Moor 1998). Dans le domaine de la prise de décision basée sur des algorithmes, qui est celui de la production automatisée d'informations, il est important de considérer que

ces technologies sont le résultat de décisions humaines prises en amont (Benjamin 2012). Une procédure algorithmique peut suivre différents chemins et donner des résultats différents, en fonction de la manière dont elle a été programmée. De mauvais jugements peuvent apparaître à différents niveaux : erreur de signal, donnée incorrecte, biais des paramètres (Angler 2013)⁴⁵. En outre, il ne faut pas oublier que l'automatisation est basée sur les principes de l'association homme-agent, selon laquelle l'être humain est responsable des résultats (Linden 2017). Cela signifie aussi que, dans ces processus, les biais liés aux préjugés humains seront probablement automatisés (Dutton & Kraemer, 1980). Mais les biais ne sont pas seulement humains, ils peuvent aussi être institutionnels ou techniques, et ils ne sont pas toujours faciles à identifier (Friedman & Nissenbaum 1996). S'ajoute la problématique relative à l'échec des procédures : bien que généralement prévisibles, elles ne sont pas à l'abri de risques liés à des objectifs irréalistes, des spécifications médiocres, des risques non gérés ou des technologies immatures (Charette 2005). Toutefois, indique Helen Vogt, responsable de l'innovation de l'agence de presse norvégienne NTB, "un logiciel ne fera pas deux fois la même erreur, à condition qu'il y ait un humain derrière pour l'entraîner"⁴⁶.

Dans certains laboratoires de recherche, les discours technocentriques peuvent rester dominants (Sabanovic 2010). Dans le domaine de la robotique, par exemple, la question posée serait moins celle du comment façonner la technologie pour la société, mais de comment adapter la société à la technologie. Plusieurs arguments peuvent expliquer cette posture déterministe : seuls des experts sont capables de comprendre comment fonctionne une technologie complexe, la dimension sociale (contexte, choix, interactions) n'est pas prise en compte au cours du développement, et la vision de l'utilisateur final reste passive. Cela étant, il appartient aux agents sociaux du monde de la technique de rester critiques à propos de leur travail. Dans cet esprit, trois ingénieurs ont signé, à Berlin en 2011, le manifeste de l'ingénierie critique (*Critical Engineering Manifesto*). Le texte, publié en ligne⁴⁷, appelle à une attitude critique, sinon responsable, de l'ingénieur. L'ingénierie y est décrite comme "le langage le plus percutant de notre époque, modélisant nos manières de nous déplacer, communiquer et penser". Le manifeste ne considère pas la technologie comme bonne ou mauvaise, mais appelle à prendre conscience de "notre dépendance à la technologie" comme un défi autant qu'une menace. Ceci peut être considéré comme un appel à un examen autocritique qui témoigne d'une forme de responsabilité morale et individuelle.

Si l'on s'intéresse aux bonnes pratiques à l'œuvre dans les domaines des statistiques et de la science des données, des préoccupations communes au monde du journalisme vont émerger, en termes de qualité des données, de fiabilité de celles-ci et de subjectivité d'une analyse, par essence, interprétative. Dans les lignes directrices pour une pratique de la statistique, l'American Statistical Association⁴⁸ entend sensibiliser sur l'activité d'analyse qui, tout en devant être

⁴⁵ "The ethics of algorithms", Martin W. Angler, SciLogs, 23/10/2013, consulté le 08/08/2014, URL : <http://www.scilog.com/algoworld/the-ethics-of-algorithms-whom-would-you-run-over/>

⁴⁶ "Your next recruit might be a robot", GEN Summit, conférence, Vienne, 15/06/2016.

⁴⁷ Source : <https://criticalengineering.org/fr>, voir aussi annexe 3 "The Critical Engineering Manifesto", p.326

⁴⁸ "Ethical Guidelines for Statistical Practice", Committee on Professional Ethics of the American Statistical Association, avril 2016, consulté le 18/10/2017, URL :

intègre, n'est pas à l'abri de biais ou de subjectivité. Dès lors, "une interprétation objective et valable des résultats exige que l'analyse sous-jacente reconnaisse le degré de fiabilité et d'intégrité des données". Toujours aux États-Unis, la Data Science Organisation propose un code de conduite qui s'appuie notamment sur les concepts de compétence, de connaissance et de rigueur du scientifique, de qualité des données et des preuves récoltées⁴⁹.

Souvent comparées à une boîte noire, les technologies d'automatisation de la production d'informations ne sont pas toujours expliquées par les intentions qui les sous-tendent. Dans la plupart des cas, cela sera dû aux préoccupations de protection des propriétaires de code (Burrell 2016). Si le code source donne de la valeur à son propriétaire (Turrili & Floridi 2007, 2009), certains projets d'automatisation de contenus journalistiques ont été mis à plat par leurs concepteurs. En mars 2018, la société finlandaise de radiodiffusion Yle a ouvert le code qui alimente son "robot" Voitto, qui couvre les résultats de matches de hockey sur glace et ceux des élections en Finlande⁵⁰. Ce code est documenté, en ce compris sur le plan de son processus général⁵¹. Un responsable du studio numérique du développement d'un moteur de rédaction pour l'agence de presse norvégienne NTB a, quant à lui, diffusé les modèles conceptuels mis en œuvre pour couvrir les résultats de rencontres sportives⁵². En Allemagne, la société AEXEA organise chaque année un événement baptisé CLUNC (*Computational Linguistics Unconference*) où elle présente les possibilités de son logiciel de génération automatique de textes et invite les étudiants en linguistique computationnelle à s'en emparer au cours d'un hackaton⁵³. Cette société, de même que la société américaine Automated Insights et que la société britannique Arria NLG⁵⁴, propose également de tester sa solution logicielle en ligne, à partir de données fournies par l'utilisateur-testeur. Mais ici, il s'agit moins d'une démarche de transparence que d'une démarche à visée commerciale.

Considérant l'influence des algorithmes sur la société, il est de plus en plus fréquemment réclamé d'en connaître le code, au nom de la transparence (Diakopoulos 2015). Supposant que ce code soit ouvert, cela nécessiterait de pouvoir le comprendre, d'autant qu'un seul algorithme est rarement mis en œuvre : le challenge n'est donc pas seulement de comprendre le code mais aussi les algorithmes en action (Burrell 2016). Aucune information n'est éthiquement neutre, car elle dépend des principes qui régissent les flux d'informations et ceux-ci concernent à la

<http://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx>

49 "Data science code of professional conduct", non daté, consulté le 18/10/2017, URL :

<http://www.datascienceassn.org/code-of-conduct.html>

50 Consulté sur Github le 21/12/2018; URL : <https://github.com/Yleisradio/avoin-voitto>

51 Voir annexe 4 "Modélisation du processus de Voitto, développé par le radiodiffuseur Yle (Finlande)", p.327

52 Source : <https://bit.ly/2AayPeT>, consulté le 21/12/2018. Voir aussi annexe 5 "Structure pour l'entraînement du "robot" de l'agence de presse norvégienne NTB", p.328

53 Voir annexe 6 "Présentation du processus de génération automatique de textes météorologiques relatifs à des destinations de vacances", p.330

54 La solution logicielle développée par Arria NLG a pour particularité de pouvoir également être utilisée avec des logiciels de *business intelligence*. Selon la société américaine Gartner, Arria Studio était leader sur le marché de la génération de textes en langue naturelle en 2019. Source : "Gartner's first ever Market Guide for NLG declares Arria 'a world leader in NLG'", communiqué reçu via courriel, le 15/07/2019. Voir aussi annexe 7, "Différenciation des solutions de génération automatique en langue naturelle et directions du marché", p.331

fois les données et les informations. De plus, la transparence résulte de choix qui ne relèvent pas seulement de l'éthique mais aussi de contraintes économiques et juridiques (Turilli & Floridi 2009). La responsabilité se place donc au niveau des architectes du système, que ce soit en matière de choix des processus, de la méthodologie appliquée et des paramètres de configuration.

Un système informatique, même s'il a pour vocation d'être autonome, est toujours le fait d'être humains qui le conçoivent et le développent. Ce constat explique que le débat éthique soit prégnant. Aujourd'hui, de nombreux ouvrages sont consacrés à un monde orchestré par une polyphonie d'algorithmes, ce qui aurait pour conséquence de rendre les individus de plus en plus dépendants et de moins en moins critiques. Le terme "algorithme" ne désigne plus seulement une procédure informatique destinée à résoudre une certaine classe de problème en un temps fini, il est également devenu un problème, en raison du manque de transparence des stratégies sociales, culturelles, politiques ou économiques qui le sous-tendent (Lévy 2017). Si un algorithme peut être défini comme une abstraction ou comme une description formelle d'une procédure informatique basée sur des règles (Thomas *et al.* 2018), ses limites sont moins techniques que sociales : elles "*sont déterminées par des engagements sociaux plutôt que par des contraintes technologiques ou matérielles*" (Dourish 2016, cité par Seaver 2017).

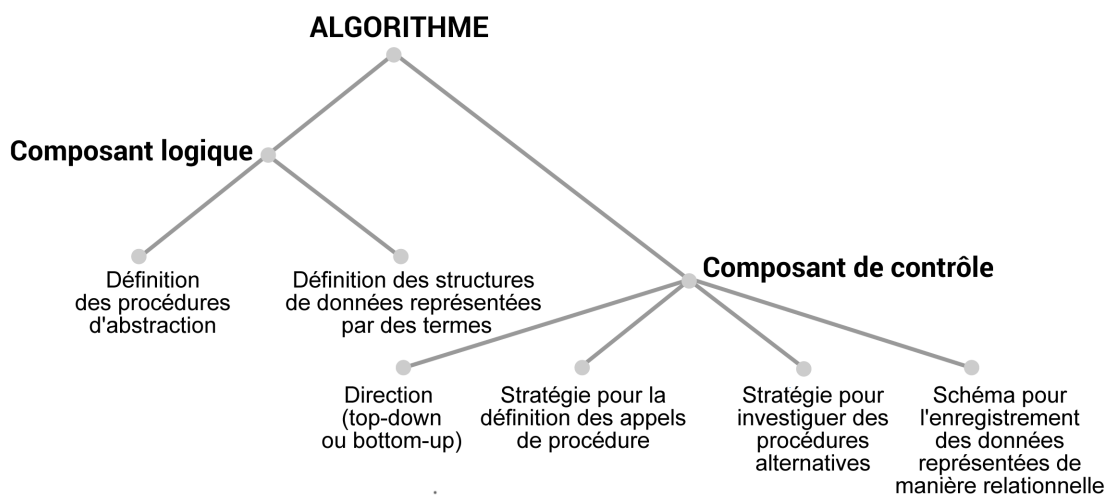


FIGURE 2.10 – Composant logique et composant de contrôle d'un algorithme (Kowalski 1979)

La définition d'un algorithme peut être comprise de manière distincte, selon que l'on soit développeur de logiciels ou que l'on en soit utilisateur : pour le premier, il désigne un ensemble d'opérations et de règles ; tandis que pour le second, il s'agit de processus incompréhensibles et impénétrables (Lewis *et al.* 2018). Boîte noire magique, intellectuellement honnête et mécaniquement efficace : dans le monde de la technique, un algorithme est souvent considéré dans une acceptation idéalisée (Thomas *et al.* 2018). Cela étant, un algorithme est un objet plutôt instable, culturellement construit et qui témoigne de pratiques culturelles (Seaver 2017). Il ne peut donc être envisagé sous le seul prisme d'une procédure informatique, et cela d'autant plus qu'il n'existe jamais une manière unique de résoudre un problème. Pour toutes ces rai-

sons, Kowalski (1979) propose de séparer ou, tout ou moins, de permettre l'identification des composants logique et de contrôle d'un algorithme (Figure 2.10). Il préconise également de séparer les structures de données des procédures qui les interrogent et les manipulent. L'opacité algorithmique jette un doute sur la sincérité des procédures informatisées, pourtant considérée comme mécaniquement fiables, précises et crédibles (Gillespie 2014) alors qu'elles cachent des valeurs et intentions humaines (Kraemer *et al.* 2011).

2.2.3 Définir un cadre commun

Les interactions entre humains et logiciels, dans le contexte de la production d'informations, peut être à la fois comprise comme étant centrée sur des humains issus de mondes sociaux aux pratiques et normes spécifiques, comme un moyen technologique de soutenir le journalisme, comme une technologie "infusée" de principes et pratiques journalistiques, et comme une technologie orientée vers le journalisme (Lewis & Westlund 2016). Cela implique de nouvelles formes de travail et de collaborations devant être considérées sous l'angle de la finalité d'une information : celle d'être destinée à des audiences envers lesquelles un média est engagé socialement. L'examen de la manière dont la responsabilité sociale et son corollaire, l'éthique professionnelle, sont envisagés dans le monde social du journalisme et dans celui de la technique témoigne de préoccupations partagées, mais aussi de points de divergence qu'il s'agirait de négocier dans le contexte de la production automatisée d'informations.

Journalisme et technologies peuvent tous deux être envisagés comme un bien public au service de la société. Cette réciprocité, à la fois matérielle et discursive, suscite des inquiétudes quant à la forme de l'autorité journalistique et aux attentes du public à l'égard de l'information (Carlson 2019). Par conséquent, les mondes du journalisme et de la technique engagent tous deux la responsabilité des agents sociaux qui y sont activement impliqués. Bien que cette responsabilité soit envisagée de manière différente, selon qu'ils relèvent d'un monde social ou de l'autre, il ne s'agit pas d'opérer une distinction aussi nette, dès que lors que l'acte de produire une information résulte d'une activité collective. L'ensemble des agents sociaux participant à ces processus va mettre en jeu des intérêts, parfois contradictoires. Ceci illustre le fait que le journalisme ne peut être considéré comme la seule "affaire" des journalistes (Pereira *et al.* 2018), d'autant qu'il est impossible d'opposer une rationalité de la technologie qui serait purement mécanique à une objectivité journalistique qui apparaît comme illusoire. L'un et l'autre n'agissent pas comme des pôles contraires, mais ces promesses constituent les fondements de leur légitimité.

Dans le monde du journalisme, le principe de responsabilité sociale repose sur la conscience que les professionnels doivent avoir quant à la puissance de la caisse de résonance médiatique. Cela implique un contrat moral guidé par les principes d'objectivité et de respect de la vérité. Si ce premier principe consiste en un idéal professionnel, il participe également à forger une identité professionnelle. Mais il est placé sous tension : d'une part, en raison d'une objectivité qui sera toujours le résultat d'un construit socioprofessionnel; d'autre part, en raison de la double nature d'une entreprise de presse, dont le caractère commercial va contribuer à pi-

loter la ligne éditoriale d'un média d'information. Dans le cadre d'une approche par données dans le journalisme, il est admis que des règles de bonnes pratiques spécifiques devraient être considérées, en raison des manipulations de données que cela implique. Sur le terrain de la production automatisée de contenus journalistiques, s'ajoute une exigence de transparence envers les audiences : ces dernières devraient être informées, dans tous les cas, du caractère non-humain de l'auteur. Cette variable échappe toutefois aux journalistes, dans la mesure où la décision de la divulgation de l'auteur (ou de sa nature) relève de stratégies managériales.

Dans le monde de la technique, la responsabilité sociale est aussi envisagée en fonction de l'impact potentiel des technologies sur les individus et leur environnement. Si des questions éthiques se sont posées aux prémices du développement de technologies informatisées, elles se sont exacerbées, au cours de cette dernière décennie, avec le développement de technologies relevant du domaine de l'intelligence artificielle et de technologies autonomes. La doctrine du double-effet, conceptualisée en 1967 par la philosophe anglo-américaine Philippa Foot (1920-2010), est revenue à l'avant plan. Celle-ci se rapporte aux effets attendus d'une action et aux effets prévus mais non souhaités de celle-ci. Foot citait l'exemple d'un conducteur de tram fonçant à toute allure qui, arrivé à un embranchement, aurait le choix entre poursuivre sa route d'un côté et tuer cinq travailleurs à l'œuvre sur des rails, ou emprunter l'autre branche et tuer un travailleur isolé. Admettons que le tram ne soit pas piloté par un être humain mais par un système informatique – ce qui est devenu un cas d'école –, la dimension éthique prend alors tout son sens.

L'exercice de la responsabilité sociale dans le monde de la technique implique un engagement individuel, sur le plan de la morale, et un engagement collectif, sur celui de normes professionnelles dont il est considéré qu'elles doivent être soumises à l'épreuve de la pratique. Mais la responsabilité de l'ingénieur est limitée, et cela d'autant plus qu'il est considéré qu'il ne peut être tenu responsable de "mauvaises utilisations" de la technologie. De plus, les conditions de la production d'artefacts technologiques ne permettent pas un engagement inconditionnel : cela implique souvent un travail d'équipe, et des procédures peuvent différer alors qu'elles aboutissent à un même résultat. Un ingénieur, qui ne dispose d'aucun pouvoir décisionnel au sein d'une organisation, est d'abord redevable envers son commanditaire et cela le place sous tension, compte tenu de la responsabilité qu'il exerce envers la société. Dans un projet d'automatisation de la production d'informations, les agents sociaux du monde de la technique ne travailleront pas nécessairement en collaboration avec des journalistes, ce qui implique que ce sera à eux de définir les processus éditoriaux qui piloteront les processus techniques. Par ailleurs, certains choix pourraient être susceptibles d'être conditionnés par des contraintes économiques. Par exemple, la personnalisation d'informations, l'une des possibilités des systèmes d'automatisation de la production d'informations, pourra être envisagée dans la perspective de vendre un service sans pour autant s'interroger sur les implications que celui-ci pourrait avoir sur les audiences. En second lieu, le concept d'objectivité apparaît comme un idéal journalistique, tandis qu'il sert d'argument technologique. Mais qu'il s'agisse de journalisme ou de technologies d'automatisation, cette objectivité apparaît comme impossible, dès

lors qu'elle implique des choix humains. Journalisme et technologies ont aussi ceci de commun qu'ils sont d'abord sociaux. Ils peuvent également être abordés par le biais de la "boîte noire", en vertu duquel ils sont le résultat de processus cachés. Ceux-ci sont caractérisés par une succession de choix impliquant des savoirs et savoir-faire relevant du domaine humain. Par conséquent, ils ne peuvent être considérés comme "neutres" ou "objectifs". Et ils impliquent une réflexion sur la manière dont des valeurs humaines sont implémentées dans la conception et les usages de ces technologies (Diakopoulos 2019 :3).

Dans la plupart des cas, il ne sera pas demandé aux journalistes de faire preuve de transparence quant aux raisons motivant les choix qu'ils posent. Cela étant, le développement de pratiques s'appuyant sur la collecte et le traitement de données ont poussé les professionnels à faire preuve de davantage de transparence, notamment en ce qui concerne les sources des données. Pour Ricardo Gutiérrez⁵⁵, les démarches journalistiques pourraient très bien faire l'objet de *making off*, de manière à dévoiler les coulisses d'une enquête, "*mais la transparence journalistique a des limites, à commencer par celle du secret des sources des journalistes. Et tous les journalistes ne souhaitent pas livrer les secrets de leur recette de fabrication*". Pour Paul Bradshaw, "*les évolutions technologiques attirent davantage d'attention sur le plan de la déontologie journalistique, en raison des conflits entre les différentes exigences éthique (...) Le journalisme a toujours fonctionné avec des algorithmes. Mais il est bon de se rendre compte qu'à certains égards, nous sommes contraints de réfléchir davantage sur ces processus lorsqu'ils deviennent encodés dans le cadre d'une programmation*"⁵⁶.

L'objectivité algorithmique se reflète dans l'idéal de l'objectivité journalistique : dans les deux cas, elle s'appuie sur des pratiques et décisions cachées, encadrées par des normes et des pratiques tant individuelles qu'institutionnelles, qui donnent une légitimité à la production de connaissances (Anderson & Kreiss 2013, Geiger 2014). En ce sens, les promesses de l'objectivité algorithmiques sont illusoire car les programmes informatiques ne sont pas développés en dehors de tout contexte social et culturel (Gillespie 2014). C'est pourquoi un processus algorithmique peut être mis en parallèle avec un processus éditorial. Cependant, les choix journalistiques vont davantage relever "*d'une expertise rigoureuse et d'un engagement intense, philosophique et professionnel, à mettre de côté préjugés et convictions politiques*" (Gillespie 2014).

Abordés sous l'angle de l'idéologie professionnelle – laquelle peut être comprise comme une vision partagée de la manière dont le journalisme et les médias devraient fonctionner dans la société (Cohen 1963, cité par Mellado 2015) –, les algorithmes de production automatisée d'informations peuvent être considérés comme l'expression d'une idéologie du journalisme, dans la mesure où un processus d'automatisation va participer à la construction du sens journalistique. Tout comme le journalisme peut échouer à remplir ses missions, la technologie peut également échouer à remplir les siennes, n'étant pas à l'abri d'erreurs (erreurs de logiciels, bugs, virus, ...). C'est en cela que la technologie – ou plutôt le code qu'elle renferme – est fra-

⁵⁵ Ricardo Gutiérrez, communication personnelle, ibidem.

⁵⁶ Communication personnelle, via courriel le 11/08/2014.

gile (McCosker & Milne 2014). De plus, les développements technologiques actuels renferment une part d'incertitude liée à une nature complexe et imprévisibles quant aux manières dont les technologies seront utilisées et aux conséquences de leurs usages. Cela témoigne d'une logique technologique qui, en raison de sa malléabilité, devrait pouvoir être expliquée. Toute procédure devrait également être fondée sur des principes éthiques (Sollie 2007).

Dans le monde de la technique, les processus à l'œuvre dans un système de production automatisée de contenus seront rarement dévoilés. Mais il ne s'agit pas tant d'une question d'opacité que de la volonté de protéger un code source qui donne de la valeur à ces technologies. Pour autant, cette exigence de transparence se justifie par l'intérêt public d'une information. Cela ne concerne pas seulement la nature de l'auteur, mais aussi la source des données et la méthodologie utilisée (Montal & Reich 2017). Hansen *et al.* (2017) abondent dans ce sens, soulignant qu'il faudrait promouvoir tant les bonnes pratiques que la transparence de divulgation. Dans le même temps, ils soulignent que transparence ne signifie pas explicabilité, et que les journalistes devraient d'abord acquérir une expertise en interne, tout en s'appuyant sur des experts du domaine, pour comprendre la manière dont fonctionnent les systèmes d'automatisation ou pour être associés à leur conception : des algorithmes renfermant des décisions éditoriales devraient pouvoir être également compris par un humain. Cela pose la question de la formation des journalistes en matière d'innovations technologiques. Mais si les journalistes font des pas vers le monde de la technique, ceux-ci devraient également être effectués en sens inverse. Lorsqu'ils participent à la chaîne de production de l'information, les agents sociaux du monde de la technique doivent admettre qu'ils font acte de journalisme et qu'ils contribuent à la mise en sens du monde. L'exercice de leur responsabilité devrait, dès lors, se traduire dans le partage des mêmes droits et devoirs que ceux qui prévalent dans le monde journalistique. En outre, s'il n'est pas éthiquement admis que les journalistes puissent travailler à des fins commerciales ou de communication, les technologies développées par les nouveaux acteurs du monde de l'information sont rarement dédiées au seul secteur des médias. Elles peuvent également être utilisées pour décrire des destinations de voyage ou tout autre type de produits ou services destinés à être vendus. Lorsqu'il est recommandé de séparer clairement la communication du journalisme, cela entre donc en contradiction avec les bonnes pratiques en matière de journalisme. Cela pose donc une exigence de transparence quant à la manière dont ces processus opèrent dans un contexte journalistique.

Les deux témoins sollicités dans le volet de cette thèse estiment que le développement de systèmes d'automatisation de contenus journalistiques doit impérativement être abordé sous l'angle de la déontologie professionnelle. Pour Ricardo Gutiérrez, cette évolution fait peur.

"Quand je vois les discussions que l'on peut avoir au Conseil de déontologie journalistique pour établir si ce qu'un journaliste a fait est respectueux ou pas de la déontologie – ce qui est très lié à la manière dont le journaliste a travaillé et au contexte de son travail –, je pense qu'il est quasi impossible de transposer cette réflexion à un logiciel. Mais dans l'absolu, les recommandations destinées aux humains devraient également valoir pour les machines. Vu la puissance de ces logiciels, le risque est d'en privatiser l'usage. Un autre enjeu consiste donc à mettre ces outils au

service de la collectivité et non au service d'intérêts privés, de manière à ne pas créer un monde à deux vitesses où il y aurait des gens bien informés et des gens mal informés"⁵⁷.

Gabrielle Lefèvre, journaliste et membre du Conseil de déontologie journalistique, souligne que les principes déontologiques de base doivent prévaloir. Elle reconnaît, cependant, que "*on va peut-être devoir avancer dans cette réflexion qui implique la responsabilité humaine des éditeurs qui, elle, ne va pas changer : c'est une question de confiance de l'opinion publique. Il faut que le politique soit averti de l'arrivée de ces systèmes, et qu'ils incluent dans les missions de service public le fait que l'information soit bien produite en fonction des principes de la déontologie journalistique. Les services publics de l'information, dans des pays comme les nôtres, ont une tâche essentielle pour préserver la qualité de l'information*"⁵⁸. Elle fait également part de ses craintes quant aux dérives que pourrait entraîner, à long terme, une généralisation de l'automatisation de la production de contenus journalistiques. D'une part, cela concerne le risque d'ubérisation du journalisme, dans le sens où la transmission de l'information ne serait plus le fait des journalistes. "*C'est un peu comme pour le taximan. Il n'est pas intéressant parce qu'il est un meilleur conducteur qu'un conducteur Uber. C'est parce qu'il est lié par un contrat à une série de codes de conduite et d'éthique professionnelle. Toute une série de luttes syndicales a permis qu'il soit payé correctement et dignement. Si on fait la même chose dans les professions intellectuelles, y aura-t-il des tentatives pour annihiler la réflexion et la mise à distance critique? Y aura-t-il une uniformisation de la pensée, dont la perspective est excessivement dangereuse? Qui formatera cette information? Le risque est de ne plus avoir de traçabilité. Qui sera responsable d'une fausse information? Celui qui la diffuse ou celui qui la produit? C'est extrêmement dangereux. Les outils sont des outils. Un mauvais ouvrier, quelle que soit la qualité de l'outil, fera du mauvais travail. Si l'outil est utilisé pour appuyer les pratiques professionnelles, celles-ci seront meilleures. Des journalistes professionnels doivent donc en garder la maîtrise. Si des linguistes ou des programmeurs veulent faire du journalisme, ils doivent être journalistes. Toutes les améliorations techniques dans les rédactions sont encadrées par le rédacteur en chef. On a donné le statut de journaliste professionnel à des gens qui font de la mise en page car ils interviennent dans les contenus. Ils sont donc aussi responsables. S'il y a une erreur, ils sont autant responsables que le journaliste dont le titre de l'article a été modifié. Il n'y aurait pas nécessairement un élargissement de la notion de journaliste mais l'obligation d'appliquer les techniques du journalisme professionnel ainsi que ses bonnes pratiques et usages*". Elle dit également s'inquiéter d'une autre dérive possible : celle d'une information consumériste, induite par une production à grande échelle, qui ne participerait plus à l'idéal journalistique de bien public.

Pour Monti (2019), la surveillance d'un processus d'automatisation de la production d'informations doit devenir une obligation éthique, arguant qu'une absence de suivi ou d'évaluation peut être considérée comme l'un des aspects les plus dangereux du phénomène. Dans le même ordre d'idées, le choix d'un jeu de données qui fera l'objet d'une automatisation doit être guidé par "*le principe éthique d'exactitude, pour choisir une source fiable, en particulier si elle pré-*

⁵⁷ Ricardo Gutiérrez, communication personnelle, op.cit.

⁵⁸ Communication personnelle, Bruxelles, le 16/11/2015.

sente une orientation politique" (Monti 2019). Là aussi, il appartient au programmeur, à l'éditeur ou au journaliste de vérifier l'exactitude et la fiabilité des données. Cela implique que les programmeurs s'engagent dans le respect de règles éthiques qui s'intègrent dans "le nouveau monde technologique du journalisme". Et elles devraient également être appliquées à toutes les formes de journalisme algorithmique. "*Compte tenu du rôle de plus en plus important des ingénieurs dans le domaine de la presse, il est (aussi) nécessaire de réfléchir aux formes de responsabilité éthique et à la responsabilité juridique des programmeurs*" (Monti 2019). Établir un cadre commun suppose aussi de voir en quoi les mondes du journalisme et de la technique convergent, et en quoi ils diffèrent (Figure 2.11).

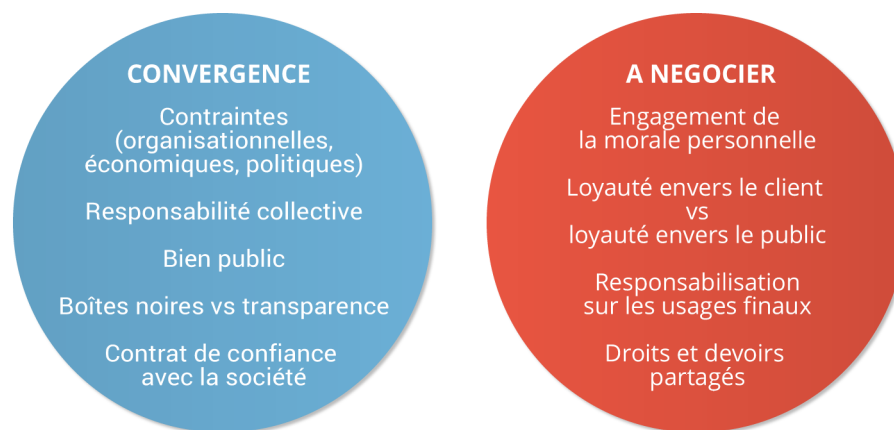


FIGURE 2.11 – Points de convergence et points à négocier entre les mondes de la technique et du journalisme (schéma de l'auteure)

En raison de la responsabilité sociale engageant un média d'information et de toutes les implications qui en découlent, c'est aux agents sociaux du monde de la technique à amener cette réflexion. Cela implique, de leur part, un changement de paradigme : avant d'être techniques, les activités liées à la production automatisée d'informations doivent être d'abord considérées comme journalistiques. Dans la perspective de collaborations fructueuses, chacun devrait également comprendre le métier de l'autre, sa logique, ses contraintes. Cette croisée des chemins entre technologie et journalisme implique donc que les points de divergence observés puissent être négociés, que ce soit sur le plan de la responsabilisation sur les usages finaux de l'information ou sur celui de la responsabilité individuelle, laquelle induit un engagement moral dans l'idéal journalistique, mais aussi sur celui d'une loyauté professionnelle devant nécessairement tenir compte du public auquel s'adresse l'information.

Dans les conclusions du mémoire consacré aux possibilités et limites de la génération automatique de textes (Doerickx 2015), figuraient dix recommandations pour encadrer de manière responsable le phénomène de la production automatisée d'informations. À l'examen de cette première partie de thèse, où sont étudiés tant les enjeux du phénomène que les dimensions de la culture du journalisme, ces recommandations sont ici précisées et étoffées. Elles sont à envisager de manière inclusive, dès lors qu'elles impliquent des journalistes, le management des entreprises de presse et des agents sociaux du monde de la technique. Elles peuvent permettre

de régler les points de divergence entre des mondes sociaux dont les vues peuvent être différentes quant à la manière dont leurs agents exercent leur responsabilité, et elles sont guidées par un souci de transparence (qui est aussi une question de confiance).

1. Les systèmes de génération automatique de contenus journalistiques devraient être envisagés, dans tous les cas, comme un appui au journalisme. Les éditeurs qui font appel à ces systèmes ne devraient pas les entendre dans une logique de réduction des coûts qui serait au détriment de l'emploi journalistique.
2. Tout acteur du monde de la technique impliqué dans la production automatisée d'informations devrait admettre qu'à partir du moment où il participe à la chaîne de production journalistique, il fait acte de journalisme. Dès lors, il devrait en partager les valeurs, dans la perspective d'un cadre communément partagé où s'appliquent les mêmes droits et devoirs que dans le monde du journalisme. De manière à faciliter une compréhension mutuelle, les journalistes devraient être tenus informés des processus à l'œuvre.
3. Les journalistes devraient être associés à la conception des artefacts d'automatisation, de manière à rester des acteurs actifs du processus de production journalistique. En tant que professionnels de l'information, cette maîtrise ne devrait pas leur échapper.
4. Les données, matériau de base à l'information, devraient être traçables. Cela devrait au moins se traduire par la mention du producteur de données. Dans tous les cas, les sources doivent être précises, fiables et à jour. Pour ce faire, des procédures de fact-checking devraient être envisagées.
5. La structure des récits devrait toujours être adaptée aux types de données traitées et à leur domaine d'application, dans la perspective d'une adéquation à leurs usages finaux.
6. Dans le souci de garantir le maintien de la qualité des productions automatisées dans le temps, celles-ci devraient faire l'objet d'un monitoring humain régulier, tant du côté des données que de celui des productions générées de manière automatique. Celui-ci doit donner lieu à une gestion opérationnelle des erreurs, ainsi qu'à une maintenance adaptée.
7. Toute activité d'automatisation qui s'appuie sur le principe de la personnalisation de contenus journalistiques en fonction du profil de l'utilisateur devrait être expliquée et approuvée, de manière explicite, par l'utilisateur.
8. Toute procédure d'automatisation de contenus journalistiques devrait faire l'objet de tests préalables auprès de l'ensemble des acteurs concernés avant de devenir effective, et cela d'autant plus lorsque la production automatisée est diffusée telle quelle aux audiences.
9. Lorsqu'un système d'automatisation de la production d'informations est conçu dans le but de soutenir des routines journalistiques, les contenus publiés devraient, dans tous les cas, faire l'objet d'une validation ou d'une médiation humaine.
10. Tout contenu journalistique généré de manière automatique devrait toujours être présenté en tant que tel aux audiences, en mentionnant explicitement la nature non-humaine de l'auteur, et sans prêter à confusion.

2.3 Dimension évaluative

La dimension évaluative de la culture professionnelle du journalisme met l'accent sur l'autonomie des journalistes, ceux-ci devant être considérés comme libres de toutes formes de pression ou entrave extérieures pour servir aux mieux l'intérêt du public (Chung & Nah 2014). S'ajoute à cet aspect celui du prestige de la profession, lié à la manière dont les audiences la perçoivent. À l'instar des deux autres dimensions examinées dans ce deuxième chapitre, cette dimension participe à forger l'identité professionnelle des journalistes (Singer 2003) mais ici, l'accent est moins placé sur les pratiques professionnelles que sur la manière dont celles-ci sont appréhendées par les journalistes et leurs publics. Le développement de technologies d'automatisation bouleverse une nouvelle fois la manière dont ce qui fait la spécificité du journaliste – laquelle fut déjà remise en question lorsque des blogueurs ou des "journalistes citoyens" se posèrent en fournisseurs de contenus à caractère informatif (Singer 2003, Dagiral & Parasie 2010) –, en induisant dans leur sillage un brouillage, voire une recomposition, des identités professionnelles. Bien que cette dimension évaluative ne soit pas liée aux deux hypothèses pilotant cette recherche – qui disposent que les usages vont dépendre de l'influence de la métaphore du robot journaliste et de l'adéquation avec les exigences et savoir-faire du journalisme –, ce chapitre serait incomplet s'il n'abordait pas la manière dont journalistes et audiences perçoivent les productions automatisées d'informations.

Dans un contexte occidental, trois recherches ont spécifiquement porté sur la manière dont les journalistes réagissaient à l'automatisation de la production d'informations. Van Dalen (2012) a étudié les réactions de journalistes lors du lancement du Statsheet Network (États-Unis), en vue de comprendre comment ils appréhendent les technologies d'automatisation. S'appuyant sur un corpus de 68 articles publiés en 2010 sur des blogs et dans la presse, il a démontré que les journalistes croient que l'automatisation de la production d'informations va changer leur façon de travailler et, par conséquent, les obligera à réexaminer leurs propres compétences. Dans une étude de cas consacrée à la start-up américaine Narrative Science, Carlson (2014) a souligné que le journalisme automatisé contribuait à redéfinir tant le travail des journalistes que leur autorité professionnelle. Son analyse reposait, en partie, sur un corpus de 63 textes publiés en ligne via des sites d'actualités, des sites spécialisés et des blogs entre 2010 et 2014. Il conclut en soulignant les bénéfices de la production automatisée d'informations, laquelle élargit considérablement la quantité d'actualités disponibles et libère les journalistes de tâches répétitives. Thurman *et al.* (2017) ont conduit une étude qualitative, au cours de laquelle dix journalistes britanniques, issus de médias différents (dont la BBC, la CNN et Thomson Reuters) et présentant différents niveaux d'ancienneté, ont été confrontés à l'usage d'un logiciel de génération automatique de textes travaillant à partir de données structurées. Les résultats soulignent que les journalistes ont perçu les caractères limités de l'automatisation, que ce soit en matière de sources d'information ou en matière de sensibilité à l'information. Les journalistes ont aussi critiqué la difficulté d'interroger les données et le manque d'angle humain dans les textes générés. Dans l'ensemble, les réactions des journalistes ont été négatives. Pour au-

tant, ils ont estimé que ces technologies sont amenées à devenir de plus en plus courantes, car elles permettent d'ajouter de la profondeur à un sujet, d'étendre les zones de couverture médiatiques et d'accélérer le processus de publication. Dans le même temps, elles présentent le risque de remplacer "*un personnel coûteux*" dont le travail est "*simplifié et long*". Elles sont également perçues comme présentant les faits "*tels qu'ils sont, sans manipulation*".

Plusieurs recherches convergent sur le fait que les audiences font peu de différence entre des textes rédigés par des journalistes et des textes rédigés de manière automatique. Leurs résultats induisent une autorité désormais partagée entre l'homme et la machine, placés pratiquement sur un même pied d'égalité avec la nuance suivante : les systèmes d'automatisation seraient plus objectifs, et les journalistes seraient de meilleurs rédacteurs. La première d'entre elle a été menée en Suède par Clerwall (2014). L'expérience a consisté à soumettre deux textes, un premier rédigé par un journaliste et un second généré de manière automatique, à des évaluateurs humains composés de 46 étudiants en journalisme. Ceux-ci devaient juger ces deux articles, qui traitaient d'actualités sportives, en fonction de douze variables : cohérence, qualité descriptive, utilisabilité, qualité d'écriture, qualité informative, clarté, qualité de lecture (agréable à lire), intérêt, ennui, précision, fiabilité et objectivité. Les résultats montrent que les textes rédigés par un journaliste obtiennent de meilleurs scores pour les variables "bien écrit", "plaisant à lire" et "clair". Les textes générés de manière automatique réalisent les meilleures performances pour les variables de la qualité informative, de la fiabilité et de l'objectivité. Les évaluateurs humains faisaient peu de différence entre ces deux textes en ce qui concerne la nature de l'auteur, attribuant majoritairement la paternité de ces articles à un humain.

Krahmer et Van der Kaa (2014) ont étudié la crédibilité du journalisme automatisé dans une recherche menée auprès de 232 locuteurs néerlandais, parmi lesquels 64 journalistes. Quatre articles traitant d'actualités sportives et financières étaient soumis à leur appréciation, dont deux générés de manière automatique et deux rédigés par des journalistes. L'objet était d'évaluer leur perception quant à l'expertise et la fiabilité des articles. Les évaluateurs avaient été informés de la présence d'articles automatisés dans l'échantillon qui leur était soumis. Les résultats montrent que la perception des évaluateurs, journalistes compris, est identique pour chaque type de texte. Dans le groupe des non-journalistes, les variables relatives à la fiabilité, la crédibilité et l'expertise étaient perçues de la même manière pour les deux types de textes. Dans le groupe des journalistes, les articles rédigés par leurs pairs étaient considérés comme plus fiables. Les articles traitant d'actualités sportives avaient obtenu les meilleurs résultats.

Graefe *et al.* (2015) ont conduit une étude, s'inspirant de celles de Clerwall et de Krahmer et Van der Kaa, auprès de 986 participants aux profils hétérogènes, dont certains ont fait mention d'une expérience en journalisme. Les résultats de cette recherche, réalisée à plus grande échelle, convergent vers ceux des deux recherches précédentes et tiennent en trois constats : (1) les articles rédigés de manière automatique sont considérés comme plus crédibles, (2) les articles rédigés par des journalistes sont plus plaisants à lire, (3) il existe peu de différences relatives à la perception des variables de crédibilité et d'expertise entre les deux types de textes sou-

mis à évaluation. Haim et Graefe (2017) ont mené deux expériences similaires en Allemagne, basées sur des enquêtes en ligne et rassemblant 618 participants. Si cet échantillon ne peut être considéré comme représentatif de la population allemande, il reflète les couches les plus éduquées de la population, compte tenu des profils répertoriés dans la base de données utilisée. Les chercheurs ont conclu que les lecteurs attendent davantage des contenus rédigés par des journalistes, en termes de lisibilité et de qualité des productions, mais pas en terme de crédibilité : selon eux, les productions automatisées devraient être plus crédibles. Les chercheurs indiquent, à ce propos, que le public posséderait une représentation générale des algorithmes comme étant capables de fournir des informations plus factuelles et donc plus "impartiales". Toutefois, les attentes en matière de qualité n'ont pas été rencontrées par les participants, bien qu'ils aient estimé que les contenus générés étaient plus crédibles, tandis les articles écrits par des journalistes étaient considérés comme plus lisibles. Le concept de qualité était défini en termes de variables largement inspirées par les méthodologies développées par Clerwall et par Kraemer et Van der Kaa.

En Finlande, Melin *et al.* (2018) ont étudié la perception du système Valtteri NLG, qui a généré des articles relatifs aux élections municipales finlandaises de 2017. 152 personnes ont été sollicitées pour évaluer six articles pré-sélectionnés générés de manière automatique, quatre articles rédigés par des journalistes, et quatre articles de leur choix générés via le programme informatique. Les évaluateurs étaient invités à se prononcer sur quatre dimensions en particulier : la crédibilité (évaluation globale de l'objectivité d'un reportage), l'appréciation personnelle (réaction affective liée aux sentiments d'un récepteur qui témoigne d'une expérience agréable), la qualité (niveau d'excellence global du reportage) et la représentativité (représentatif d'une actualité). Dans l'ensemble, les articles rédigés par des journalistes ont enregistré les meilleurs résultats. Le système a, quant à lui, obtenu le meilleur score pour la dimension de la crédibilité. Par ailleurs, les évaluateurs ont attribué la paternité de 21% des textes rédigés de manière automatique à un auteur humain, et de 10% des textes rédigés par des journalistes au programme informatique.

Wölker et Powell (2018) se sont concentrés sur la question de la crédibilité d'articles générés de manière automatique. 300 lecteurs, dont la particularité était de disposer d'une nationalité européenne, se sont exprimés à propos de textes générés de manière automatique et de textes rédigés par des journalistes. Les résultats de cette recherche concluent que les lecteurs font peu de différence sur cet indicateur particulier. Toutefois, les compte-rendus sportifs rédigés de manière automatique ont été jugés plus crédibles. Les évaluateurs se sont également exprimés sur des contenus rédigés par le couple homme-machine et les résultats enregistrés furent jugés concluants. Les chercheurs indiquent également que programmes informatiques et journalistes humains peuvent désormais être placés sur le même pied d'égalité, en ce qui concerne la crédibilité de leurs productions. Ils soulignent également le potentiel du partenariat homme-machine "car il permet aux journalistes humains d'améliorer un contenu automatisé avec des opinions originales et des contenus différents qu'un ordinateur ne peut pas créer".

Conclusions de la première partie

Le phénomène de la production automatisée d'informations ne peut être dissocié de son contexte de développement : c'est dans cet esprit qu'a été abordée la première partie de cette thèse. L'étude proposée dans le premier chapitre témoigne d'imaginaires technologiques nourris par une série d'expériences antérieures où la représentation de l'objet technique a tantôt pu être perçue comme source d'anxiété professionnelle, tantôt comme une opportunité pour développer de nouvelles compétences et, partant, participer à la "réinvention" du journalisme. Confrontés aux innovations technologiques depuis une quarantaine d'années, les journalistes ont souvent fait preuve de résistance : d'une part, parce que l'innovation est un facteur de stress, induisant une charge de travail accrue qui participe à la détérioration des conditions de travail; d'autre part, parce que l'innovation a pu remettre en cause l'idée même de ce qu'est le journalisme, tant sur le plan de ses pratiques, que sur celui de ses idéaux professionnels, alors que la pression du temps et le *multitasking* rendent difficiles la production d'un journalisme de qualité.

Les politiques de convergence, menées dans les rédactions dès l'aune des années 2000, ont eu pour principale conséquence le développement de nouvelles compétences moins axées sur les tâches traditionnelles des journalistes que sur la maîtrise de logiciels visant à mettre en forme le récit médiatique en textes, en sons et en images. Il est désormais admis que les journalistes doivent être capables de satisfaire les exigences d'une logique multimédia. En adoptant les technologies numériques, les journalistes ont modifié les aspects centraux des processus éditoriaux et des routines professionnelles, malgré des attitudes de résistance. Ces résistances peuvent être considérées comme un refus du changement, y compris en termes organisationnels, dès lors que les innovations technologiques les ont placés sous la contrainte du multisupport. Elles ont donc participé à forger des imaginaires plutôt dystopiques.

Dans le même temps, le développement d'une approche par données dans le journalisme, dont la production automatisée d'informations est ici considérée comme un avatar, a mobilisé des imaginaires technologiques liés à un journalisme s'inscrivant dans le temps long de l'enquête. Ceux-ci véhiculent la représentation d'un journalisme d'excellence, perçu comme élitiste en raison du savoir-faire et des compétences techniques qu'il requiert. Circonscrite au territoire des États-Unis et caractérisée par l'application au journalisme de méthodes issues des sciences sociales, l'approche par données dans le journalisme est longtemps restée le fait d'une poignée de journalistes autodidactes et passionnés. Les bases de données informatisées y étaient considérées comme des outils permettant de gagner du temps, que ce soit dans la collecte ou dans

l'analyse de données. Elles ont donné lieu à des partenariats homme-machine qui participent à la construction d'imaginaires plutôt utopiques. L'approche par données dans le journalisme a gagné en popularité à l'échelle mondiale dès le milieu des années 2000, mais elle est restée limitée en ressources et outils, ses pratiques nécessitant l'apprentissage de compétences techniques spécifiques. Elle s'est également fondée sur le mythe d'une information qui serait plus précise, fiable et "objective".

En Belgique francophone, terrain sur lequel se déploie les deux études empiriques présentées dans la deuxième partie de cette thèse, l'approche par données dans le journalisme est envisagée de manière modeste, par manque de temps, de moyens et de retour sur investissement. Ses acteurs sont peu nombreux et il s'agit essentiellement d'autodidactes, formés sur le tas. Le manque de formations spécialisées, couplé à un manque d'intérêt des médias d'information pour des pratiques qui requièrent du temps et des compétences pointues, ne favorisent pas son développement. Cette situation témoigne d'un décalage non seulement avec les pratiques anglo-saxonnes, mais aussi avec celles observées dans plusieurs pays européens, comme en France où ces pratiques sont de plus en plus intégrées dans les rédactions. À cela, s'ajoute le rapport difficile entre les journalistes et les mathématiques, qui ne favorise pas non plus le développement de pratiques journalistiques s'appuyant sur les données en tant que matériau de base à l'information. Parallèlement, sur le plan des pratiques numériques, la professionnalisation gagne doucement certaines rédactions, à travers la constitution de pôles multidisciplinaires où sont intégrés des développeurs, mais où le traitement de données ne constitue par le cœur de leurs activités.

L'évolution des technologies numériques a donné lieu à de nouvelles formes de collaborations, et de nouveaux acteurs sont apparus sur le terrain de l'information. S'ils ne sont pas traditionnellement liés au monde du journalisme, ils participent à l'émergence d'une presse liminaire, dans l'esprit d'une "réinvention" des récits journalistiques en ligne. Dans le même temps, de nouveaux profils sont apparus dans les rédactions, essentiellement aux États-Unis : ceux de journalistes hybrides, à mi-chemin entre les mondes du journalisme et de la technique. Souvent issus des communautés *open source*, leur culture de la transparence est controversée dans le monde du journalisme. Leur double nature professionnelle n'est pas sans poser des questions d'ordre identitaire pouvant apparaître difficiles à négocier – journaliste ou développeur? –, alors que les rédactions tirent encore peu parti du potentiel de ces profils particuliers, qui restent encore rares. En Belgique francophone, un tel profil a été rencontré dans le contexte de cette recherche et il bénéficie du titre de journaliste professionnel, témoignant de la logique inclusive d'une commission d'agrément qui estime que toute activité technique, dès lors qu'elle implique une démarche éditoriale, doit nécessairement être soumise aux mêmes droits et devoirs.

Si les journalistes témoignent d'une longue histoire de relations ambiguës avec les technologies, celles-ci sont exacerbées par le développement des technologies d'automatisation de la production d'informations, à propos desquelles l'influence de l'usage de la métaphore du

"robot journaliste" ne peut être considérée comme neutre. En premier lieu, cette métaphore nourrit des imaginaires rationnels et irrationnels à propos de la fin de l'emploi journalistique. À ce jour, aucun système d'automatisation n'a donné lieu à des licenciements mais l'on retrouve des traces, en France, de pertes de collaborations de journalistes indépendants dans le contexte de l'automatisation de résultats électoraux. Les inquiétudes sont d'autant plus prégnantes que les conditions d'exercice du journalisme connaissent une dégradation constante. Ceci s'inscrit dans le contexte plus large de la crise économique que traversent les médias d'information, qui voient tant leurs audiences payantes que leurs revenus publicitaires s'éroder depuis plusieurs années. En ce sens, la pratique du journalisme est devenue fragile parce que les entreprises de presse le sont aussi.

La dimension symbolique des technologies d'automatisation de la production d'informations ne peut donc être dissociée de l'usage d'une métaphore qui place la machine et l'humain sur un même pied d'égalité. Cela va induire des représentations utopiques, considérant que ces technologies vont permettre de renforcer le rôle des journalistes en leur apportant la possibilité de traiter des volumes de données de plus en plus importants; mais aussi dystopiques, fondées sur l'impossibilité, pour les journalistes, de concurrencer des machines qui seront toujours plus rapides et plus productives. L'analyse du discours médiatique des journalistes, dès lors qu'ils sont amenés à traiter du sujet de la production automatisée d'informations, témoigne de la dualité de ces imaginaires : entre la promesse d'un monde meilleur pour les journalistes et les craintes de la disparition de l'emploi humain, voire de l'identité professionnelle du journaliste. La métaphore du robot contribue à forger des imaginaires rétifs, en raison de la forte charge représentationnelle qu'elle implique. Aussi, il a été observé un besoin, voire une nécessité, pour certains journalistes francophones, de proclamer leur identité et de souligner leur spécificité humaine : "*Cet article n'a pas été écrit par un robot*". Toutefois, le positionnement discursif des journalistes reste globalement en retrait de l'objet de leur discours, tandis que les enjeux professionnels les plus souvent abordés sont ceux d'une relation homme-machine placée sous tension, l'automatisation étant à la fois considérée comme source d'opportunités et de menaces.

Les technologies d'automatisation sont souvent développées en dehors des rédactions, par de nouveaux acteurs qui ne sont pas traditionnellement liés au monde du journalisme mais à celui de la technique. Les journalistes ont été peu associés à ces projets ou de manière limitée. L'argument selon lequel ces technologies vont prendre en charge des tâches répétitives et chronophages, pour permettre aux journalistes de retrouver du temps pour un travail plus valorisant, est régulièrement mis en avant pour justifier l'introduction de ces technologies dans un média. Cependant, leur développement contribue à redéfinir le rôle voire l'identité du journaliste, et cela d'autant plus qu'aucune profession ne peut se résumer à la somme de ses tâches. Le journalisme, ce n'est pas qu'une activité occupationnelle : c'est un savoir-faire, un jeu de relations humaines impliquant des sources d'information, un ensemble de valeurs partagées, une idéologie professionnelle. C'est aussi la manière dont le journalisme fait sens dans un monde social. Ces différents aspects culturels et sociaux ne peuvent être automatisés, et en tout cas pas

en l'état des technologies actuelles qui sont essentiellement fondées sur des systèmes à base de règles, une forme rudimentaire d'intelligence artificielle. Cela étant, les performances des systèmes d'automatisation placent les journalistes face à leurs forces et à leurs faiblesses, et elles remettent en question leurs compétences dans un environnement numérique où les données sont devenues omniprésentes. Aussi, lorsqu'une innovation technologique est susceptible d'affecter des normes et pratiques professionnelles, les journalistes peuvent être invités à redéfinir ce qui caractérise leur métier.

Dans les milieux académiques, de plus en plus de voix défendent l'idée selon laquelle les journalistes devraient développer une forme de pensée computationnelle, en vue de comprendre les processus informatisés et à favoriser les dialogues entre journalistes et informaticiens. Si cette forme de pensée consiste à décomposer des problèmes logiques en séquences, celle-ci peut être mise en parallèle avec des routines journalistiques caractérisées par une succession de choix (sources, angle, récit) en vue de résoudre un problème (celui de raconter une actualité). De la même manière, un processus de production automatisée d'informations peut également être considéré comme un processus éditorial à part entière, visant à répondre aux questions du "quoi dire" et du "comment le dire". Les choix posés ne sont pas très différents de ceux d'un processus éditorial traditionnel, dès lors qu'il s'agit de formaliser un ensemble de règles, de routines et de procédures institutionnalisées qui sont sous-tendues par une expertise professionnelle. Dans la mesure où ils traduisent des intentions humaines et où ils impliquent des jugements de valeurs, aucun de ces choix ne peut être considérés comme "neutre" ou "objectif". C'est pourquoi les agents sociaux du monde de la technique devraient également développer une forme de pensée journalistique, en vue de favoriser une meilleure compréhension des exigences d'un processus éditorial et journalistique. Mais cela implique un changement de paradigme : avant d'être techniques, les activités de production automatisée d'informations devraient d'abord être considérées comme journalistiques, alors que les nouveaux acteurs impliqués dans le développement d'artefacts d'automatisation de contenus journalistiques n'estiment généralement pas faire acte de journalisme.

Ces considérations trouvent un prolongement dans le deuxième chapitre, où il apparaît que l'intégration des savoir-faire et exigences journalistiques dans les artefacts d'automatisation ne peut être réduite à une seule compréhension mutuelle des différents agents sociaux impliqués dans ces processus. Le prisme de la dimension cognitive de la culture du journalisme démontre que cela implique une expertise en matière de traitement et d'analyse des données. Ces activités impliquent une interprétation humaine. Si l'objectivité apparaît comme un idéal dans le monde journalistique, elle est un argument dans le monde de la technique. Dans les deux cas, celle-ci ne pourra être pleinement réalisée : un processus éditorial, qu'il soit automatisé ou non, procède de choix humains. Si le challenge qui est posé par le développement d'un artefact d'automatisation de la production d'informations est double, à la fois technique et journalistique, il ne peut être rencontré qu'à la condition de disposer de données fiables, exactes et exploitables, sans quoi les objectifs journalistiques ne peuvent être rencontrés. Bien que le concept de qualité soit difficile à définir, en raison de son caractère multidimension-

nel, il peut être envisagé à partir des pratiques professionnelles mobilisées dans le cadre d'une approche par données dans le journalisme. C'est sur ce plan que les pratiques techniques et journalistiques vont se rejoindre : des données de mauvaise qualité ne peuvent donner lieu à une information de qualité. Pour autant, il ne faudrait pas oublier que le point de départ de tout processus d'automatisation de la production d'informations – à l'instar de n'importe quel autre type de projet d'automatisation de contenus – implique l'expertise d'un domaine d'application et que celle-ci appartient au professionnel de l'information.

À l'examen de la dimension normative des dimensions culturelles de l'automatisation de la production d'informations, il apparaît que les mondes du journalisme et de la technique se rencontrent, dès lors qu'ils envisagent leurs activités comme relevant d'un idéal de bien public au service de la société. Pour autant, l'exercice de leur responsabilité sociale respective ne va pas sans soulever de nombreux points de divergence qu'il s'agit de régler, considérant que les agents sociaux du monde de la technique font acte de journalisme dès lors qu'ils sont impliqués dans un processus éditorial. Bien qu'ils s'en défendent, les agents sociaux du monde de la technique devraient reconsidérer leur position et s'engager davantage dans le monde du journalisme. Autrement dit, ils devraient en partager les mêmes droits mais aussi les mêmes devoirs, dans la mesure où l'information est une activité qui engage la responsabilité sociale de l'ensemble de ses acteurs. Pour les agents sociaux du monde de la technique, il s'agit aussi de reconnaître leur responsabilité individuelle et de témoigner d'une loyauté professionnelle qui ne soit plus l'exclusive de leur employeur.

De manière à baliser un cadre normatif communément partagé et à encourager de bonnes pratiques, dix recommandations sont proposées dans cette thèse. Elles portent tant sur la manière d'implémenter et de développer un système d'automatisation de la production d'informations, que sur des exigences de transparence ; et elles sont à envisager en incluant tous les acteurs impliqués dans un projet d'automatisation d'informations. À l'examen de la dimension évaluative de la culture professionnelle du journalisme, ceci apparaît d'autant plus comme une nécessité : plusieurs recherches empiriques ont ainsi démontré que les audiences font peu de différences entre des textes rédigés par des journalistes et des textes générés de manière automatique. Elles soulignent également que l'autorité journalistique est désormais partagée entre l'homme et la machine. Cela induit, pour les journalistes, un repositionnement autant professionnel qu'identitaire, l'un des principaux enjeux abordés dans le premier chapitre de cette thèse, dès lors qu'il s'agit de (re)trouver l'équilibre dans une relation homme-machine véhiculant des représentations ambivalentes.

Comment les journalistes vont-ils faire valoir leurs spécificités ? Comment vont-ils garder la main sur des procédures informatisées qui souvent leur échappent ? Les conditions d'un partenariat gagnant-gagnant vont dépendre de la manière dont les journalistes vont envisager celui-ci, tant sur le plan des imaginaires mobilisés que sur celui dont ils vont mettre en avant une expertise qui ne peut être automatisée. Il s'agit là de la principale limite des systèmes d'automatisation de la production d'informations, qui ne font jamais qu'imiter une activité journalistique,

sans être capables d'y adjoindre des informations de contexte, et sans faire preuve de créativité en raison de la standardisation induite par leurs processus. Les systèmes d'automatisation de la production d'informations sont également limités en matière de communication complexe et de pensée experte, bien qu'ils prennent en charge un travail intellectuel et cognitif (Diakopoulos 2019 : 27). C'est sur ce terrain que les journalistes peuvent marquer leur différence, et démontrer leur complémentarité avec des systèmes susceptibles de les libérer de tâches répétitives et chronophages, ou les aider à traiter de manière rapide de larges ensembles de données. Car il s'agit avant tout, de donner du sens à l'information générée : c'est là le propre de toute activité journalistique, qu'elle soit automatisée ou non.

Deuxième partie

Structuration des usages

Introduction aux études de cas

Les objets techniques définissent dans leur configuration une certaine partition du monde physique et social (...) en même temps qu'ils deviennent des médiateurs obligés dans toutes les relations que nous entretenons avec le "réel".

Madeleine Akrich (1987)

Comment les représentations sociales et les pratiques culturelles des journalistes participent-elles à structurer les usages professionnels d'artefacts d'automatisation de la production d'informations? Cette question a été soumise à l'épreuve du terrain empirique, dans deux études de cas réalisées dans deux rédactions belges francophones, sur une période de trois ans. La première expérience porte le nom de "Bxl'air bot". Elle se rapporte à un système d'automatisation développé dans le contexte de cette recherche pour le mensuel *Alter Échos*, lequel avait pour objet de soutenir un travail journalistique au long cours en s'appuyant sur des données publiques relatives à la qualité de l'air en région bruxelloise. La seconde expérience a pour nom de code "Quotebot". Il désigne un système d'automatisation qui vise à soutenir les routines quotidiennes des journalistes du service "Investir" du quotidien *L'Echo*, en leur fournissant des brouillons à partir de données relatives aux marchés boursiers. Ses autres particularités sont d'avoir été développé par un prestataire externe au média, la start-up française Syllabs, et d'avoir été financé par le fonds Google pour l'innovation numérique dans les médias (Digital News Initiative, DNI).

La matérialité de ces deux dispositifs sociotechniques, qui peuvent être envisagés à la fois comme des objets et des outils du journalisme, s'exprime dans un croisement entre journalisme et technologie, au sein duquel un processus de médiation sociale et culturelle va être mis en œuvre (Manovich 2013, Cox *et al.* 2013, Berry 2011, Anderson 2011, Anderson & De Maeyer 2014, Rodgers 2015). Aussi, ne peuvent-ils être considérés comme s'appuyant seulement sur la transformation d'entrants en extrants : comme n'importe quelle autre forme de technologie, ils ne peuvent être séparés du contexte socioculturel dans lequel ils se sont déployés, et ils ne peuvent être compris qu'à travers leurs dimensions sociales, culturelles et organisationnelles qui forment leur contexte d'usage ((McCarthy & Wright 2004, Geiger 2014, Proulx 2006,

Akrich 1993a, Akrich 1993b). Les référents sociaux et culturels des acteurs impliqués dans ces deux projets sont hétérogènes, mais ces derniers sont considérés comme s'inscrivant dans le domaine des communautés de pratique, où des valeurs communes sont partagées, tout en côtoyant des valeurs individuelles (Schmitz-Weiss & Domingo 2010).

Bxl'air bot	Quotebot
Données publiques diffusées en temps réel	Données privées (payantes) diffusées en temps réel
Automatisation de la collecte et du traitement de données par la chercheuse	Automatisation de la collecte et du traitement de données par la start-up française Syllabs
Rédaction peu habituée à traiter avec des données chiffrées (peu d'appétence)	Rédaction habituée à traiter avec des données chiffrées (appétence)
Magazine de niche (social), moins de 700 abonnés	Quotidien généraliste spécialisé en économie et finance, plus de 72.000 lecteurs
Structure : association sans but lucratif	Structure : société commerciale
6 journalistes impliqués dans l'expérience	6 journalistes impliqués dans l'expérience + responsable du développement et de la transformation numérique
Pas d'extension de l'expérience	Extension de l'expérience au quotidien néerlandophone <i>De Tijd</i>
La rédaction a été sollicitée dans le contexte de cette recherche	La rédaction a sollicité la chercheuse en tant que consultante
Outil du journalisme : soutenir un travail journalistique s'inscrivant dans le temps long	Outil du journalisme : soutenir des routines journalistiques s'inscrivant dans le temps court de l'immédiateté
Objet du journalisme : matérialité via la plateforme automatisée www.bxlairbot.be	Objet du journalisme : matérialité via la diffusion des productions soumises à la validation des journalistes dans la rubrique "Market Live" du site web lecho.be (information boursière généraliste), et via une diffusion dans l'espace de gestion personnalisée d'un portefeuille d'actions virtuel réservé aux abonnés (information servicielle)
Durée de l'expérience : un an	Durée du projet : initialement un an, étendu à deux ans

Tableau 2.3 – "Bxl'air bot" et "Quotebot" : caractéristiques communes et particulières

Ces deux études de cas se déploient sur le terrain de la presse belge, où le phénomène connaît ses premiers développements. Elles s'inscrivent dans le contexte d'un marché étroit et économiquement fragile, où les pratiques journalistiques relevant d'une approche par données sont limitées. Bien que tout semble séparer les deux rédactions étudiées, que ce soit en termes de taille ou de modèle d'affaires, le point de convergence est à trouver dans la manière d'envisager les artefacts technologiques, en tant qu'appuis à des activités journalistiques. L'étude des

usages des artefacts d'automatisation tient compte de ces deux environnements de développement, lesquels présentent donc des caractéristiques communes et spécifiques (Tableau 2.3). Dans les deux rédactions étudiées, la construction sociale des artefacts d'automatisation s'appuie sur une approche inclusive où les agents du monde du journalisme, en tant qu'utilisateurs finaux, ont été invités à devenir des acteurs de l'innovation technologique, en s'impliquant activement dans la conception des artefacts destinés à soutenir leurs routines professionnelles. Lorsque des utilisateurs finaux sont embarqués dans un processus de conception, devenant de ce fait des acteurs de l'innovation, cela peut être considéré comme une première forme d'usage (Akrich 2006).

Sur le plan conceptuel et théorique, le modèle SCOT (*Social Construction of Technology*)⁵⁹ a servi de fil conducteur pour l'analyse des observations relatives au développement des deux systèmes d'automatisation. Ce modèle a été mis en parallèle avec le modèle incrémental en cascade du cycle de vie du logiciel (Ghezzi *et al.* 2002), de manière à suivre le développement de chaque artefact sur le plan de leurs processus techniques et sociaux (Figure 2.12).



FIGURE 2.12 – Mise en correspondance des étapes du modèle en cascade du cycle de vie du logiciel avec le modèle SCOT

Considérant que chacun de deux artefacts d'automatisation de la production d'informations résulte d'un construit social (Bijker, Hugues & Pinch 1987, Flichy 1995), et qu'ils disposeront chacun de leur propre matérialité (Berry 2011, Manovich 2013, Cox & McLean 2013), leur développement est également envisagé sous l'angle d'une pratique culturelle structurant des interactions au sein d'environnements médiés (Geiger 2014, MacKenzie 2006). Il s'agit donc de proposer un point de vue complémentaire pour mieux comprendre un phénomène encore peu étudié dans la perspective de ses usages.

L'analyse des usages (et des non-usages) s'adosse, quant à elle, à une tradition de recherche ancrée dans la sociologie des usages. La formation des usages, qui relève de mécanismes complexes, dépend de constructions individuelles et collectives, ainsi que de l'imaginaire lié à l'objet lui-même (Jouët 2000), dont la logique résultera d'une confrontation entre l'instrument, sa fonction et le projet de l'utilisateur (Perriault 1989, cité par von Pape et Martin 2010). Pour autant, les usages ne pourront être déduits mécaniquement des choix effectués lors du processus de conception (Akrich 1990). Bien que la sociologie des usages soit largement mobilisée dans le cadre de l'analyse des différentes formes d'usages observées, la problématique des non-usages est également abordée dans la perspective plus générale d'une sociologie de l'innovation.

⁵⁹ Voir "Cadre conceptuel et théorique", p.12

Étant donné que leur nature et leur contexte de développement présentent des caractéristiques sociales, culturelles et organisationnelles qui leur sont propres, l'étude des deux systèmes d'automatisation n'a pas été déployée tout à fait à l'identique dans les deux rédactions étudiées. Aussi, des méthodes particulières de recherche ont-elles été développées dans chaque étude empirique. Celles-ci concernent, plus précisément, l'évaluation de l'adéquation aux savoirs et savoir-faire journalistiques⁶⁰. Dans l'expérience "Bxl'air bot", les données publiques nourrissant le système d'automatisation étaient extraites de pages web : elles présentaient donc de potentiels problèmes de qualité. Considérant que sans données fiables et précises, les informations générées ne pouvaient pas être fiables et précises, le modèle conceptuel d'évaluation développé dans le deuxième chapitre de cette thèse⁶¹ a été mis en œuvre. Dans cette première étude de cas, la problématique de la qualité des données se posait donc en préalable aux usages qui, pour autant, ne furent pas nécessairement mobilisés. Dans l'expérience "Quotebot", s'agissant de brouillons automatisés pouvant être publiés tels quels ou faire l'objet d'un retraitement journalistique, il s'agissait de voir en quoi les textes générés de manière automatique correspondaient aux exigences posées par les journalistes lors de la conception de l'artefact. Il s'agissait également de voir en quoi l'expérience des textes générés de manière automatique était satisfaisante (ou non) pour ses utilisateurs finaux. C'est pourquoi le modèle d'évaluation de cet artefact s'appuie sur des concepts issus des champs de l'expérience utilisateur et de la linguistique computationnelle. Ces méthodes particulières de recherche sont détaillées au début de chaque chapitre, dans la section relative au contexte de l'expérience.

⁶⁰ Ceci fait référence aux dimensions cognitive et normative de la culture professionnelle du journalisme, développées dans le chapitre 2, "Entre deux mondes : les dimensions culturelles de l'automatisation" p.96

⁶¹ Voir p.114

3 | L'artefact "Bxl'air bot" dans les usages de la rédaction d'*Alter Échos*

"Bxl'air bot" est une application automatisée qui a diffusé en temps réel, pendant un an, des informations relatives aux mesures des taux de polluants atmosphériques réalisées en région bruxelloise, dans des représentations textuelles et visuelles. Nourri par des données publiques diffusées en ligne, le système d'information avait également pour objet l'enregistrement de ces données, en vue de fournir aux journalistes des niveaux d'analyse complémentaires pour nourrir une démarche de journalisme au long cours. Bien que les technologies d'automatisation puissent convenir pour effectuer des tâches répétitives, elles ne peuvent se substituer à l'humain, dès lors qu'il s'agit d'apporter au récit journalistique une dimension contextuelle indispensable à l'explication des données, que ce soit en amont (à propos des causes de la pollution atmosphérique) ou en aval (à propos de ses conséquences).

Le projet "Bxl'air bot" a été proposé à la rédaction du mensuel *Alter Échos* dans le contexte de cette recherche. Il ne s'agissait pas d'un terrain inconnu, en raison de notre engagement professionnel, depuis 2015, dans diverses collaborations avec l'agence Alter, sa structure faitière. Plusieurs constats avaient nourri cette démarche : il n'existait pas, au moment d'engager cette recherche, d'artefacts d'automatisation de la production d'informations en Belgique francophone ; la problématique de la qualité de l'air à Bruxelles fait partie des sujets traités par la rédaction du mensuel ; des données existent mais leurs modes de diffusion relèvent de dispositifs communicationnels ou scientifiques ; une expérience d'automatisation de la production d'informations traitant de la qualité de l'air est menée en Allemagne, depuis 2014, par le quotidien *Berliner Morgenpost*, démontrant que ce type de données fait partie du champ des possibles dans le contexte de la production automatisée d'informations¹. Cette étude de cas a été envisagée en tant que double opportunité : d'une part, il s'agissait de faire l'expérience d'un développement où des challenges spécifiques se posent sur les plans technique et journalistique ; d'autre part, elle permettait de disposer d'un objet d'étude original dans une rédaction présentant peu d'appétence pour les technologies numériques ainsi que pour une approche par données dans le journalisme. Notre position embarquée, bien que caractérisée par une tension entre engagement et distance², a permis de disposer d'un poste d'observation privilégié et, ce faisant, d'accéder à un matériel de recherche qu'il n'eût pas été possible d'obtenir autrement.

¹ Voir aussi p.104

² Voir "Implication personnelle", p.28

Deux cheminements méthodologiques ont été entrepris dans cette étude de cas. Le premier relève d'une démarche de recherche-action expérimentale, où sont alliées théorie et mise en pratique (Goyette & Lessard-Hébert 1987), en vue de développer une application automatisée qui tienne compte des exigences d'exactitude et de précision, deux des enjeux majeurs dans le cadre d'une approche par données dans le journalisme. Il s'agissait également de comprendre comment rencontrer ces exigences, avant d'expliquer les situations sociales (Barbier 1996 :14), ce qui constitue le cœur de ce troisième chapitre et le second cheminement méthodologique. Le développement de la plateforme automatisée a été nourri par la littérature relative aux bonnes pratiques en matière de conception d'interfaces et d'ergonomie – dont l'étymologie provient du grec *ergon*, qui signifie "travail" et *nomos*, qui veut dire "loi" –, une discipline qui se rapporte à la conception de dispositifs technologiques adaptés aux caractéristiques et aux limites des êtres humains (Geslin 2002). S'il a consisté en une mise en pratique de notre expérience professionnelle antérieure à cette expérience de recherche, ce développement de l'artefact a également permis d'éprouver le modèle conceptuel d'évaluation de la qualité des données, présenté dans le deuxième chapitre de cette thèse³.

La collecte des données empiriques comprend :

- des méthodes d'enquête par questionnaire soumis aux journalistes participant à cette expérience (trois formulaires accessibles en ligne, dont deux de manière anonyme)⁴ ;
- douze entretiens semi-dirigés (deux entretiens avec chacun des cinq journalistes salariés de la rédaction et une journaliste indépendante spécialisée dans les questions relatives à la qualité de l'air, le second entretien ayant été réalisé en fin d'expérience⁵) ;
- des observations participantes organisées dans un cadre formel (participation à trois réunions de rédaction et organisation d'une formation d'initiation à l'utilisation de la base de données de l'artefact et d'un tableur, dans la perspective d'un traitement humain des résultats de "Bxl'air bot") ;
- 84 échanges de courriels avec la coordinatrice de la rédaction, et 55 échanges de courriels avec la journaliste indépendante impliquée dans le projet.

En vue de documenter le contexte de l'expérience, trois entretiens portant sur l'histoire et le développement d'*Alter Échos* ont été réalisés avec trois fondateurs de l'Agence Alter (Arnaud Grégoire, Thomas Lemaigre et Pierre Verbeeren). Dans la mesure où la compréhension du domaine d'application – qui relève du champ scientifique – constituait un préalable au développement de l'artefact, des sources académiques ont permis de le documenter. Plusieurs experts ont également été sollicités : deux représentants du BRAL, une association bruxelloise qui milite pour "un Bruxelles durable", et deux représentants de la Cellule Interrégionale de l'Environnement (CELINE), l'organisme public qui assure la mission de diffusion des données relatives à la qualité de l'air en Belgique. Ces échanges, qui avaient également pour objet de documen-

³ Voir p.114

⁴ Voir annexe 8, "Formulaires d'évaluation diffusés via Google Forms dans le cadre de l'expérience 'Bxl'air bot'", p.332

⁵ Voir annexe 9, "Guide d'entretien pour l'évaluation finale de l'expérience 'Bxl'air bot'", p.334

ter un domaine d'application avec lequel nous étions peu familière, n'ont pas été mobilisés dans ce chapitre, en dehors des informations relatives au cycle de vie des données qui furent essentielles pour le prototypage de l'artefact.

3.1 Contexte de l'expérience

La problématique de la qualité de l'air est devenue une préoccupation prégnante au cours de ces dernières années, en région bruxelloise, considérant que la pollution atmosphérique représente un risque environnemental majeur pour la santé⁶. L'Union Européenne contraint ses membres à respecter des normes en matière de taux de polluants atmosphériques, lesquelles consistent en des moyennes horaires ou annuelles à ne pas dépasser⁷. Fruits d'un compromis entre les états membres de l'UE, elles consistent en des limites *a minima* (Tableau 3.1). Chaque pays est soumis à l'obligation d'organiser les mesures de ces taux de polluants sur son territoire, de les communiquer au public, et d'en informer la Commission européenne via un *reporting* annuel. Le respect de ces normes engage donc la responsabilité d'une politique nationale, alors que la Belgique se trouve régulièrement en situation d'infraction, en ce compris en Région de Bruxelles-Capitale.

PM10	PM2.5	O3	NO2
50 µg/m ³ (moyenne sur 24 heure, 35 jours maximum/an)	Pas de valeur à court terme	120 µg/m ³ (moyenne sur 8 heures, pas plus de 25 jours en moyenne/3 ans)	40 µg/m ³ (moyenne annuelle)
40 µg/m ³ (annuelle)	25 µg/m ³ (moyenne annuelle)	180 µg/m ³ (moyenne horaire, seuil d'information), 240 µg/m ³ (moyenne horaire, seuil d'alerte)	200 µg/m ³ (valeur limite horaire, pas plus de 18 fois/an)

Tableau 3.1 – Normes de l'Union européenne relatives aux taux de polluants atmosphériques

Compte tenu de l'incidence sanitaire de la formation d'ozone (O3) ainsi que des émissions de dioxyde d'azote (NO2) et de particules fines (PM 10 et PM 2.5), l'Organisation Mondiale de la Santé (OMS) publie également ses propres recommandations. Celles-ci consistent en des

⁶ Source : Organisation Mondiale de la Santé, "Qualité de l'air ambiant et santé", consulté le 24/06.2017, URL : <http://tiny.cc/qualiteairOMS>

⁷ Directive 2008/50/CE du Parlement européen et du Conseil du 21 mai 2008 concernant la qualité de l'air ambiant et un air pur pour l'Europe.

taux moyens, par type de polluant, qu'il conviendrait de ne pas dépasser. Seules les particules fines de type *black carbon* (carbone suie) ne font l'objet d'aucune recommandation spécifique, alors qu'elles sont reconnues comme potentiellement plus dangereuses pour la santé que les autres types de microparticules (Janssen *et al.* 2011). Ces recommandations portent sur des moyennes flexibles dans le temps (Tableau 3.2). Plus strictes que les normes européennes, elles n'ont aucune valeur contraignante.

PM10	PM2.5	O3	NO2
50 µg/m ³ (moyenne sur 24 heures)	25 µg/m ³ (moyenne sur 24 heures)	100 µg/m ³ (moyenne sur 8 heures)	200 µg/m ³ (moyenne horaire)
20 µg/m ³ (moyenne annuelle)	10 µg/m ³ (moyenne annuelle)	Pas de recommandation	40 µg/m ³ (moyenne annuelle)

Tableau 3.2 – Recommandations de l'OMS relatives aux polluants atmosphériques

Recommandations de l'OMS et normes européennes ont ceci de commun qu'elles présentent un caractère évolutif dans le temps : chacune des institutions est susceptible de revoir ses taux régulièrement, témoignant du caractère arbitraire de la norme. Le domaine d'application de la qualité de l'air a aussi pour particularité de se trouver régulièrement au centre de controverses. À Bruxelles, des associations de défense de l'environnement interpellent régulièrement le monde politique, tantôt remettant en cause la méthode de transmission des données relatives aux taux de polluants, tantôt réclamant des engagements plus stricts, dans un souci de santé publique. Les médias d'information traitent régulièrement de ces actualités, qu'il s'agisse de relayer des préoccupations citoyennes, de dresser un bilan politique ou d'alerter sur des phénomènes tels que les smogs ou les pics d'ozone. Ce faisant, ils participent à faire écho aux controverses (Cefaï 1996). Dans une perspective sociologique, la problématique de la qualité de l'air présente les caractéristiques d'un problème public, lequel consiste en la transformation d'un fait social en un enjeu de débat public et/ou d'intervention étatique. "*Le fait se métamorphose alors en problème au terme d'un processus (...) (définissant) une expérience comme une offense, une situation génératrice de litige*" (Neveu 1999). Un problème public "*naît de la conversion d'un fait social en objet de préoccupation et de débat, éventuellement d'action publique*" (Neveu 2015).

3.1.1 Chaîne sociotechnique et cadre technologique

La chaîne sociotechnique constituée autour de "Bxl'air bot" est principalement composée des membres de la rédaction d'*Alter Échos* qui, dans cette expérience, sont envisagés comme les premiers utilisateurs finaux de l'application en ligne. Mais elle est aussi composée de groupes sociaux hétérogènes qui participeront ponctuellement à la construction sociotechnique de l'artefact (le groupe social du monde l'administration publique et les audiences d'*Alter Échos*) ou à sa diffusion (le groupe social du monde associatif, et le groupe social des journalistes non apparentés à la rédaction qui ont fait l'usage de l'artefact en tant que sujet de reportage). Ces agents sociaux sont considérés comme secondaires dans le contexte de cette recherche, qui traite des pratiques journalistiques, et ne font donc pas l'objet d'un traitement particulier. Toutefois, il était important de les mentionner, de manière à comprendre la dynamique qui s'est créée autour de l'artefact. Dans ce réseau, notre rôle a été celui de porte-parole et de médiatrice du système d'information (Figure 3.1).

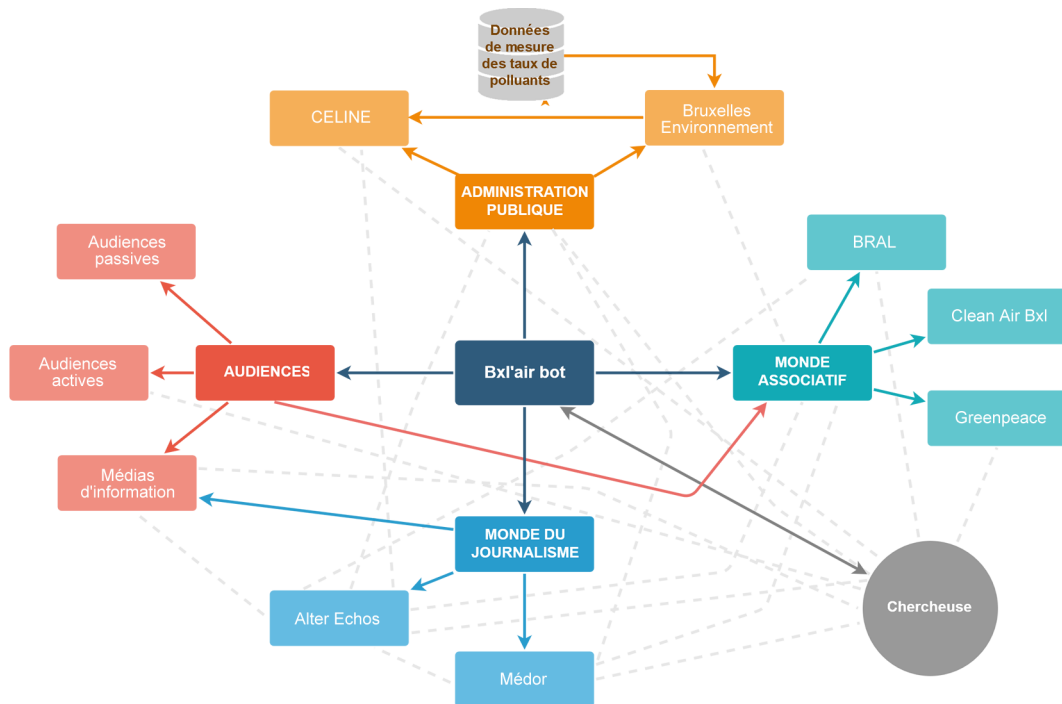


FIGURE 3.1 – Schéma conceptuel de la chaîne sociotechnique formée autour de l'artefact "Bxl'air bot"

La cartographie du réseau témoigne de l'ampleur que celui-ci a pu prendre au cours de l'expérience et de la variété d'interactions entre les différents groupes sociaux gravitant autour de l'artefact "Bxl'air bot". Il s'agit là du positionnement de ce dernier en tant qu'objet du journalisme. Les agents sociaux considérés comme appartenant à des groupes secondaires ont été essentiellement sollicités dans le contexte de cette recherche (agents du monde associatif et du monde de l'administration publique), en tant qu'experts du domaine d'application. Vue sous la perspective des usages de l'artefact, en tant qu'outil du journalisme, la chaîne sociotechnique est exclusivement composée des cinq journalistes de la rédaction d'*Alter Échos*; d'une

journaliste indépendante spécialisée dans les questions relatives à la qualité de l'air qui, au moment de lancer cette recherche, travaillait également sur ce sujet pour le trimestriel *Médor*; de la chercheuse et de l'artefact "Bxl'air bot". Aucun des journalistes impliqués dans cette expérience ne présente un profil professionnel lié à une approche par données dans le journalisme, et ne peut être considéré comme particulièrement technophile. Ces journalistes évoluent dans le contexte d'une rédaction pouvant être considérée comme occupant une place particulière dans le paysage médiatique belge, tant par sa structure que par ses activités de niche, engagée dans le choix des sujets qu'elle traite mais défendant des idéaux professionnels de "neutralité".

3.1.1.1 Le magazine *Alter Échos*

Alter Échos est une émanation de l'Agence Alter, née dans le giron de la rédaction bruxelloise du magazine liégeois C4⁸. Nous sommes en alors 1995. À l'époque, le projet rassemble une poignée de jeunes journalistes ou de militants engagés "*qui avaient envie d'en découdre avec la société*", explique Pierre Verbeeren⁹, directeur d'Alter jusqu'en 2002, aujourd'hui secrétaire général de l'ONG Médecins du Monde. Arnaud Grégoire, un autre fondateur du projet, aujourd'hui producteur indépendant, précise que l'envie était "*de faire les choses autrement (...), d'avoir une démarche sociale et de traiter de l'innovation sociale*"¹⁰. "*On a trouvé des gens qui avaient des choses à raconter, qui avaient une lecture très dure de la société*", précise Pierre Verbeeren. De leurs réunions, qui s'égrenèrent pendant quelques mois rue Potagère à Saint-Josse, va naître un premier projet. "*On s'était dit qu'on allait reproduire le modèle liégeois, en faisant de l'éducation permanente et de la formation. On voulait aussi créer de l'emploi. La toute première ASBL créée dans ce cadre s'appelait EFPE, pour 'Éducation, Formation, Production, Emploi'. Certains trouvaient ce nom affreux. On a été voir l'administration bruxelloise avec notre projet et avons reçu un subside de 650.000 francs belges (environ 16.000 euros, non indexés, ndlr) pour le volet production du projet, où il s'agissait de travailler sur les enjeux de l'insertion socioprofessionnelle, qui étaient nouveaux à l'époque*", poursuit Pierre Verbeeren.

L'ASBL Agence Alter est créée dans la foulée. Le numéro zéro d'*Alter Échos* – "Les échos d'Alter" – est daté du 26 novembre 1996. Le premier numéro paraît le 24 février 1997. Il prend la forme d'un bulletin d'information envoyé gratuitement aux professionnels du secteur social. À la Une, l'équipe signe une lettre destinée à ses futurs lecteurs : "*Alter Échos est conçu comme un outil de travail. Votre outil de travail. Il veut rencontrer des préoccupations prioritaires : votre efficacité et les droits des publics qui vous rencontrent au quotidien*". Pour Pierre Verbeeren, c'était "*un travail de dingue. On écrivait nos articles et la nuit, on rassemblait les photocopies du numéro pour les mettre sous enveloppe*". Une seule personne était alors salariée à temps partiel : Thomas Lemaigre, un journaliste qui fait également partie des fondateurs d'*Alter Échos*.

Les envies militantes exprimées lors des réunions qui présidèrent à la naissance du projet trouvent une expression journalistique qui consiste "à faire parler les faits" : "*On s'était dit qu'il*

⁸ Ce mensuel avait été créé en 1992 par des journalistes au chômage pour un public de chômeurs. Il existe toujours aujourd'hui.

⁹ Communication personnelle, Bruxelles, 04/12/2018.

¹⁰ Communication personnelle, Bruxelles, 11/09/2017.

ne fallait pas prendre position. Si on racontait la réalité, ce serait d'un subversif décapant", souligne Pierre Verbeeren. L'Agence Alter fait alors ses débuts en tant qu'agence de presse. *"L'idée était de produire un journal, tous les quinze jours, pour un lectorat d'abonnés. Une partie de ces contenus était faxé aux rédactions, sous la forme de dépêches. Les abonnements ont très vite décollé. Il y avait besoin, à l'époque, d'une information factuelle, bien faite et indépendante des traditionnels piliers que sont les syndicats ou les administrations",* se rappelle Thomas Lemaigre¹¹, salarié pendant seize ans à l'Agence Alter – successivement comme journaliste, rédacteur en chef et directeur –, aujourd'hui codirecteur de *La Revue Nouvelle*.

À l'époque, le modèle économique de l'association dépend exclusivement de subsides. Il s'agit là de sa plus grande faiblesse. Le nombre d'employés de l'Agence évoluera en fonction des projets subsidiés, et ceux-ci ne vont pas toujours relever du domaine du journalisme. *"On a eu assez vite d'autres boulots qui se trouvaient à la frontière avec le journalisme, voire qui ne relevaient plus du tout du journalisme. On a fait des comptes-rendus de colloques, des notes synthèse, des briefings. On disait 'oui' à tout : cela permettait de dégager des marges pour faire notre gazette. À l'époque, on s'était dit que nos ressources financières devaient être de trois types : subsides, abonnements et contrats pour développer des outils sur les questions sociales",* explique Thomas Lemaigre. Pour Pierre Verbeeren, il fallait y voir la force de la jeunesse. *"On ne doutait de rien. En 2003, quand j'ai quitté Alter, on avait réussi à créer seize emplois et on travaillait avec une quarantaine d'indépendants. Nous avons fort diversifié nos activités. On publiait 'Alter Educ', consacré au secteur de l'éducation; 'Alter Business News', qui traitait de sujets relatifs à la responsabilité sociale des entreprises; 'Alternative Business', qui consistait en un travail d'accompagnement des chômeurs dans la création de leur activité; et 'Alter EU', une sorte de société de consultance en recherche et innovation, qui travaillait avec des gens qui avaient besoin que l'on réfléchisse sur leur secteur. On tournait avec environ 1,2 million d'euros, dont 300.000 ou 400.000 euros en recettes propres".*

Au fil des ans, les activités de l'Agence Alter se poursuivent sur le mode de la diversification de ses activités – posant, par la même occasion, un débat sur ce que les journalistes peuvent ou non accepter comme mission, sur le plan éthique –, tandis qu'*Alter Échos* est publié à un rythme bimensuel. Sa maquette évoluera une première fois en 2001, à l'occasion de son numéro 107, tout en restant dans une logique de feuillets photocopiés, dont la couleur des pages change à chaque numéro. Elle adoptera une nouvelle physionomie en 2007 puis en 2012 et en 2018 (Figure 3.2), année au cours de laquelle le magazine adopte un rythme mensuel et démarre une diffusion en librairies. Vingt ans après son lancement, le lectorat d'*Alter Échos* n'a pas beaucoup évolué : il compte aujourd'hui entre 600 et 700 abonnés. À l'occasion du vingtième anniversaire de l'agence Alter, Aude Garely, qui avait intégré l'équipe en 2005, avant de la diriger de 2012 à 2015, retracera le fil d'une histoire marquée par le lancement de plusieurs projets qui ne perdront pas dans le temps. Elle épinglera celui des premiers numéros spéciaux réguliers d'*Alter Échos*, en 2012-2013, *"sur des thématiques à la frontière du social, qui font (de la) revue de plus en plus ouverte à d'autres secteurs que les secteurs sociaux traditionnels (par exemple mobilité et*

¹¹ Communication personnelle, Bruxelles, le 19/09/2017.

social, alimentation durable et social, énergie et social, logement, ...)". Elle soulignera la particularité d'Alter, "une forme hybride d'indépendance grâce à ce qu'on pourrait appeler la multi-dépendance ou la multi-activité (...) au niveau financier (...) C'est peut-être ça la seule chose qui n'a pas varié en vingt ans et qui fait la force de l'Agence, et aussi sa fragilité. (...) Ça veut dire une exigence de dingue pour financer dix emplois à ce jour (..) avec l'objectif toujours primordial de ne perdre ni son âme ni sa marge de manœuvre, et donc un délicat équilibre à préserver (...) Une particularité aussi, c'est le fait d'avoir dans une si petite structure des métiers complémentaires (journalisme, recherche) qui amène des façons d'intervenir que personne d'autre n'a, et qui comportent des valeurs ajoutées, et des tensions lorsqu'ils coexistent dans un même projet : sur les enjeux de déontologie, de mode d'intervention, de liberté, d'indépendance"¹².



FIGURE 3.2 – Évolution de la maquette d'Alter Échos en 1997, 2001, 2007, 2012 et 2018

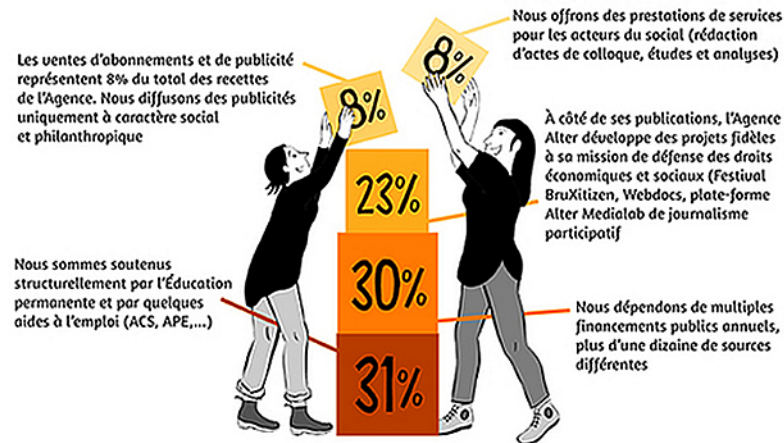
De l'équipe des débuts, qui a longtemps œuvré bénévolement au projet, il ne reste plus personne. En 2018, la rédaction se composait de cinq journalistes salariés, en ce compris la coordinatrice du magazine (anciennement coordinatrice de la rédaction, qui quittera la rédaction en février 2019), tandis que l'Agence Alter comptait quatre salariés de plus (un directeur financier, un coordinateur général, deux responsables de projets). Après avoir traversé plusieurs crises liées au management de l'association, l'équipe d'Alter a souhaité adopter un mode de gestion collaboratif. Sur le plan financier, sa dépendance aux fonds public a rendu le projet fragile. S'il ne s'agit pas de la première crise qu'elle traverse, celle qu'elle connaît actuellement apparaît comme l'une des plus sérieuses. En 2018, le budget annuel de l'association s'élevait à 850.000 euros. 76% de celui-ci était consacré aux salaires (Figure 3.3)¹³.

Le magazine *Alter Échos* dispose d'un site web depuis 2004. En 2017, il a fait l'objet d'une refonte complète, avec pour objectif de mettre davantage en valeurs les contenus, qui consistent essentiellement en des reprises d'articles publiés dans le magazine. Environ deux-tiers des articles publiés sont accessibles via un abonnement payant. Les statistiques démontrent que c'est sur ce terrain que le magazine rencontre le plus large lectorat. "La consultation du site est stable par rapport à l'année 2016 et en augmentation de 15% par rapport à 2015. Le nombre total de sessions enregistrées en 2017 est de 120.206, soit une moyenne mensuelle de 10.017 sessions par mois.

¹² Aude Garely, document transmis via courriel, le 18/09/2017.

¹³ Source : Kiss Kiss Bang Bang, "Campagne de soutien à la revue Alter Échos", consulté le 03/01/2018, URL : <https://www.kisskissbankbank.com/fr/projects/campagne-de-soutien-a-la-revue-alter-echos/T1\textendash2>

En termes de visites uniques, cela représente 92.118 internautes, soit une moyenne mensuelle de 7.676 visiteurs uniques par mois", indique la coordinatrice de la rédaction¹⁴. Par ailleurs, chaque publication est annoncée via une lettre d'information électronique. En décembre 2018, celle-ci comptait 4.771 abonnés. En ce qui concerne les réseaux sociaux, *Alter Échos* comptait, à la même époque, 5.872 abonnés à sa page Facebook et 2.272 abonnés à son compte Twitter.



À QUOI SERT L'ARGENT DES LECTEURS D'ALTER ÉCHOS ?

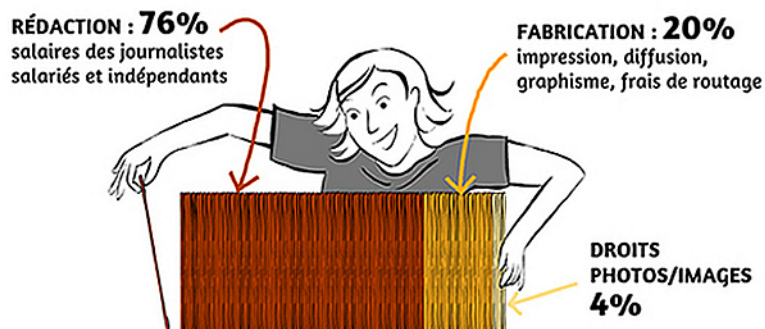


FIGURE 3.3 – Modèle économique du magazine *Alter Échos* en 2018

Alter Échos fut parfois le lieu de prises de positions tranchées, mais les idéaux d'hier n'ont pas forcément disparu. Les journalistes ne se considèrent pas comme militants, mais ils revendiquent un engagement pour davantage de justice sociale. Lors d'un débat sur la presse alternative organisé au Théâtre National à Bruxelles, le 28 septembre 2017 dans le cadre du Festival des Libertés, le journaliste Julien Winkel expliquera que si les moyens et la force de frappe du média sont différents, il ne considère pas qu'il s'agisse là d'une presse alternative. "Je ne me sens pas en opposition avec la presse traditionnelle. On fait le même boulot et on peut traiter les mêmes sujets. Les médias traditionnels peuvent avoir un traitement de l'information que l'on pourrait aussi trouver chez nous. Je ne me sens mal à l'aise avec cette division entre ce qui serait de la presse alternative et de la presse traditionnelle (...) On peut être engagé sans faire d'opinion. On peut être engagé dans ses choix mais on n'est pas des militants dans le traitement des sujets. Dans un sens, on fait de la presse d'opinion mais différemment qu'en affichant un étendard".

¹⁴ Communication personnelle, via courriel, le 06/12/2018.

Il ajoutera que le modèle économique du magazine n'entrave pas la liberté d'informer de la rédaction : "*Être financé par les pouvoirs publics ne nous a jamais posé problème, même si l'on sent parfois des velléités de recadrage*".

3.1.1.2 Les membres de la rédaction

Dans le contexte de l'expérience "Bxl'air bot", deux agents sociaux ont été particulièrement mobilisés dans la construction sociale de l'artefact : Sandrine Warsztacki, en sa qualité de coordinatrice et porte-parole de la rédaction ; et Céline Gautier, en sa qualité de journaliste indépendante spécialisée dans les questions relatives à la qualité de l'air. Les quatre autres membres de la rédaction ont participé à la réalisation d'un dossier thématique sur le sujet de la pollution atmosphérique, publié en avril 2018, où les données de l'application automatisée ont été traitées. Ces six journalistes ont participé aux trois réunions de rédaction organisée dans le cadre de la préparation de ce dossier.

Sandrine Warsztacki, coordinatrice de la rédaction.

Entrée en fonction en 2010, Sandrine Warsztacki (37 ans au moment de l'expérience) fut l'une des deux principales interlocutrices dans le cadre du projet "Bxl'air bot". Au sein du réseau, elle est la porte-parole de la rédaction. Formée en journalisme et en anthropologie, elle a d'abord travaillé comme chargée de communication à la Mission locale de Schaerbeek. Elle s'est ensuite réorientée vers le journalisme, sous le statut d'indépendante, pour différentes publications – parmi lesquelles *Métro* (quotidien gratuit), *Victoire* (supplément hebdomadaire du quotidien *Le Soir*) et *Imagine demain le monde* (magazine d'écologie et de société). Elle est également cofondatrice du magazine *Médor* et membre du Conseil de déontologie journalistique (représentante des journalistes).

Ses usages professionnels des technologies numériques portent essentiellement sur l'agrégation de contenus ou via les réseaux sociaux (Twitter, en particulier). "*Dans de rares cas, je m'en sers pour des appels à témoins sur un sujet précis, mais je ne trouve pas que cela donne lieu à des résultats extraordinaires*", indique-t-elle¹⁵. Elle s'intéresse également à une approche par données dans le journalisme qui serait davantage abordée en termes de source d'informations. Toutefois, elle déplore le manque de temps à pouvoir consacrer à de la formation continuée. "*Je n'ai pas spécialement les compétences pour le faire mais je commence à avoir le réflexe. Auparavant, je n'aurais jamais pensé qu'un tableau Excel pouvait constituer une source d'informations*". Concernant l'édition d'informations en ligne, elle utilise régulièrement le système de gestion de contenus WordPress, reconnaissant que les outils actuels "*sont assez intuitifs à utiliser mais, encore une fois, cela demande du temps. Chez Alter, on se bat pour avoir un modèle économique qui nous permette de faire du journalisme de fond, qui nous prend déjà pas mal de temps. Pouvoir chipoter, je vois ça comme un luxe : le temps, c'est quelque chose qui me manque*". Les valeurs défendues par *Alter Échos* sont, pour elle, celles de l'équité sociale, dans le sens d'une "*défense de l'accès aux droits économiques et sociaux*". Elle estime que le mensuel *Alter Échos* occupe une place particulière dans le paysage médiatique francophone belge, dans la mesure où il s'agit d'un espace dédié au journalisme d'enquête et aux reportages longs

¹⁵ Communication personnelle, Bruxelles, le 29/08/2018.

formats. Dans le même temps, elle reconnaît *"un manque d'impact. On est une petite structure et on manque de moyens. Tout ce qui concerne la communication va toujours passer au second plan, faute de temps"*. En février 2019, elle quittait la rédaction pour devenir coordinatrice de la rédaction du magazine *En Marche*, publié par les Mutualités Chrésiennes.

Cédric Vallet, journaliste (jeunesse, aide à la jeunesse et migrations).

Entré à la rédaction en 2010, Cédric Vallet (39 ans au moment de l'expérience) dispose d'une formation initiale en sciences politiques et en droits de l'Homme. Il a débuté son parcours professionnel dans une ONG s'occupant de réfugiés, avant de s'orienter vers le journalisme. Ses usages des technologies de l'information et de la communication se limitent à des activités de traitement de texte, de recherche sur internet, de veille passive sur les réseaux sociaux (Facebook et Twitter) et d'édition d'articles via le système de gestion de contenus WordPress. *"Je n'ai pas d'intérêt particulier pour ces choses, sauf lorsque cela peut me servir pour mon travail"*¹⁶. Il a toutefois suivi des formations à l'utilisation d'outils du web : *"Ce sont deux ou trois formations que je n'ai jamais mises en application, faute de temps"*. S'il est régulièrement amené à travailler avec des données, cela n'est pas systématique. Il affirme, à ce propos, que *"les chiffres ne me passionnent pas"* et font *"un peu peur"*. Il définit les valeurs du média comme étant celles d'une presse indépendante qui défend, en toile de fond, des droits économiques et sociaux.

Manon Legrand, journaliste (logement et culture).

Diplômée en histoire contemporaine et en journalisme, Manon Legrand (29 ans au moment de l'expérience) a rejoint la rédaction d'*Alter Échos* à temps partiel, en mars 2015, pour travailler – entre autres – sur le site web (mises à jour et rédaction de contenus spécifiques). Si elle dit ne pas éprouver un *"intérêt fort"* pour les technologies du web, la cadette de la rédaction estime toutefois que *"on ne se rend pas compte à quel point, dix ans de différence, cela joue. WordPress et Facebook n'ont pas de secret pour moi"*¹⁷. Elle utilise régulièrement des outils en ligne ainsi que des outils de traitement de textes, d'édition d'images et de montage vidéo. Si elle a suivi des formations spécifiquement dédiées aux outils du web, elle met peu en pratique ces acquis. *"On ne s'y met pas car cela demande du temps. J'ai déjà pensé à réaliser des entretiens en vidéo mais je ne vois pas ce que cela apporterait de plus. Mais il n'y a rien à faire, je ne parviens pas à me mettre à Twitter"*. Par contre, elle ne s'est jamais essayée à une approche par données dans le cadre de son métier : *"Je n'aime pas trop les chiffres. Ça m'effraie. Je n'y ai jamais été confrontée et je n'ai jamais cherché à le faire"*. Les valeurs portées par *Alter Échos* sont, selon elle, celles de la justice sociale, de la citoyenneté, du respect des droits, de la dignité des personnes, de l'indépendance et de la transparence. Depuis le départ de Sandrine Warsztacki, Manon Legrand assure la coordination du magazine.

Marinette Mormont, coordinatrice du supplément "Focales", journaliste (social et santé).

Historienne de formation, elle a travaillé comme chercheuse puis comme journaliste avant de rejoindre la rédaction en 2012. Ses usages des technologies sont, selon elle, *"old school"* : trai-

¹⁶ Communication personnelle, Bruxelles, le 26/09/2017.

¹⁷ Communication personnelle, Bruxelles, le 26/09/2017.

tement de textes, édition d'articles via WordPress, tableur Excel, logiciel de gestion de notes Evernote. *"J'insère des photos et des vidéos dans les pages web du site mais je n'en manipule pas les programmes d'édition. J'utilise également l'outil ThingLink¹⁸. D'une manière générale, j'utilise le matériel à ma disposition"*¹⁹. Elle a suivi des formations dédiées aux outils du web, mais elle dit ne pas avoir eu l'occasion de les utiliser car *"cela demande du temps pour s'y exercer et on laisse toujours ça de côté"*. En matière de réseaux sociaux, elle se définit comme une *"dino-saure"*, dans le sens où elle n'en utilise aucun, estimant *"malsain cette espèce de mélange entre privé et professionnel"*. Elle considère également qu'elle manque de temps pour ces outils, qui lui paraissent *"énergivores"*. Marinette dit parfois travailler avec des données mais *"pas de manière élaborée. J'ai déjà réalisé deux ou trois tableaux avec un graphiste pour illustrer des chiffres dans un article, mais ce n'est pas quelque chose de courant et cela va dépendre des sujets. La pression du temps empêche de développer tout cela. Personne, ici, n'est très spécialisé là-dedans. On reste dans notre zone de confort qui est l'écrit car on est y tous formatés. Si j'avais plus de temps, cela me plairait de développer tout cela"*. Les valeurs d'Alter Échos sont, pour elle, celles de l'indépendance et de *"la promotion des droits des personnes en défaut de droits"*.

Julien Winkel, journaliste (emploi et formation).

Employé à l'Agence Alter depuis 2008, il a été attaché de presse en politique puis journaliste indépendant (chroniques culturelles et musicales, piges pour Amnesty International), et est titulaire d'un master en journalisme. Son usage d'outils technologiques se limite à celui du logiciel Word. Il n'est pas présent sur les réseaux sociaux, et se considère comme un journaliste *"à l'ancienne"*²⁰, c'est-à-dire que *"j'enrichis mon carnet d'adresses, parle aux gens et collecte facilement des données. Je me rends aussi sur le terrain, et essaie d'être attentif à tout ce qui se passe pour trouver des sujets"*. S'il a suivi de courtes formations dédiées aux outils du web, il ne les a pas mises en pratique car *"on n'en a pas besoin pour le boulot et le serveur est naze. Je n'ai pas la culture d'utiliser des outils pour des enrichissements. Je ne suis pas non plus un 'vieux schnoll' du support numérique. On fait des longs formats chez Alter, ils sont bien foutus. Par rapport à d'autres journalistes, je me considère comme un incapable technologique. Je suis un dinosaure, alors que je n'ai même pas quarante balais"*. Spécialisé dans les matières relatives au chômage et à l'emploi, il a l'habitude de travailler avec des données chiffrées. Toutefois, il ne considère pas cela comme relevant du domaine du datajournalisme. Alter Échos n'est pas, pour lui, un média alternatif : *"On fait juste du journalisme. On prend le temps, on essaie d'être le plus didactique possible. On est dans du 'slow journalism'. Je n'ai pas l'impression que nous sommes très originaux mais notre positionnement, relatif aux matières que l'on traite, est particulier et innovant."* Pour lui, l'originalité de l'Agence Alter réside dans sa structure organisationnelle, *"ne fût-ce que par les fonds dont on dispose : on est le seul organe de presse qui est financé par des subsides et sans aide à la presse"*. Les valeurs portées par le magazine sont, selon lui, celles de la défense des droits sociaux *"dans toute une palette de sensibilités différentes. Mais nous ne sommes pas des militants. On essaie de rester objectifs, honnêtes"*.

¹⁸ Ce service en ligne permet de réaliser des images interactives pouvant être intégrées dans les pages d'un site web.

¹⁹ Communication personnelle, Bruxelles, le 26/09/2017.

²⁰ Communication personnelle, Bruxelles, le 26/09/2017

Céline Gautier, journaliste indépendante.

Céline Gautier a débuté sa carrière en 1999 comme journaliste pour différents médias belges, mais c'est à *Elle Belgique* qu'elle a fait l'essentiel de sa carrière, de 2010 à 2015. C'est à cette époque qu'elle a commencé à travailler comme indépendante. Elle fait partie des membres fondateurs du périodique d'investigation *Médor*, dont elle est présidente du conseil d'administration, et elle collabore avec *Alter Échos* depuis 2014. Elle est membre du Conseil de déontologie journalistique depuis 2015, et elle enseigne ponctuellement le journalisme dans une haute école, ainsi que dans le cadre de formations AJPro (le service de formation continuée de l'Association des journalistes professionnels). Elle est titulaire d'une licence en langues et littératures romanes, et d'une agrégation de l'enseignement secondaire supérieur. Elle n'estime pas présenter un profil web ou data. "Je comprends l'intérêt de l'open data mais je n'ai pas une pratique de ces questions et de ces outils. Je fais partie des gens qui ne sont pas encore conscients de toutes les possibilités de ce domaine"²¹. Aux débuts de l'expérience "Bxl'air bot", elle venait de recevoir une bourse du Fonds pour le journalisme, pour une enquête portant sur les concentrations en dioxyde d'azote dans le métro bruxellois, destinée au trimestriel *Médor*.

Les journalistes de la rédaction d'*Alter Échos* ne témoignent pas d'une appétence particulière que ce soit pour le traitement de données – bien que certains d'entre eux aient affirmé en percevoir l'utilité – ou pour les technologies numériques. Leurs pratiques professionnelles s'inscrivent moins dans ces champs que dans celui dans une forme journalisme idéalisé où l'écrit domine et où le travail de terrain concentre l'essentiel de leurs activités. Trois membres de la rédaction se considèrent davantage comme des journalistes "à l'ancienne", alors que la moyenne d'âge de la rédaction est de moins de quarante ans. Dans l'ensemble, l'usage des technologies numériques est perçu comme un facteur mobilisateur de temps : que ce soit en raison de leur caractère chronophage, du manque de temps à consacrer à la formation continuée ou du manque de temps pour mettre en pratique les éléments vus lors de formations. L'imaginaire de la rédaction est également forgé par les valeurs portées, à leurs yeux, par le magazine.

3.1.2 Conception de l'artefact technologique



Le processus de conception de "Bxl'air bot" correspond à la phase de flexibilité interprétative du modèle SCOT. Il s'est déroulé en trois temps : la compréhension des flux de données disponibles, la définition de l'architecture du système d'information et le design de son interface, en vue de proposer un prototype à la rédaction d'*Alter Échos*. Tout au long de ce processus, le groupe social de l'administration publique va jouer un rôle important, dans la mesure où il va permettre d'identifier avec précision le cycle de vie des données. Il jouera également un rôle

²¹ Communication personnelle, Bruxelles, le 23/05/2017.

de validation scientifique de l'application, bien que celle-ci posera des choix journalistiques susceptibles d'entrer en contradiction avec la manière dont l'autorité envisage la communication des données. En premier lieu, considérant qu'il s'agit d'un problème de santé publique, les valeurs de référence tiennent compte des recommandations de l'Organisation Mondiale de la Santé (OMS), plus restrictives que les normes européennes. En second lieu, toutes les opérations statistiques réalisées à partir des données collectées incluent les mesures de la station de mesure Arts-Loi, exclue du *reporting* annuel obligatoire auprès de la Commission européenne car elle ne correspond pas aux prescriptions de l'UE pour la surveillance de la qualité de l'air en matière de santé publique²², en raison de sa trop grande proximité avec un carrefour important. D'un point de vue scientifique, les mesures effectuées à proximité de cet axe ne sont pas considérées comme représentatives de l'exposition de la population à la pollution de l'air. Toutefois, sur le plan journalistique, l'inclusion de ces données apparaissent comme un choix pertinent. Lors d'un entretien exploratoire avec la journaliste indépendante spécialisée dans les questions relatives à la qualité de l'air, qui fut réalisé le 23 mars 2017 lors de la phase de prototypage de l'application, celle-ci a confirmé ces choix tout en soulignant qu'une autre station de mesure (sur les dix en activité au moment de l'expérience) lui paraissait problématique : celle située sur le site de l'Observatoire d'Uccle où, à l'inverse, il y a peu de circulation automobile. Pour elle, l'inclusion ou l'exclusion d'une station de mesure constitue un moyen de fournir une vue plus représentative de la qualité de l'air bruxellois.

3.1.2.1 Accessibilité des données



Les mesures des taux de polluants atmosphériques sont organisées, en région bruxelloise, par Bruxelles Environnement (Institut bruxellois pour la gestion de l'environnement, IBGE). Cet organisme, qui dépend de l'autorité régionale bruxelloise, gère le réseau des stations de mesure. Il assure également la prise en charge des communications de ces mesures au public, de même que la Cellule Interrégionale de l'Environnement (CELINE). Ces deux organismes communiquent aussi à propos de l'indice de la qualité de l'air, qui consiste en une formule scientifique tenant compte des moins bonnes mesures des taux de polluants²³. Toutefois, Bruxelles Environnement souligne que cet indice constitue "*un outil de communication qui n'est pas approprié à une interprétation scientifique du phénomène de la pollution atmosphérique*"²⁴. Bruxelles Environnement privilégie, sur son site web, un mode de communication qui apparaît comme rassurant (Figure 3.4), par l'utilisation de couleurs chaudes (les nuances entre le jaune et le rouge sont dominantes), d'une iconographie faisant référence à l'univers des cartoons (utilisation de *smileys*) et d'un langage simplifié. La communication de CELINE (Figure

²² Directive 408/50/CE.

²³ Le mode de calcul de cet indice (BelAtmo) a été modifié au cours de l'expérience (BelAQI), nécessitant une mise à jour du code de l'application "Bxl'air bot".

²⁴ Source : <http://qualitedelair.brussels>

3.5) s'inscrit, quant à elle, dans une logique scientifique, où les mesures des taux de polluants sont accessibles via cinq pages web distinctes. Si elles sont également présentées sous la forme de tableaux et de cartes interactives, les représentations cartographiques sont davantage mises en avant. Elles consistent dans des cartes de Belgique, dont la colorimétrie correspond à un niveau donné de pollution par type de polluant. La région bruxelloise ne fait donc pas l'objet d'un traitement isolé.



FIGURE 3.4 – Page d'accueil du site web de Bruxelles Environnement

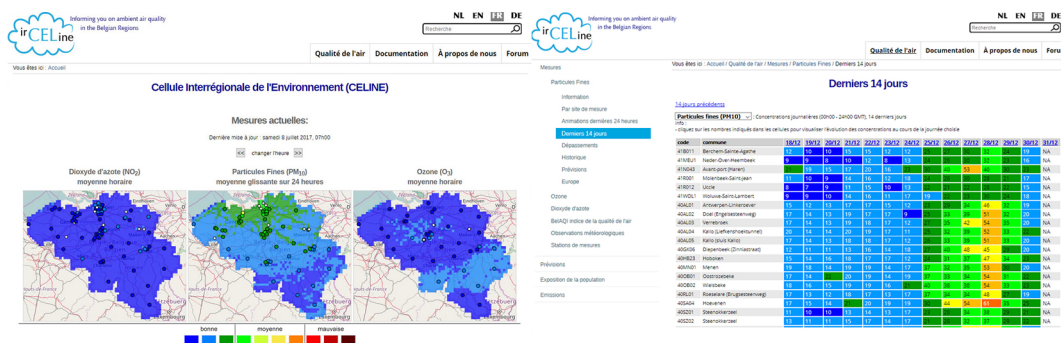


FIGURE 3.5 – Diffusion des mesures des taux de polluants sur le site de CELINE

Lors de l'étude préliminaire relative à l'accessibilité des données, en vue de leur récupération, il est apparu qu'aucun de ces deux organismes ne propose des données en téléchargement ou sous la forme d'un flux. Leur mode de diffusion pose potentiellement des problèmes de qualité, en raison du format mouvant d'une page web. Ces données consistent en des moyennes glissantes, devenant fixes par tranche de vingt-quatre heures. Seules les moyennes fixes sont diffusées par CELINE, dans des tableaux retraçant l'historique des mesures réalisées, lesquels peuvent être consultés dans des pages web distinctes du site.

La piste des portails *open data*, qui proposent en accès libre des données publiques, a été explorée mais elle n'a pas donné lieu à des résultats probants. Qu'il s'agisse du portail fédéral, de celui de la Ville de Bruxelles ou de celui de la Commission européenne, les problèmes ren-

contrés sont ceux d'une absence de valeur et de documentation, d'un étiquetage ambigu et d'un manque d'actualité. CELINE mentionne bien l'existence d'un web service permettant de récupérer un flux de données en temps réel, mais sans préciser comment y accéder. Quatre liens vers ce web service²⁵ ont été communiqués en cours d'expérience par un représentant de CELINE mais aucun d'entre eux n'a pu être exploitable dans le cadre de ce projet.

3.1.2.2 Évaluation de la qualité des données

La nécessité de disposer de données relatives tant aux moyennes glissantes, pour la diffusion d'informations en temps réel, qu'aux moyennes fixes, pour un enregistrement de l'historique des données, a abouti au parti pris d'agréger les données via les pages web du site de CELINE. Cinq pages web ont été identifiées comme pertinentes pour la récupération des données. Elles correspondent chacune à un type de polluant particulier. De manière à évaluer les problèmes pouvant se poser dans le cadre de la collecte de ces données, en vue de les anticiper en amont, ces pages web ont été soumises à l'épreuve du modèle conceptuel d'évaluation développé dans le deuxième chapitre de cette thèse²⁶ (Tableau 3.3. et Tableau 3.4).

Axe	Type d'évaluation	Évaluation (Oui/- Non/Non Pertinent)
Axe documentaire	Mention de la licence d'utilisation	O
	Identifiant unique	N
	Présence de métadonnées	N
	Conformité aux métadonnées	N
Axe d'encodage	Pas de problème d'encodage	N
	Pas de surcharge du code HTML	N
	Pas de données dupliquées	O
Axe normatif	Application des standards (adressage, géolocalisation, unités de mesure, URL normalisés, métadonnées, ...)	NP
Axe sémiotique	Pas d'incohérences orthographiques	O
	Étiquetage des colonnes explicite et non ambigu	O
	Pas de valeurs manquantes	N

Tableau 3.3 – Évaluation du challenge technique

²⁵ Voir annexe 10, "Données en temps réel diffusées par CELINE (web service)", p.335

²⁶ Voir p.114

Dimension	Type d'évaluation	Évaluation (Oui/-Non/Non Pertinent)
Intrinsèque	Exactitude (correction syntaxique)	O
	Précision des valeurs (pas d'anomalies)	N
	Actualité (date et fréquence de mise à jour)	O/N
Contextuelle	Provenance (source authentique)	O
	Nombre de données approprié (pas de lignes vides)	O
	Complétude (pas de valeurs manquantes)	N
	Pertinence	O

Tableau 3.4 – Évaluation du challenge journalistique

Modèle général d'évaluation de la qualité des données

- **Source.** Le diffuseur des données n'en est pas le producteur original ou la source authentique, mais son statut permet de l'y assimiler. La source est donc considérée comme fiable.
- **Accès.** Les données sont accessibles en *open data*. Cela signifie que ces données peuvent être réutilisées, bien que le format proposé puisse porter à discussion en matière de qualité des données fournies.
- **Documentation.** Les données ne sont pas documentées dans un fichier de métadonnées mais elles sont expliquées et commentées, de manière aussi disparate qu'abondante, dans de nombreuses pages du site web. Des informations de contexte sont fournies, telles que celles relatives aux normes européennes. La structure des tableaux de données est compréhensible, s'organisant en mesures par stations. L'étiquetage des colonnes est explicite et sans ambiguïté (il s'agit de dates).
- **Automatisation.** Si le format d'une page web est libre et exploitable, la récupération des données suppose une opération de "*parsing*", qui permet de cibler des cellules spécifiques du tableau pour en récupérer les valeurs. Toutes les cellules présentent une surcharge de code HTML, liée à leur mise en forme. Dans certaines d'entre elles, des valeurs sont manquantes ou non connues. Elles sont signalées par la mention "NA". Chaque type de polluant est repris dans un tableau qui lui est dédié et mis à jour tout au long de la journée, sans pour autant que la fréquence des mises à jour soit mentionnée.
- **Pertinence journalistique.** Les données permettent de suivre les dépassements relatifs aux normes européennes et aux recommandations de l'OMS, et de suivre l'évolution des mesures station par station. Elles permettent de donner un sens à l'analyse, en établis-

sant les moments les plus sensibles de la journée. Un suivi de l'évolution des données dans le temps consiste en un point d'entrée pertinent pour dégager des sujets journalistiques.

Bien qu'il ait été utile dans la conception et le design de l'architecture de l'application, ce modèle d'évaluation a, toutefois, montré sa principale limite : celle de l'impossibilité de définir la fréquence exacte de la mise à jour des données. La compréhension du cycle de vie des données s'est donc posée comme une nécessité, celle-ci ne pouvant se faire autrement qu'en sollicitant des experts du domaine d'application, en l'occurrence des représentants de CELINE. Par ailleurs, au terme de trois semaines d'observation des données diffusées en ligne, il est également apparu que les valeurs étaient susceptibles d'être modifiées dans le temps, les pages web présentant un indéniable caractère mouvant. Dès lors, il était aussi important de connaître le processus de validation des données. Les informations fournies par les représentants de CELINE ont permis de mieux comprendre comment s'organisent la création, la diffusion et la validation des données (Figure 3.6).

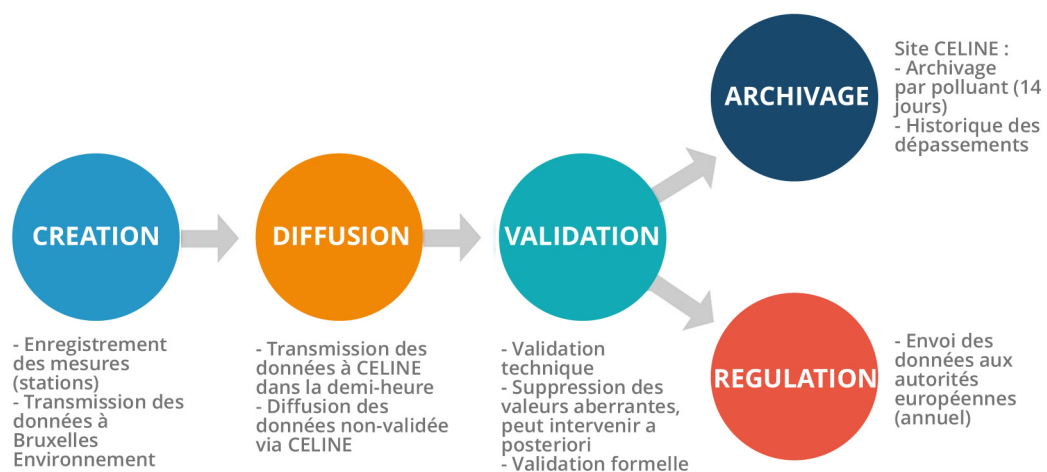


FIGURE 3.6 – Cycle de vie des données relatives aux mesures des taux de polluants atmosphériques en région bruxelloise

À la source des données, se trouvent douze stations de mesure des taux de polluants atmosphériques – dont deux étaient inactives au moment de l'expérience. Leur répartition géographique a pour objectif de représenter la diversité des zones urbaines bruxelloises. Les mesures sont réalisées tout au long de la journée, et elles sont collectées par Bruxelles Environnement qui les transmet à CELINE environ une demi-heure après leur réception²⁷. CELINE les publie, dès leur réception, sur son site internet. Des interruptions peuvent se produire lors du transfert des données, par exemple en raison d'une panne d'électricité ou de serveurs informatiques. Par ailleurs, toutes les stations ne mesurent pas nécessairement les mêmes types de polluants. Toutes les valeurs diffusées consistent en des données non-validées, et elles peuvent faire l'objet d'ajustements dans le temps. Aussi, une première validation a pour objet de vérifier que la

²⁷ Représentants de CELINE, communication personnelle, Bruxelles, le 12/04/2017.

station de mesure se trouve dans un état de fonctionnement "correct". Une deuxième validation est relative à un contrôle effectué en interne, dont l'objet est de détecter les anomalies et, le cas échéant, de supprimer les valeurs jugées anormales ou aberrantes (par exemple, un taux d'ozone élevé en hiver, alors que la formation de ce polluant nécessite un ensoleillement abondant et des températures élevées). Ces deux modes de validation expliquent que des valeurs puissent être manquantes dans les tableaux présentant les données. Une ultime validation a lieu *a posteriori* : il s'agit des données qui seront transmises à l'autorité européenne, mais elles ne sont pas communiquées au public.

Le cycle de vie des données relatives aux mesures de la qualité de l'air en région bruxelloise peut donc être décomposé en cinq temps : création, diffusion des données non-validées, validation des données, transmission des données aux autorités européennes et archivage. Sa compréhension a été déterminante quant aux choix posés lors de la conception de l'application, en particulier sur le plan technique car l'enregistrement des moyennes fixes, par station de mesure et par type de polluant, devait rencontrer des objectifs journalistiques d'exactitude et de précision.

3.1.2.3 Architecture du système d'information



La définition de l'architecture du système d'information a consisté en une traduction des premiers choix éditoriaux posés par la chercheuse, lesquels ont ensuite été confrontés à la rédaction d'*Alter Échos*, en vue de les adapter à leurs besoins et exigences éditoriales.

L'application automatisée s'est appuyée sur deux modes de collecte des données : en temps réel, s'agissant de générer un bulletin d'information mis à jour à chaque nouveau chargement de la page par l'utilisateur ; et en différé, s'agissant de l'enregistrement des moyennes fixes des taux de polluants, en vue de nourrir une analyse relative à leur évolution dans le temps. Quatre formats de pages ont été définis au cours de cette étape (Figure 3.7) :

- (1) une page "Bulletin du jour", page d'accueil de l'application web ;
- (2) une page relative aux informations de contexte, en vue d'expliquer et de documenter le projet ;
- (3) un tableau de bord présentant une analyse statistique évoluant dans le temps, au fur et à mesure des enregistrements ;
- (4) un page permettant d'accéder aux archives mensuelles des données enregistrées.

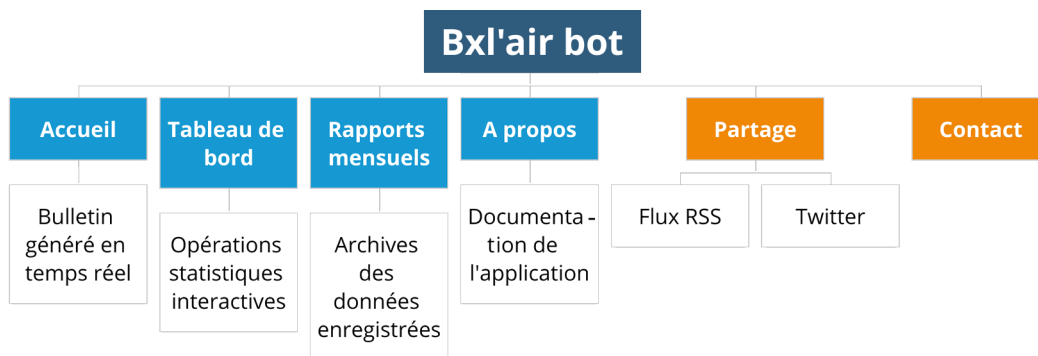


FIGURE 3.7 – Première version de l'arborescence de l'application "Bxl'air bot"

Dans la perspective d'une information servicielle, il a été prévu qu'un flux RSS soit également généré de manière automatique – il enregistrerait, dans un format structuré, la mise à jour des informations publiées en page d'accueil du site, sur une base horaire – en vue d'alimenter, de manière automatique, un compte Twitter dédié qui donne de la visibilité au projet²⁸.

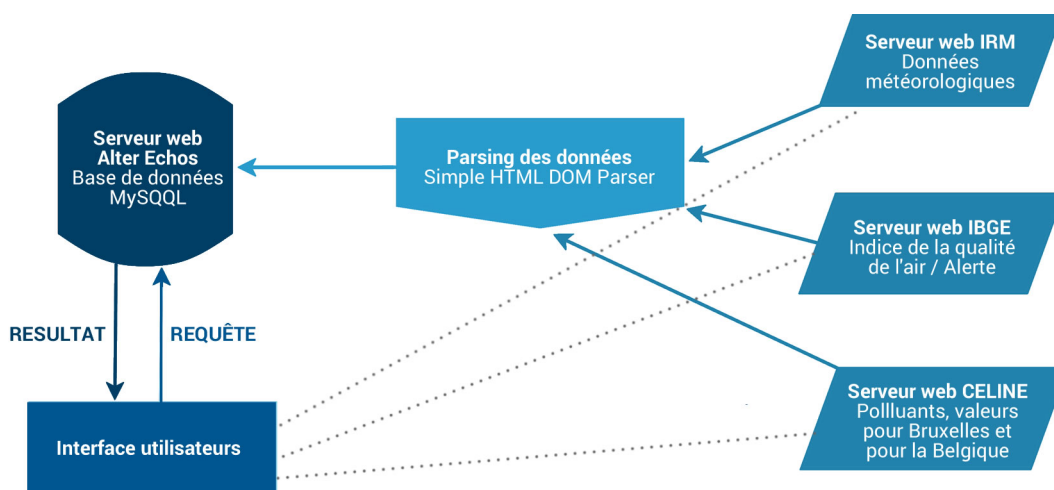


FIGURE 3.8 – Modélisation du flux de données de "Bxl'air bot"

Dans cette architecture, trois types de données ont été prises en compte. Elles ont donné lieu à une première modélisation des flux de données (Figure 3.8) :

- les données relatives aux mesures des différents types de polluants (particules fines, dioxyde d'azote et ozone) publiées sur le site de CELINE;
- les données relatives à l'indice de qualité de l'air et au niveau d'alerte, diffusées en temps réel sur le site de Bruxelles Environnement et sur celui de CELINE;
- les données relatives aux observations météorologiques – un facteur influençant la formation de certains polluants, comme l'ozone – publiées en temps réel sur le site internet de l'Institut Royal Météorologique (IRM). Toutefois, les données de l'IRM n'étaient

²⁸ URL : <https://twitter.com/bxlairbot>

pas encore officiellement libérées au moment de la conception de l'application. Deux autres alternatives avaient alors envisagées : les prévisions de l'institut météorologique néerlandais et celles de l'institut météorologique finlandais, accessibles en *open data*. Toutefois, les prévisions ne correspondent pas forcément aux observations. Dans un souci de précision, il a été décidé de s'appuyer sur les observations diffusées sur le site de l'IRM, compte tenu d'un cadre légal favorable à une libre réutilisation des données produites par les autorités publiques²⁹.

Les modes de narration ont été envisagés de manière complémentaire, l'objet étant de permettre une analyse des données qui soit accessible dans une forme textuelle (génération automatique de textes en langage naturel) et graphique (évolution des taux de polluants dans le temps et analyses statistiques). À ce stade de la conception, il était prévu que les textes présentent une synthèse des observations en cours (temps réel) et des observations en matière de dépassement, par type de polluant, des normes européennes et des recommandations de l'OMS. Sur le plan du processus de génération automatique de textes, le traitement des données s'est appuyé sur un système à base de règles ("si... alors... ou si... alors.... ou...") et les textes générés le sont à partir de gabarits de texte (ou *templates*) définis en amont.

Les graphiques avaient pour objet de proposer une représentation de l'évolution des données dans le temps. La visualisation de données a été ici envisagée comme un outil de communication de données abstraites par l'utilisation d'une interface visuelle interactive (Manovich 2011). Mais il s'agissait surtout de présenter des données quantitatives, de telle sorte à ce que l'on puisse y percevoir "*des tendances ou des anomalies, une constance ou une variation, que d'autres formes de présentation, textes et tableaux, ne permettent pas*" (Friendly 2008). Cette présentation s'appuyait sur l'usage de statistiques descriptives – nombre de jours de dépassements des normes et recommandations, valeurs maximales par type de polluant, moyennes observées par station de mesure –, dans un tableau de bord pouvant être consultés de manière dynamique.

Définition des standards

Les standards techniques ont été définis en tenant compte d'un travail général de conception de l'application et des bonnes pratiques dans ce domaine. Le parti pris a été de travailler à partir de *frameworks*³⁰, qui consistent en des espaces de travail modulaires où sont proposés des composants prêts à l'emploi. Ceux-ci étaient de deux types : graphiques (pour la réalisation d'une interface modulable en fonction de la taille de l'écran de l'utilisateur et pour la

²⁹ Loi du 4 mai 2016, publiée au Moniteur Belge le 3 juin 2016, qui consiste en une transposition de la directive 2013/37/UE du Parlement européen et du Conseil du 26 juin 2013 concernant la réutilisation des informations du secteur public, source : http://tiny.cc/loi_open_data.

³⁰ "*Un framework est un espace de travail modulaire qui consiste en une collection de fichiers (...) et de conventions permettant le développement rapide d'applications. Il fournit suffisamment de briques logicielles et impose suffisamment de rigueur pour pouvoir produire une application aboutie et facile à maintenir. Ces composants sont organisés pour être utilisés en interaction les uns avec les autres*", définition proposée par TechnoSciences.net, consulté le 10/07/2017, URL : <http://www.techno-science.net/?onglet=glossaire&definition=1471>

génération de visualisations de données interactives), et techniques (pour la récupération automatique des données – ou *parsing* – à intervalles réguliers). Sans entrer dans des détails trop techniques, précisons que le langage de développement sur lequel s'appuyait l'application web automatisée est le PHP³¹, car il permet d'interagir avec une base de données via le langage de requête MySQL³², dans des délais d'exécution rapides. La structure de la base de données reflétait le type de données agrégées en fonction de leur source. Elle comportait quatre tables : air (indice de la qualité de l'air et niveau d'alerte), polluants (mesures moyennes des différents polluants atmosphériques mesurés en région bruxelloise et en Belgique, pour établir des comparaisons), stations (mesures réalisées dans les dix stations bruxelloises en activité au moment de l'expérience), météo (observations météorologiques).

3.1.2.4 Prototypage de l'interface



La qualité du code informatique³³ est un élément important à prendre en compte pour garantir un fonctionnement optimal de l'interface. L'expérience du code procède d'une démarche de mise en visibilité des processus définis au moment de la phase de conception, et sa pratique induit qu'il soit à la fois un objet, dont la matérialité est opaque pour l'utilisateur final, et un acte, dans le sens d'un travail (McCosker & Milne 2014).

La couleur dominante de la charte graphique était le bleu, consistant en une variante de la couleur du logo d'*Alter Échos*. De nombreux travaux en psychologie des couleurs, que ce soit dans le domaine de la communication ou dans celui de l'art, ont mis en avant le rôle déterminant de la couleur dans la perception d'une interface. Le bleu est considéré comme généralement mieux perçu que des couleurs chaudes, et il notamment pour vertu de relaxer l'utilisateur (Bonnardel *et al.* 2011). Une icône rappelant un baromètre a été choisie pour appuyer le nom de l'application, qui consiste en un jeu de mot en dialecte bruxellois (un "Brusseleer" signifie "un Bruxellois")³⁴. L'identité graphique de l'interface a également été conditionnée par les contraintes d'accessibilité liées aux usages des internautes – on parle d'expérience utilisateur ou d'UX pour *user experience* (Unger & Chandler 2012). Cela s'est traduit dans une mise en page fluide et sans surcharge, un étiquetage clair et non ambigu des éléments de navigation, l'utilisation d'une iconographie qui fait sens pour l'utilisateur, et une interface conçue pour être déclinée de manière optimale sur tous les formats d'écran. Les règles d'accessibilité pour le design d'interfaces, qui relèvent du domaine des interactions hommes-machine³⁵, ont été

³¹ Acronyme de "Hypertext Preprocessor".

³² Les données sont stockées dans une base de données phpMyAdmin, couramment utilisées dans le contexte de sites web dynamiques.

³³ Le code informatique de l'application "Bxl'air bot" a évolué tout au long l'expérience pour des raisons d'optimisation, de réajustement du *parser* en fonction des données ciblées ou parce que de nouvelles fonctionnalités ont été ajoutées à l'interface.

³⁴ Voir annexe 11, "Première version de l'interface de l'application 'Bxl'air bot'", p.337

³⁵ *Human-Computer Interaction* (HCI).

prises en pratique (Figure 3.9). Celles-ci disposent que l'acceptabilité globale d'un système informatique consiste essentiellement à déterminer si celui-ci "est assez bon pour satisfaire tous les besoins des utilisateurs" (Nielsen 1999). Elle est une combinaison entre son acceptabilité sociale et son acceptabilité pratique (Nielsen 1994).

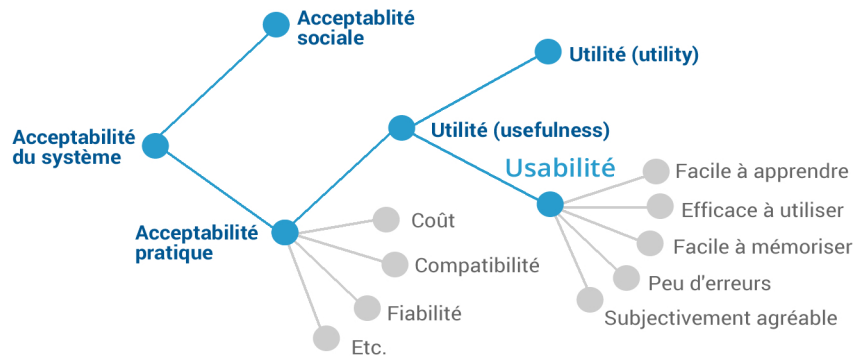


FIGURE 3.9 – Modèle des attributs d'acceptabilité d'un système (Nielsen 1999)

Les graphiques ont été travaillés de manière à permettre leur interactivité, que ce soit via la sélection de variables spécifiques ou via celle des mesures réalisées au cours d'un mois en particulier. Dès lors qu'il s'agissait de montrer des évolutions dans le temps, les graphiques en courbe se sont imposés comme le mode de visualisation le plus approprié pour représenter des données continues (Stray *et al.* 2013). Ceux-ci portaient sur l'évolution de l'indice de la qualité de l'air et sur celle des mesures du taux de chaque polluant. Ces visualisations peuvent être considérées comme des affichages symboliques qui révèlent les données (Tufte 2001). Ils sont régis par une série de bonnes pratiques, qui relèvent notamment du domaine de la sémiologie graphique (Bertin 2013). Les visualisations de données étaient titrées et légendées, de manière à pouvoir être comprises lorsqu'elles sont détachées de leur contexte. Les concepts ont été regroupés pour présenter des données comparables entre elles.

Les générations automatiques de textes se sont appuyées sur des gabarits qui consistent en des chaînes de caractères pré-écrites. Différentes règles ont été appliquées dans leur codage, en fonction de la valeur des données ou de leur variation. Ce mode de conception a été privilégié car elle facilite le contrôle du contenu ainsi que la gestion des erreurs (Leppänen *et al.* 2017). Les gabarits de texte ont été spécifiquement développés dans le cadre de cette recherche, et ne sont reproductibles que dans le seul contexte du domaine d'application de mesures des taux de polluants atmosphériques. Il ne s'agissait pas de produire des textes en grande quantité mais bien de générer deux modèles de textes dont la forme allait évoluer en fonction des valeurs des données qui les nourrissent³⁶ :

- **En page d'accueil** : il s'agissait d'un bulletin traitant des données recueillies en temps réel depuis les sites internet de Bruxelles Environnement (indice de la qualité de l'air, ni-

³⁶ Voir annexe 12, "Bxl'air bot : génération automatique de textes en langue naturelle" p.338

veau d'alerte), de CELINE (valeurs des mesures du taux de particules fines, d'ozone et de dioxyde d'azote), et l'IRM (température et type de temps³⁷). Dans la logique du temps réel, cette page était rafraîchie à chaque nouveau chargement de l'utilisateur, donnant ainsi lieu à une nouvelle version du texte généré. Le modèle était linguistique, c'est à dire basé sur un catalogue d'expressions de référence. De deux à quatre expressions de référence ont été rédigées pour chacun des dix syntagmes principaux que comportait le texte (selon le moment de la journée et les données disponibles et leurs valeurs, le texte généré compte de 10 à 15 syntagmes). La structure du texte a été inspirée par des récits journalistiques et des communications d'autorités traitant de la qualité de l'air. Pour assurer un minimum de variabilité au bulletin généré en temps réel, un catalogue de synonymes a été ajouté dans le code de l'application. Le modèle développé affichait aléatoirement les expressions de référence et les synonymes, en fonction de la valeur des données. 1.024 syntagmes différents étaient possibles, bien que le taux de variation de ces syntagmes n'était pas très élevé dans les faits³⁸.

- **Dans la page "Statistiques"** : il s'agissait d'un modèle de phrases à trous se modifiant dans le temps en fonction des moyennes ou totaux effectués à partir des données enregistrées dans la base de données. Ce texte avait pour fonction de proposer une synthèse des analyses réalisées depuis le début des enregistrements, le 1^{er} avril 2017.

Dans le souci du respect du principe d'identification formelle³⁹, la nature de l'auteur des textes générés de manière automatique a été explicitement mentionnée : "Ce texte a été généré de manière automatique à partir de données publiques extraites en temps réel", et "Ce texte a été généré de manière automatique à partir de données publiques enregistrées de manière quotidienne, depuis le 01/04/2017".

3.1.2.5 Maintenance du système

La gestion et la prévention des erreurs, dans le souci de satisfaire l'exigence journalistique de précision, a nécessité un monitoring humain organisé sur une base hebdomadaire. L'objet était de s'assurer que les valeurs enregistrées dans de la base de données correspondaient bien à celles diffusées par CELINE. Sur le plan opérationnel, cette activité de monitoring s'est notamment traduite dans une comparaison des valeurs agrégées avec celles diffusées par les sources de données, via une interface de contrôle développée en parallèle (Figure 3.10).

³⁷ URL : <http://www.meteo.be/meteo/view/fr/123386-Observations.html>

³⁸ Cette affirmation est fondée sur les résultats du calcul de la distance de Levenshtein (ou distance d'édition), qui mesure la similarité entre deux chaînes de caractères, à partir d'un échantillon composé d'un texte source et d'un texte cible. Ce même échantillon a également été soumis à un test de lisibilité, via le calcul du score Flesch-Kincaid, qui calcule le degré de lisibilité d'un texte. Plus celui-ci est élevé, plus le texte est lisible. En moyenne, les textes ont obtenu un score de 69, ce qui correspond à un niveau d'acceptabilité satisfaisant. D'autres tests de lisibilité ont également été appliqués, tous convergeaient vers un résultat plutôt positif. Les formules de calcul de ces indicateurs automatiques figurent à l'annexe 18, p.351, les tests ont été réalisés sur un site web personnel : <https://ohmybox.info/linguistics/fr>

³⁹ Voir les recommandations proposées dans le chapitre 2 consacré aux dimensions culturelles de l'automatisation de la production d'informations, p.144

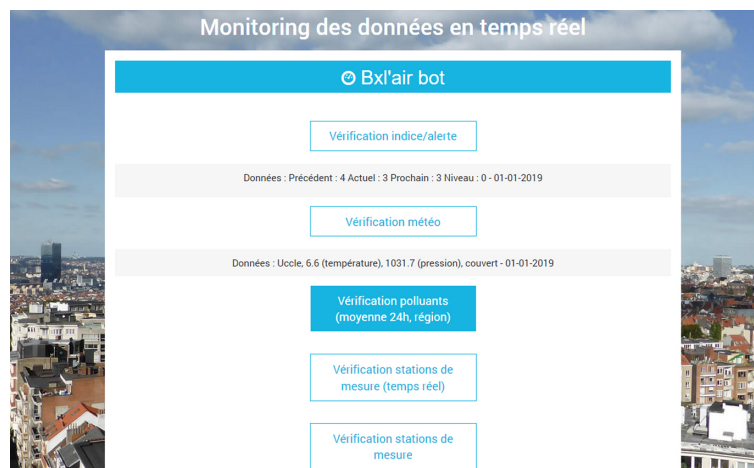


FIGURE 3.10 – Interface pour le monitoring des données diffusées en temps réel

Trois types d’anomalies ont été régulièrement constatés ⁴⁰ : des modifications *a posteriori* des valeurs relatives aux moyennes fixes des taux de polluants, des modifications *a posteriori* de valeurs initialement manquantes, et des modifications dans la structure d’une des pages web à la source des données. Elles ont également pu découler de problèmes techniques qui ont entraîné une absence de valeurs, ou la diffusion de valeurs négatives, comme cela fut constaté le 29 mai 2017 en raison d’un problème technique. Autre exemple : le 2 juillet 2017, plus aucune valeur n’était diffusée. Cette fois, il s’agissait d’une panne d’électricité qui a eu une conséquences directe sur l’interface de l’application : pas de données, pas d’enregistrements... et donc pas d’information. Bien que le traitement des données en temps réel ait été conçu pour s’adapter à l’éventualité de valeurs manquantes ou à des anomalies liées à des valeurs négatives, le dispositif ne permettait pas de résoudre le caractère potentiellement évolutif des données relatives aux moyennes fixes des taux de polluants. À au moins cinq reprises, au cours des douze mois de l’expérience, des interventions directes ont dû être effectuées dans la base de données, en raison de l’évolution de valeurs dans le temps. Ceci illustre l’une des difficultés majeures dans la mise en place d’un tel projet : celui du temps que nécessite un monitoring humain, malgré l’automatisation.

3.2 Construction sociale de l’artefact technologique



Les premiers besoins de la rédaction ont pu être identifiés à partir du moment où le prototype de l’application a été confronté à la coordinatrice de la rédaction et à la journaliste indépendante spécialisée dans les questions relatives à la qualité de l’air. Ce processus, qui participe à celui de la construction sociale de l’application, peut être compris comme une première

⁴⁰ Voir annexe 13, "Échantillon d’anomalies détectées dans le cadre du monitoring des données", p.340

forme d'usage, où vont se mettre en place des mécanismes d'ajustement réciproque entre l'objet et son environnement social (Akrich 2010).

"Par la définition des caractéristiques de son objet, le concepteur avance un certain nombre d'hypothèses sur les éléments qui composent le monde dans lequel l'objet est destiné à s'insérer. Il propose un 'script', un 'scénario' qui se veut prédétermination des mises en scène que les utilisateurs sont appelés à imaginer à partir du dispositif technique et des prescriptions (...) qui l'accompagnent. Mais tant qu'il ne se présente pas d'acteurs pour incarner les rôles prévus par le concepteur (ou en inventer d'autres), son projet reste à l'état de chimère : seule la confrontation réalise ou ir-réalise l'objet technique"(Akrich 2010).

Dans cette expérience, la confrontation avait pour objet de mettre en place un dispositif inclusif, en vue de stimuler des interactions à travers une participation active (Flanagin *et al.* 2010). Théoriquement, celles-ci vont dépendre du cadre sociotechnique des agents sociaux (Bijker *et al.* 1987), de la structure organisationnelle à laquelle ils appartiennent, et de leurs référents sociaux et culturels : c'est en cela que le processus technologique est flexible (Napoli 2014, Klein & Kleinman 2002). L'application "Bxl'air bot" est donc ici entendue comme résultant d'un assemblage sociotechnique, où le processus de confrontation permet de se représenter ou d'imaginer la manière dont l'artefact pourrait être utilisé sur un plan journalistique. En ce sens, la confrontation relève d'un processus mental de projection dans des usages finaux. En amont de cette confrontation, quatre types de choix avaient donc déjà été posés : (1) la définition du domaine d'application, (2) l'identification des sources et des types de données disponibles, (3) la définition des processus d'extraction des données, (4) les premières formes de traitement automatisé. Les scénarios d'usage définis lors de la conception de l'artefact étaient essentiellement liés aux modes de représentation des données et à leurs modèles d'analyse. L'étape suivante, visant une stabilisation sociotechnique, a consisté dans une adaptation et une extension des fonctionnalités de l'artefact (Akrich 2006). Si nous avons joué un rôle de médiation entre la rédaction et l'artefact technologique, l'application automatisée a assuré un rôle de médiation avec les données agrégées. Ces activités de médiation, qu'elles soient sociales ou techniques, ont participé au façonnage de l'artefact.

Les premières demandes de modifications ont été exprimées fin février 2017. À ce stade de l'expérience, seule la coordinatrice de la rédaction était impliquée dans ce processus de confrontation, en sa qualité de porte-parole de la rédaction. À cette occasion, elle a proposé que l'interface intègre une carte agrégeant les données des dix stations de mesure bruxelloises, en vue de fournir une représentation géographique des mesures des taux de polluants atmosphériques. Sur le plan journalistique, elle souhaitait examiner si les quartiers bruxellois les plus précarisés étaient ceux les plus touchés par la pollution atmosphérique. Elle s'interrogeait également sur la possibilité de comparer les mesures enregistrées par une station située dans une commune défavorisée de la région avec celles d'une station située dans une commune favorisée. Deux autres points ont également été abordés lors de cette première confrontation : la question de l'hébergement de la plateforme, sur un serveur qui lui est propre (www.bxlairbot.be), et le moment du lancement officiel du projet, dans le numéro du magazine paraissant à la mi-avril 2017.

La "pollucarte" de la Région bruxelloise

Toutes les stations localisées sur cette carte ne mesurent pas nécessairement les mêmes polluants atmosphériques. Pour l'ensemble des polluants, à l'exception du dioxyde d'azote (moyenne glissante pour 1 heure), il s'agit de moyennes glissantes pour 24 heures, elles évoluent donc pendant ce laps de temps, d'heure en heure, avant de devenir une moyenne fixe. Toutes les données figurant sur cette carte interactive proviennent de la Cellule interrégionale d'environnement (CELINE) et sont exprimées en microgrammes/mètre cube ($\mu\text{g}/\text{m}^3$).



FIGURE 3.11 – "Pollucarte" de la région bruxelloise

Sur le plan technique, l'implémentation de ces demandes s'est traduite par la création d'une nouvelle table dans la base de données, dont l'objet était de stocker des enregistrements effectués par tranche d'une demi-heure, puisque cela correspond à la fréquence de mise à jour des pages web sur le site de CELINE. Ceux-ci portaient sur les mesures des différents types de polluants par station de mesure, étant entendu que toutes les stations ne mesurent pas les mêmes types de polluants. En ce qui concerne l'interface, un nouveau *framework* a été utilisé pour générer une carte interactive, intitulée "Pollucarte" (Figure 3.11). Dans l'esprit de faire sens, les marqueurs des stations de mesure changeaient de couleur en fonction du respect ou non des limites fixées par les normes européennes et les recommandations de l'OMS : bleu lorsqu'aucun dépassement n'était constaté pour au moins un des polluants, orange lorsqu'une valeur limite était atteinte, rouge lorsqu'une valeur limite était dépassée. Concernant le dioxyde d'azote, la valeur moyenne annuelle a été choisie comme référence compte tenu de l'impact sur la santé de ce type de polluant et de l'écart important avec la valeur moyenne journalière (où l'on passe de 40 à 200 $\mu\text{g}/\text{m}^3$). Elle a été mise en contexte via un texte expliquant ce parti pris. L'heure de la dernière mise à jour a également été précisée. Un texte d'explication sur les normes européennes et les recommandations de l'OMS a également été ajouté dans un onglet de l'application en ligne, en vue de fournir des informations de contexte. Une autre précision a été apportée à propos de l'absence de valeurs limites pour les particules fines de type *black carbon* (carbone suie), à propos desquelles la recherche a mis en avant le caractère nocif pour la santé (Janssen *et al.* 2011). La demande portant sur une comparaison entre deux stations de mesure a été traduite sous la forme d'un tableau où étaient comparées les mesures des dix derniers jours pour les stations d'Uccle et de Molenbeek, lequel qui figurait sous la "Pollucarte", s'agissant d'un focus sur deux stations de mesure en particulier.

Dans la foulée de ces travaux, un module a été développé pour le système de gestion de contenus (CMS) utilisé par *Alter Échos*. Il avait pour objet de permettre l'affichage du bulletin du jour de "Bxl'air bot" sur le site web du magazine, à partir d'une procédure automatisée (Figure 3.12). Ce module a fait l'objet d'une validation avant son implémentation. Son positionnement dans la colonne du site, qui permet l'affichage de contenus additionnels communs à toutes les pages, variera en fonction des priorités éditoriales accordées à cet espace. À son lancement, le 18 mai 2017, ce module se trouvait en tête de colonne.

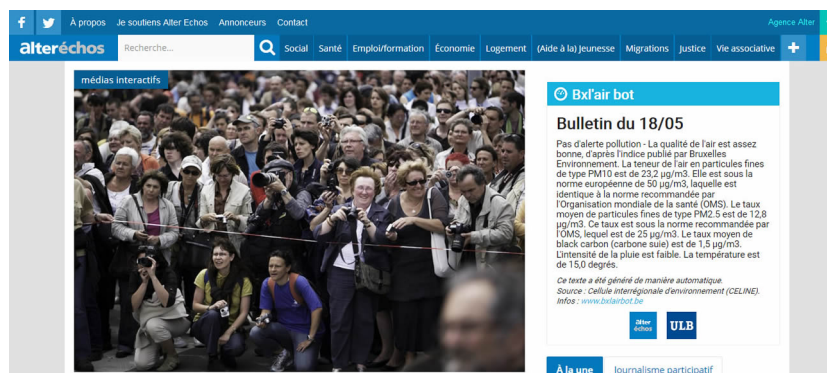


FIGURE 3.12 – Affichage du bulletin du jour de "Bxl'air bot" sur le site d'*Alter Échos*

Lors de ce processus de confrontation, la journaliste indépendante, qui collabore habituellement avec le magazine *Alter Échos* en ce qui concerne le traitement des questions relatives à la qualité de l'air, expliquera qu'elle avait essayé de récupérer les données publiées sur le site de CELINE : *"J'avais fait des captures d'écran de la page d'accueil avec l'indice de qualité de l'air. Ce qui ne servait à rien mais, à l'époque, je ne savais pas où j'allais"*⁴¹.

"Bxl'air bot" a été lancé officiellement le 25 avril 2017. La coordinatrice de la rédaction lui consacre alors un article, publié sur le site web d'*Alter Échos*. Intitulé "Un robot à la rédac", il joue sur la métaphore du robot pour présenter *"une nouvelle recrue"* : *"La rédaction d'Alter Échos a la fierté d'accueillir un petit nouveau dans ses rangs. Il démontre une mémoire infailible des chiffres et travaille 24 h sur 24 sans jamais prendre de pause. Ses dadas : la mesure de la pollution atmosphérique et la cartographie interactive en ligne. Il compose des textes, que l'on saluera, il faut l'avouer, davantage pour leur objectivité et leur précision factuelle que leurs qualités littéraires. Le nom de notre nouvelle recrue ? Il s'appelle Bxl'air bot et c'est... un robot ! Et plus précisément le premier newsbot belge lancé par un média belge francophone. Cette application de datajournalisme a pour objet d'auditer quotidiennement la qualité de l'air dans la capitale (découvrir «Bxl'air bot, le robot qui surveille la qualité de l'air bruxellois») sur la base des données publiées quotidiennement par CELINE, la cellule interrégionale de l'environnement"*⁴². Il s'agit d'une adoption symbolique s'opérant sur un plan discursif et où l'usage de la métaphore

⁴¹ Communication personnelle, Bruxelles, le 24/03/2017.

⁴² "Un robot à la rédac", *Alter Échos*, 25/04/2017, URL : <https://www.alterechos.be/un-robot-a-la-redac/>

du robot confère un aspect sympathique au projet, tout en mettant en avant une image positive du média. S'il y est bien indiqué qu'il s'agit d'une application automatisée, c'est aussi à une approche par données qu'il est fait référence pour légitimer, sur le plan journalistique, la nature de l'expérience.

Lors d'une deuxième rencontre sollicitée par la journaliste indépendante, dans le cadre de son enquête pour le trimestriel *Médor*, celle-ci a évoqué la possibilité de retracer l'historique des dépassements, disponibles sur le site de CELINE mais pour chaque type de polluant et sans comparaison directe possible. Elle demandait également d'intégrer la moyenne des stations bruxelloises concernant le dioxyde d'azote, laquelle n'est pas proposée sur le site de CELINE, contrairement aux autres types de polluants : "*Il y a quand même 9 stations de mesure. Même si deux sont invalides ou à l'arrêt, ça fait quand même pas mal de données. Je serais assez intéressée d'isoler les résultats de la station d'Uccle (qui, au plus j'y pense, au plus me semble être un gros foutage de gueule)*", précise-t-elle⁴³. Ce nouvel entretien a permis d'identifier la manière dont elle envisage d'intégrer l'artefact dans ses pratiques professionnelles : "*Je n'ai pas un profil web du tout. Toutes les possibilités de l'open data, j'en comprends bien l'intérêt mais je n'ai pas une pratique de ces questions et de ces outils. C'est une opportunité, je découvre tout le potentiel de l'outil en le pratiquant. Comme je sais que l'outil concerne les données, je n'ai pas encore cherché à en faire quelque chose car l'outil va le faire lui-même. Ce qui m'intéresse, c'est comment ça fonctionne et pas comment interpréter les données. J'ai plus été voir comment ça marche. (...) C'est un outil qui me permet de relâcher la pression, il y a une conservation des données qui est fiable, scientifique et beaucoup plus précise que ce que je peux faire de manière bricolée. C'est l'outil qui crée l'opportunité plutôt que répondre à un besoin qui était présent*"⁴⁴. Cette journaliste a précisé ses demandes, en termes de fonctionnalités du dispositif, le mois suivant. Elles portaient sur le calcul de moyennes spécifiques par types de polluants. Leur intégration a été envisagée de deux manières : avec la création d'une nouvelle page au sein de l'application, intitulée "Dépassements et normes", qui présentait un historique des dépassements des normes européennes; et avec le développement d'un module supplémentaire dans la page des statistiques, lequel calculait la moyenne des stations par type de polluant et par mois⁴⁵.

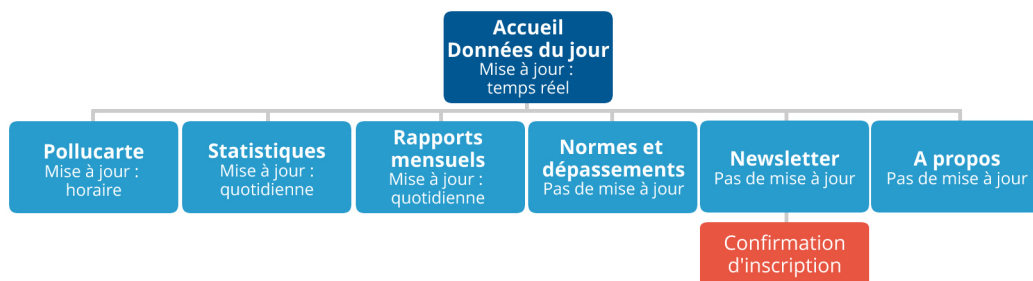


FIGURE 3.13 – Arborescence finale de l'application automatisée

⁴³ Communication personnelle, via courriel, le 30/05/2017.

⁴⁴ Communication personnelle, Bruxelles, le 25/05/2017.

⁴⁵ Voir annexe 14, "Tableau de bord (modules d'analyses statistiques, p.342

Il ne s'agissait pas des derniers développements visant à adapter l'artefact à ses usages finaux qui, par la nature même d'une application librement accessible en ligne, a également composé avec une audience non-journalistique. C'est ainsi qu'un système d'alerte a été développé, à la faveur de la sollicitation d'un internaute qui posait la question de la mise en place d'un système d'alerte automatisé portant sur les dépassements des recommandations de l'OMS ou des normes européennes⁴⁶. Pour la coordinatrice de la rédaction, il s'agissait également d'une opportunité de communication. Ce système d'alerte a pris la forme d'une newsletter envoyée à ses souscripteurs⁴⁷ dès que l'une des recommandations de l'OMS se trouvait en situation de dépassement. Le système a été mis en service le 7 juin 2017⁴⁸ et le premier envoi automatisé a eu lieu six jours plus tard. Cela étant, le système n'était pas tout à fait automatique : si l'exécution du code était réalisée toutes les heures, cela signifie que les destinataires étaient susceptibles de recevoir un e-mail d'alerte jusqu'à vingt-quatre fois par jour en cas de dépassements continus. Il s'agissait donc là d'une limite qui n'a pu être levée autrement que par l'action humaine de validation.

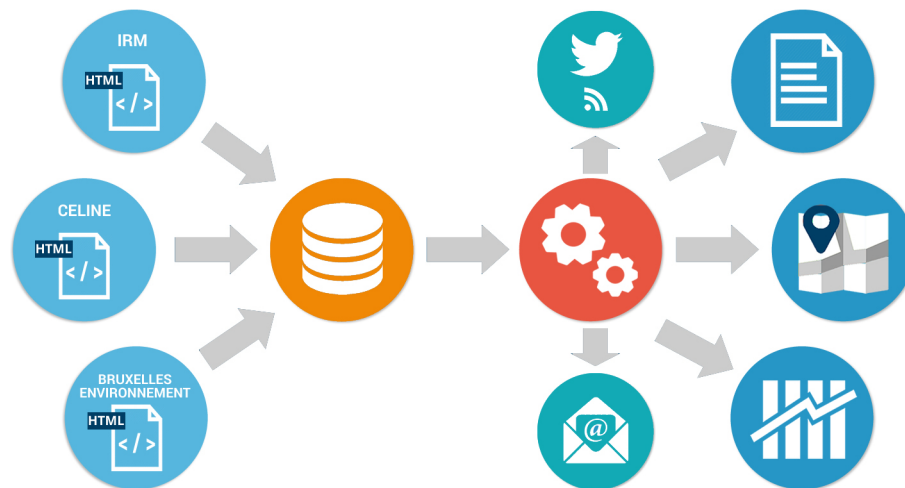


FIGURE 3.14 – Processus d'automatisation des enregistrements et productions automatisés

Pendant trois mois, les interactions avec la coordinatrice de la rédaction et la journaliste indépendante ont permis de faire évoluer l'application, jusqu'à sa stabilisation, tant en termes d'architecture que de fonctionnalités (Figure 3.13, Figure 3.14 et Figure 3.15). Interrogée au sujet de l'interprétation de ses demandes ayant donné lieu à des modifications de l'application, la coordinatrice de la rédaction considère que cela rencontre ses souhaits : "*Pour moi, le sys-*

⁴⁶ Contenu du courriel reçu le 25/05/2017 : "*Tout d'abord, bravo pour votre initiative. Je suis allergique respiratoire et je trouve cela génial! Lutter contre cette pollution passe par la sensibilisation et une information de qualité sur le phénomène. Une question cependant : prévoyez-vous de développer une app ou un flux mail/sms pour prévenir les gens des pics éventuels?*"

⁴⁷ Il comptera un total de 50 abonnés, dont deux membres de la rédaction d'*Alter Échos*. Cela témoigne d'une audience générale relativement modeste : sur une période de douze mois, l'application en ligne "Bxl'air bot" a enregistré une moyenne mensuelle de 110 visiteurs uniques, dont 67,08% proviennent de Belgique (ces données sont fournies via le service Google Analytics, aucune distinction ne pouvant être opérée entre des audiences journalistiques et non-journalistiques).

⁴⁸ Voir annexe 15, "Modèle de système d'alerte automatisé", p.343

tème est stable mais, à la rédaction, nous sommes encore dans une phase où l'on cherche encore comment tirer parti de ces données et avec quelles autres données les croiser. Il s'agit plutôt d'une question relative au comment le faire vivre, journalistiquement parlant⁴⁹.

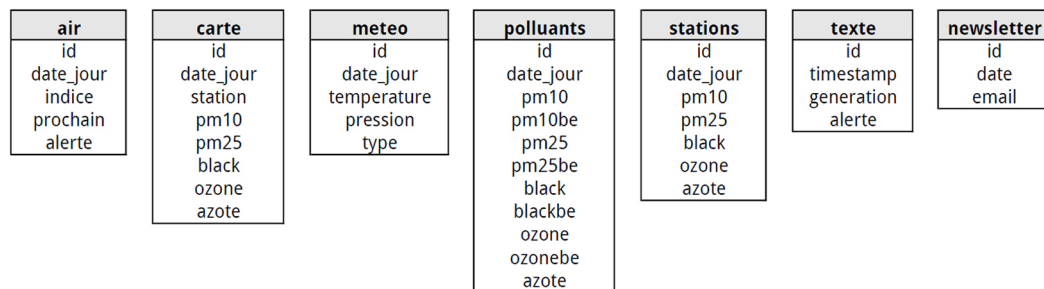


FIGURE 3.15 – Architecture de la base de données

3.2.1 Premières résistances et nouvelles formes d'usages



Le premier contact avec l'ensemble de la rédaction, dans le contexte de cette expérience, s'est produit lors d'une réunion consacrée au projet, le 30 mai 2017. Son objet était de baliser le terrain d'un dossier consacré à la qualité de l'air, à paraître fin avril de l'année suivante. S'il était question de "Bxl'air bot" dans un premier temps, les discussions se sont davantage orientées sur les sujets et angles susceptibles de faire l'objet d'un traitement au regard de la thématique : santé, logement, mobilité, inégalités sociales, mesures citoyennes des taux de pollution, exposition des enfants aux abords des écoles... À cette occasion, un journaliste dira : "À la limite, on n'a pas besoin du robot". Les deux journalistes qui se sont impliquées dans la conception de l'artefact s'en feront alors les porte-paroles : "Le robot permet d'avoir une vue dans le temps", "C'est un point de départ", "Tout le travail de collecte et d'archivage est inédit car les données ne sont pas accessibles, il faudrait plutôt réfléchir à ce qu'il est intéressant de faire avec des données portant sur une période d'un an". Par ailleurs, un consensus s'est dégagé sur la nécessité de réaliser un travail préalable auprès d'experts, compte tenu de la variété des approches journalistiques envisagées.

Dans la foulée de cette réunion, un questionnaire axé sur la perception de l'application "Bxl'air bot" a été envoyé aux six journalistes impliqués dans le projet. Intitulé "Perception et usages du newsbot", ses réponses ont été récoltées de manière anonyme, afin de ne pas influencer sur les opinions des répondants⁵⁰. Il était accompagné d'un second formulaire, "Faire connais-

⁴⁹ Communication personnelle, Bruxelles, le 07/06/2017.

⁵⁰ Voir annexe 8, "Formulaires d'évaluation diffusés via Google Forms dans le cadre de l'expérience 'Bxl'air bot'", p.332

sance", dont l'objet était de recueillir des informations relatives à la situation professionnelle de chaque membre de la rédaction. Quatre journalistes sur six ont répondu à ce dernier questionnaire, dont les deux qui se sont impliqués dans le projet dès ses débuts. Le questionnaire d'évaluation a, quant à lui, été complété par cinq journalistes. Deux d'entre eux n'avaient jamais entendu parler du phénomène du "robot journalisme" avant le début de l'expérience "Bxl'air bot".

Concernant les trois autres répondants, un journaliste s'est dit sceptique à propos du phénomène, tandis qu'un autre estimait que l'outil est intéressant pour traiter des données mais qu'il ne remplacera pas le travail de contextualisation et de "mise en sens" du journaliste. La métaphore du robot est abordée par une autre journaliste, qui indique : "*Plutôt méfiante (robots qui piquent le boulot des journalistes) et opinion amusée aussi (les robots c'est marrant)*". Un répondant a indiqué que son opinion n'avait pas été modifiée en cours d'expérience, un autre que son opinion avait été modifiée en cours d'expérience, et un troisième a coché les deux cases "oui" et "non". Ils justifient leur position de la manière suivante : c'est au journaliste de donner du sens à l'information, les possibilités du dispositif n'ont pas été perçues d'entrée de jeu, et si l'outil peut s'avérer efficace, il pose bel et bien la question du remplacement du journaliste par une machine (Figure 3.16).

Les infos fournies par BXLairBot prendront leur sens grâce au travail journalistique de contextualisation et d'interpellation du pouvoir politique mené sur base de ces données

Je n'avais pas perçu les possibilités que cela impliquait en termes de traitement des données quantitatives (p. ex. sur le long termes)

Oui car le terme "robot" correspond donc à de l'agrégation de données et cela peut être un outil très utile et efficace pour les courageux qui se lancent là-dedans. Mais "non" aussi car la méga menace du robot qui piquera ma place est toujours bien présente

FIGURE 3.16 – Influence de l'expérience sur la représentation du phénomène du journalisme automatisé

Tous les répondants ont affirmé s'être rendus sur le site web de l'application, dont deux à une fréquence d'une fois par mois, un à une fréquence d'une fois par semaine, et deux plus d'une fois par semaine. Un répondant a précisé s'y être rendu plusieurs fois dans le cadre d'un travail journalistique. Un autre répondant a indiqué qu'il manquait un choix : "*Il n'y a pas 'moins d'une fois par mois', et pourtant c'est ce que j'aurais coché. Car pas méga fan de données brutes, de tableaux et de stats. Même si je comprends bien l'utilité du truc*". L'interface a été jugée plutôt claire et lisible, et les informations plutôt accessibles (Figure 3.17). Les données ont généralement été qualifiées de bien expliquées. Les mêmes opinions ont été constatées en ce qui concerne la compréhension des tableaux et infographies. Le traitement de l'information a été perçu comme plutôt pertinent. L'évaluation de la navigation était plus partagée, bien qu'aucun répondant n'ait exprimé avoir rencontré un problème particulier à ce propos.

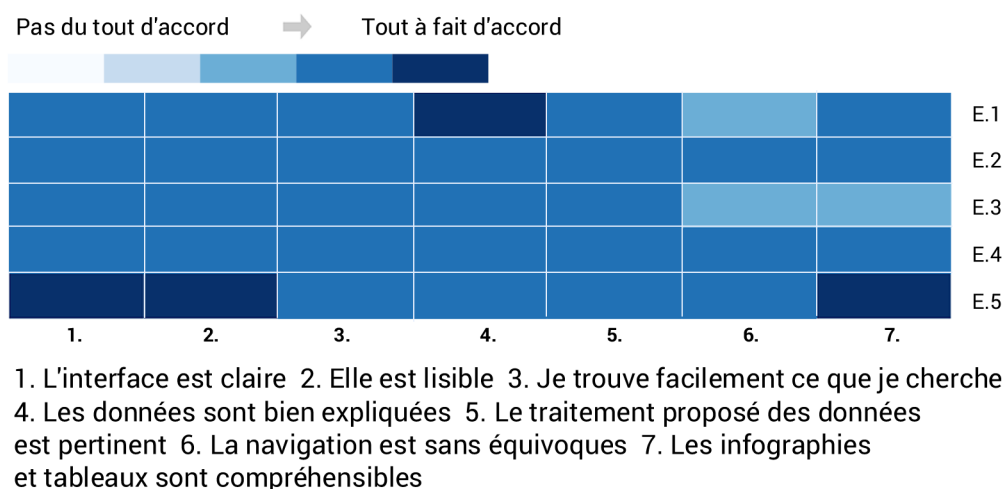


FIGURE 3.17 – Évaluation de l'interface de "Bxl'air bot"

Les journalistes étaient également invités à s'exprimer à propos de leur perception du bulletin du jour, généré de manière automatique en fonction des données en temps réel. Considéré comme plutôt explicite, intelligible et précis, le texte n'a pas été considéré comme particulièrement plaisant à lire ou comme bien écrit. Le caractère objectif du récit a été admis par quatre répondants, tandis qu'un autre répondant était partagé sur la question (Figure 3.18).

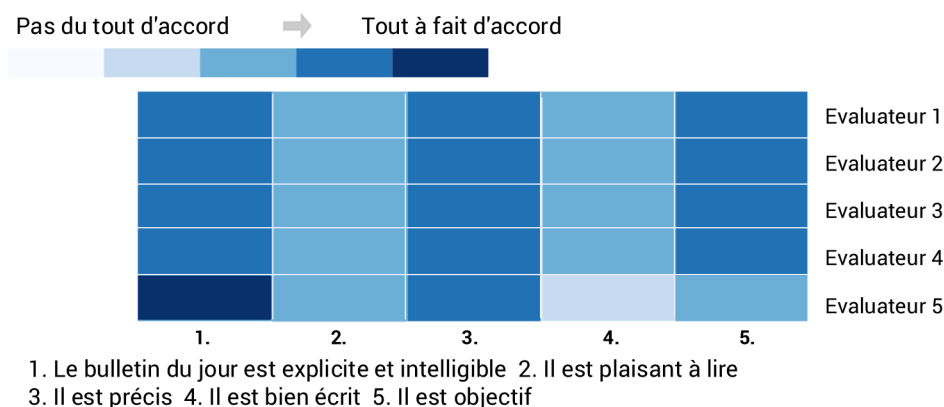


FIGURE 3.18 – Évaluation du bulletin généré en temps réel

Tous les répondants n'ont pas poursuivi le questionnaire jusqu'à son terme, s'agissant de questions ouvertes les invitant à s'exprimer sur des améliorations susceptibles de les soutenir dans leur travail d'enquête. Une seule réponse a été enregistrée pour chacune de ces deux questions : "*L'enjeu pour moi est de comprendre le potentiel de l'outil, de comprendre avec quelles données croiser les infos apportées par le bot pour faire ressortir les enjeux de santé publique. À ce stade, je n'ai pas de réponse à la question des améliorations*" et "*Dans l'immédiat, trouver des statistiques non disponibles ailleurs. À plus long terme, croisement avec d'autres bases de données*". Par ailleurs, une zone pour un commentaire libre avait été prévue : deux répondants ont

choisi de s'y exprimer. Le premier a indiqué que sa participation à cette expérience lui a permis de s'ouvrir plus largement à la réflexion de l'évolution de la profession, mais que la démarche consistant à partir d'un outil pour trouver de l'information est compliquée. Le second a souligné la nécessité d'une " *personne réelle*" derrière le dispositif technologique pour mieux en connaître ses possibilités.

Pour répondre à un souhait qui fut exprimé, au sein de la rédaction, à propos de la manière dont fonctionne la base de données de l'application, une formation a été organisée l'après-midi du 17 juillet 2017. Quatre membres de la rédaction y ont pris part. Le public était exclusivement féminin (les deux journalistes masculins de la rédaction se trouvaient en vacances à cette époque). Cette formation était articulée comme suit : introduction générale sur ce qu'est une donnée, introduction à ce qu'est une base de données phpMyAdmin, introduction aux manipulations de base sur le logiciel Excel et bonnes pratiques en matière de visualisation de données. L'objectif était que chaque participant soit capable de se connecter à la base de données, de réaliser des recherches simples avec l'outil intégré dans le logiciel, d'exporter des tableaux de données, de les ouvrir dans un tableur, d'y effectuer des analyses de base, et de choisir la visualisation de données la plus appropriée. Il s'agissait donc de proposer un traitement des données sans passer par la médiation du dispositif automatisé et de son interface, une démarche pouvant être interprétée par une envie ou un besoin de contrôle humain sur les données brutes.

À l'issue de cette demi-journée, les participantes ont reçu un questionnaire d'évaluation, dont les réponses ont été recueillies de manière anonyme⁵¹. La majorité d'entre elles était motivée par la possibilité d'intégration d'une approche par données dans leurs pratiques professionnelles. Une seule répondante a fait part d'attentes spécifiques liées à l'expérience en cours : "*Comprendre 'Bxl'air bot' pour mieux voir comment l'utiliser dans le dossier. Appréhender des outils qui peuvent m'être utiles dans mon travail*". Cette attente a été rencontrée, affirme cette répondante. L'ensemble des participantes a estimé que les explications étaient claires, que le support écrit qui leur a été fourni était suffisant et que le niveau de formation était "correct".

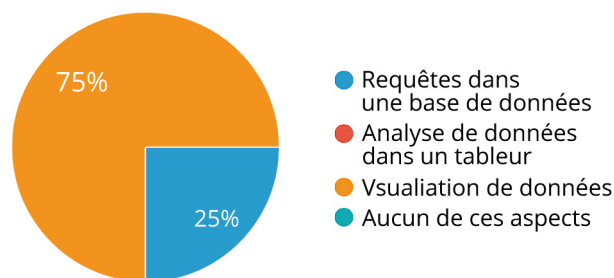


FIGURE 3.19 – Aspects susceptibles de trouver un prolongement pratique

Parmi les points abordés susceptibles de trouver un prolongement dans le cadre de pratiques professionnelles (Figure 3.19), la visualisation de données a été la plus plébiscitée (3 répon-

⁵¹ Voir les questionnaires d'évaluation diffusés via Google Forms à l'annexe 8, p.332

dantes). Deux participantes ont indiqué qu'elles souhaitent approfondir les aspects liés à la visualisation de données et deux autres, ceux relatifs à l'analyse dans un tableur (Figure 3.20).

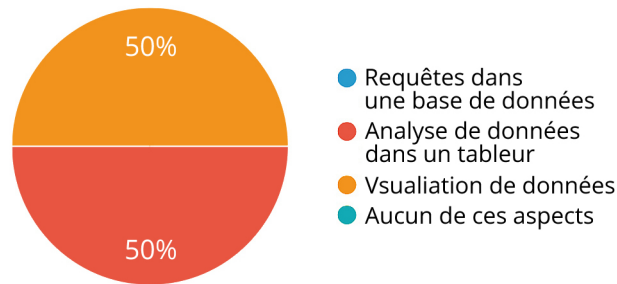


FIGURE 3.20 – Aspects que les journalistes souhaitent approfondir

Trois participantes sur quatre se sont dites "Moyennement satisfaites" de cette demi-journée de formation, tandis qu'une autre s'en est dite "Très satisfaite". Deux participantes ont commenté leur réponse en épinglant le niveau inégal des participantes, qui n'a pas permis d'approfondir certains aspects, et le manque d'informations sur les données manipulées.

Les premières publications relatives aux données de "Bxl'air bot" ont lieu en septembre 2017, d'une part sur le site web d'*Alter Échos* et d'autre part, dans le trimestriel *Médor*. Toutes deux sont signées par la journaliste indépendante. Le premier article, "Qualité de l'air : Nuage de données" (daté du 25/09/2017), a pour objet de faire le point à mi-parcours de l'expérience⁵². "À six mois d'existence, il livre de premières informations inédites", écrit-elle. Il s'agit d'un article qui met en contexte la problématique (première partie) et en lumière l'apport du dispositif d'automatisation (deuxième partie). Les données récoltées dans ce cadre y font l'objet d'une analyse journalistique : "Le Bxl'air bot a collecté suffisamment de données pour livrer de premiers résultats interpellant. Pour le dioxyde d'azote, le robot n'a vu passer, depuis sa naissance, aucune valeur supérieure à 168 µg/m³. Jusqu'ici, on n'atteint donc jamais le pic de 200 µg/m³ par heure, autorisé 18 fois par an par l'Europe. En revanche, si on regarde les moyennes générales, on explose complètement le plafond annuel fixé par l'Europe à 40 µg/m³". Elle souligne également que les données récoltées sont comparées "non seulement avec les seuils tolérés par l'Europe mais aussi avec les recommandations, plus sévères, de l'Organisation mondiale de la santé (OMS)".

La journaliste conclut son article sur l'actualité relative à la problématique de la qualité de l'air en région bruxelloise : "La compilation des données par le robot, entamée en avril dernier, se fait à un moment crucial pour la qualité de l'air à Bruxelles. Le 1^{er} janvier 2018, entre en effet en vigueur la zone de basse émission ou Low Emission Zone (LEZ) dans toute la Région. (...) Progressivement, les véhicules les plus polluants seront interdits d'entrée sur le territoire (...) À partir du mois de janvier (ou après la phase de rodage du système), le Bxl'air bot pourra calculer, mois après mois, l'effet de cette mesure politique". Le dispositif d'automatisation est donc également présenté comme un potentiel instrument de contrôle sur les politiques régionales.

⁵² URL : <https://www.alterechos.be/qualite-de-lair-nuages-de-donnees/>

L'enquête qu'elle publie dans le trimestriel *Médor*, qui traite de la qualité de l'air dans le métro bruxellois – "*Déni de pollution. Bruxelles, un air louche*"⁵³ – comporte deux articles. La journaliste intègre les données et analyses de l'application "Bxl'air bot" dans le premier volet de cette enquête, qui consistent en deux paragraphes en fin de page. Elle indique que la question des stations de mesure en région bruxelloise peut être objectivée grâce à ce dispositif, étant donné que celui-ci prend en compte les mesures effectuées à la station Arts-Loi, lesquelles ne sont pas prises en compte dans le calcul des moyennes régionales. Elle souligne que le dispositif automatisé "*permet aux journalistes d'obtenir des informations inédites (...) Depuis avril dernier, ce logiciel collecte les données de CELINE sur la qualité de l'air, édite ses propres bulletins et recense tous les dépassements. (...) Essayons. Robot, peux-tu comparer la moyenne des concentrations de dioxyde d'azote sur les mois d'avril, mai et juin 2017 avec et sans les stations d'Uccle et d'Arts-Loi? Réponse de la machine : la moyenne de toutes les stations est de 62,77 µg/m³ (l'Union européenne et l'OMS ont fixé la limite annuelle à 40 µg/m³)*". Ce tutoiement avec l'artefact marque à la fois une familiarité (qui serait plutôt sympathique) et une distance induite par la situation de dialogue, bien que les informations générées par le dispositif soient intégrées dans son analyse tout en n'étant pas livrées telles quelles. S'il peut s'agir d'une forme d'appropriation, celle-ci se réfère davantage à un usage de l'outil en tant que source d'informations supplémentaire.

Lors de la préparation de ces deux articles, la journaliste a exprimé des demandes spécifiques relatives à des traitements non proposés par le système d'information, et qui portent sur des calculs de moyennes spécifiques⁵⁴. Dès lors qu'il s'agissait de demandes se rapportant à des cas particuliers, elles ont été traitées via des requêtes manuelles dans la base de données. Ceci témoigne d'une première limite du dispositif : s'il délivre des analyses sur un plan général, tous les cas de figure n'ont pas pu être anticipés, faute d'une projection précise dans les usages finaux du dispositif. Nous observons donc un décalage entre la manière dont l'artefact a été conçu et les usages attendus, lesquels se précisent au fur et à mesure d'une démarche d'enquête journalistique. Ceci avait pu être perçu lors d'un premier entretien avec la journaliste indépendante, qui affirmait alors que l'application automatisée avait créé un nouveau besoin, sans pour autant définir dans quelle mesure les productions automatisées pourraient lui servir sur un plan opérationnel. Par ailleurs, il s'agissait, pour elle, d'une porte d'entrée sur une approche par données dans le journalisme : "*Maintenant, je me rends compte que la technique peut amener un plus, une complémentarité. Lors de la rédaction de mon article pour Médor, j'ai consulté le site à plusieurs reprises : c'est le lien où l'on trouve les informations les plus complètes et la meilleure synthèse. Les données sont disponibles par ailleurs mais elles sont super mal présentées. (...) Je me trouvais alors dans une approche où je me disais que je ferais avec ce qui est disponible. Je n'étais pas dans une approche hyper créative par rapport aux données. Cela m'incite à m'y intéresser davantage et me permet de me rendre compte du potentiel du datajournalisme. L'application est sans doute plus proche des personnes qui recherchent l'information, tandis que CELINE n'est pas pour le grand public et que le site de Bruxelles Environnement est à*

⁵³ *Médor* n°8, automne 2017.

⁵⁴ Communication personnelle, via courriel le 12/09/2017.

moitié bien fait. Elle a donc créé des besoins, tout en révélant un potentiel ou des possibilités"⁵⁵. Au terme de huit mois d'expérience, elle a confirmé cet intérêt pour le journalisme de données : *"Cela m'a donné envie d'explorer le potentiel du data, même si je ne veux pas me mettre moi-même au codage. Je développe actuellement avec Médor un petit projet de datajournalisme autour de la qualité de l'air (carte interactive pollution/emplacement des écoles)"*⁵⁶. Ce témoignage met en lumière un premier impact de l'expérience sur ses pratiques professionnelles.

S'agissant de la coordinatrice de la rédaction, qui ne traite habituellement pas des questions environnementales, cette expérience s'inscrivait dans une double perspective : celle de l'innovation numérique et celle de la participation à un projet de datajournalisme. *"Je ne connaissais pas bien ces robots, le seul dont j'avais entendu parler était celui du Monde pour les élections. Notre métier est en train de se bouleverser, et les technologies sont pour nous une menace ou un outil. J'étais curieuse de participer à cette expérience car elle participe à une évolution du métier. Dans mon rôle, mon objectif est aussi de faire aussi évoluer Alter Échos sur le web, avec une équipe qui se trouve un peu dans une fracture numérique et où les filles sont plus sensibles aux technologies. Le fait que l'Agence Alter ait toujours été innovante, couplé au fait que son modèle économique soit un peu précaire, nous a toujours poussé à l'expérimentation. C'est un peu dans notre ADN. Je suis aussi très intéressée par le journalisme de données. C'est l'occasion de comprendre comment cela fonctionne de manière concrète, et de me former à ça en me disant que je pourrai utiliser ces acquis par la suite"*⁵⁷.

Elle indique également que, pour la rédaction, il ne s'agissait pas d'un projet facile à comprendre : *"surtout quant à la manière dont ça peut les aider dans leur travail. On m'a fait part d'incompréhensions quant à la manière dont fonctionne un projet de datajournalisme mais je n'ai pas constaté de résistance. Le questionnement est de voir comment cet outil va pouvoir être utilisé. Ce n'est pas lié au robot mais à la manière de travailler avec des données car on n'en a pas l'habitude ici"*. En ce qui la concerne, à ce stade de l'expérience, il s'agirait donc davantage d'une adoption symbolique de l'objet – relative à son acceptation mentale (Verra *et al.* 2012) – couplée à un usage utilitaire de l'outil. Considérant que l'adoption d'une innovation technologique implique au moins deux décisions, celle de l'acceptation de l'idée et celle de l'utiliser ou non (Klonglan & Coward 1970), il s'agit d'une forme d'appropriation mais qui ne s'inscrit pas dans le cadre de pratiques professionnelles. Cette adoption symbolique est aussi celle d'une communication au public, qui contribue à donner une image innovante et positive du média.

⁵⁵ Communication personnelle, Bruxelles, le 23/05/2017.

⁵⁶ Communication personnelle, via courriel le 26/12/2018.

⁵⁷ Communication personnelle, Bruxelles, le 07/06/2017.

3.2.2 Clôture et évaluation de l'expérience



La deuxième réunion de rédaction relative à la préparation du dossier "qualité de l'air" a été organisée le 6 mars 2018. Elle s'est déroulée dans le contexte du passage d'*Alter Échos* à un rythme mensuel (et non plus bimensuel) et de l'adoption d'une nouvelle maquette. Dans la foulée, chaque numéro sera désormais piloté par une rédaction en chef tournante, composée d'un duo de journalistes. Cette réunion a permis de faire le point sur les dernières actualités du domaine d'application, mais aussi d'ébaucher des premières pistes pour un sommaire. Seule la journaliste indépendante a indiqué qu'elle allait travailler en s'appuyant sur les données et analyses fournies par "Bxl'air bot", en élargissant le propos aux pics atypiques ayant pu être observés ou à l'impact de journées de grève sur les mesures observées. La question de comparaisons possibles avec des mesures effectuées par des mouvements citoyens a aussi été évoquée. Pour cette journaliste, il s'agissait de "*se demander ce que le robot peut apporter*", indiquant qu'il faudrait "*travailler en fonction des causes et des conséquences*".

La troisième et dernière réunion de rédaction consacrée au dossier "qualité de l'air" a été organisée le 17 avril 2017, soit un mois avant la publication du numéro. Le chemin de fer y a été passé en revue : quatre pages sur quatorze allaient être consacrées aux données et analyses de "Bxl'air bot". La journaliste indépendante a indiqué que la station de mesure Arts-Loi sera prise en compte dans son article et les graphiques qui allaient l'accompagner, et que les mesures y seront comparées avec les recommandations plus sévères de l'Organisation Mondiale de la Santé, confirmant les choix journalistiques posés lors de la conception de l'artefact. "*Il serait aussi intéressant de voir s'il y a un effet particulier le dimanche, en comparaison avec les taux observés pendant la semaine*", a-t-elle ajouté, émettant l'idée de réaliser une "*journée type du respirant bruxellois*". À ce stade, tous les sujets n'étaient pas encore déterminés, en raison de la variété des possibilités de traitement : entre la problématique de l'isolation du bâti bruxellois, celle des inégalités sociales face à la pollution, celle des mesures politiques visant à réduire les taux de polluants atmosphériques, celle des véhicules diesel, ou encore celle des mesures citoyennes.

Dans le cadre de la préparation de son article, la journaliste indépendante va poser de nouvelles demandes spécifiques : "*Comme j'imagine que le robot trie les résultats, calcule et pige plus vite que moi (et que c'est à ça qu'il sert), je t'envoie la totale (...)* :

- résultats pour tous les polluants confondus, heure par heure, pour les 5 jours de la semaine d'une part, pour les 2 jours du WE d'autre part. Objectif : faire une journée-type du respirant bruxellois et indiquer à quelles périodes il vaut mieux ne pas respirer.
- moyennes mensuelles de l'indice de la qualité de l'air de l'IBGE + moyennes mensuelles par polluant (hors BC). Objectif : comparer la com' officielle avec résultats réels et avec les normes OMS et Europe, voir si la situation a changé depuis l'introduction du nouveau mode de calcul

- moyenne annuelle (pour la période de Bxl'air bot) du dioxyde d'azote avec Arts-Loi et moyenne sans Arts-Loi. Objectif: comparer avec et sans Arts-Loi, en regard des normes Europe et OMS.

- moyenne annuelle du dioxyde d'azote (c'est le polluant sur lequel c'est le plus flagrant, non?) pour les jours de la semaine, moyenne du samedi, moyenne du dimanche + résultats pour certaines journées : 25/12/17 (Noël), 17/09/17 (Journée sans voiture), 09/10/17 (fermeture Hermann-Debroux), 10/10/17 (grève STIB + fermeture Hermann-Debroux) + météo pour ces journées. Objectif: comparer la moyenne semaine, samedi et dimanche + certains jours particuliers (en nuanciant avec la météo)"⁵⁸.

À l'instar des demandes particulières exprimées lors de la rédaction de ses deux premiers articles intégrant les productions de "Bxl'air bot", ces nouvelles demandes ont été traitées manuellement via des requêtes dans la base de données ou dans un tableur Excel. Cette contribution humaine n'est pas perçue en tant que telle par la journaliste, qui impute les résultats qui lui sont fournis au seul système d'information : "Quelle efficacité (...) J'aurais bien besoin d'un robot dans ma vie"⁵⁹.

DOSSIER

Pollution, l'air de rien



16 — 19

Un an avec un robot

Pendant douze mois, *Alter Échos* a accueilli à la rédaction le Bxl'air bot, une application de datajournalisme qui nous a aidés à traiter des données sur la qualité de l'air. Le résultat? De l'info inédite sur la pollution à Bruxelles.

20 — 22

Sommes-nous tous égaux face à la pollution de l'air?

Une étude menée dans plusieurs villes européennes analyse le lien complexe entre pollution de l'air et inégalités sociales. En Belgique, la question commence à être timidement étudiée.

23

« La pollution atmosphérique est sous-estimée en Wallonie »

Greenpeace poursuit en justice les gouvernements wallon et flamand pour la pollution de l'air. En attendant le jugement, attendu fin de l'année, Joeri Thijs (Greenpeace) revient sur les manquements de la Wallonie.

24 — 25

Air wallon : Carlo Di Antonio « prend des dispositions »

Le ministre wallon de l'Environnement Carlo Di Antonio estime que les mesures de la pollution de l'air sont suffisantes en Wallonie. Conscient du problème de santé publique, il a toutefois pris plusieurs dispositions pour améliorer la qualité de l'air.

26 — 27

Low Emission Zone : la zone qui cache la forêt

Depuis le 1^{er} janvier, les vieux diesels n'ont plus droit de cité à Bruxelles où une zone de basse émission a été instaurée. Une mesure suffisante? La ministre Fremault ouvre la porte à une interdiction totale du diesel après 2025.

28 — 29

Souffle citoyen sur Molenbeek

Les citoyens sont de plus en plus nombreux à se mobiliser contre la pollution de l'air. Zoom sur deux initiatives menées dans la commune de Molenbeek.

FIGURE 3.21 – Sommaire du dossier publié dans *Alter Échos* n°464

"Pollution, l'air de rien" est le titre final du dossier publié à la mi-mai 2018 dans l'*Alter Échos* n°464. "Bxl'air bot" apparaît dans la présentation du sommaire de ce dossier : "C'est avec cette même envie de connaissance que nous avons accueilli à la rédaction d'*Alter Échos* un robot de

⁵⁸ Communication personnelle, via courriel, le 23/04/2018.

⁵⁹ Communications personnelles, via courriel, les 25/04/2018 et 04/05/2018.

collecte et d'analyse de données liées à la qualité de l'air. Il nous permet aujourd'hui de poser un œil critique sur les informations communiquées par les autorités bruxelloises". Ce sujet ouvre le dossier, il s'agit du seul à traiter des données récoltées (Figure 3.21). Cet article est également publié en ligne, à la date du 14 mai 2018⁶⁰. L'illustrateur de ce numéro, Adrien Herda, fait usage de la métaphore d'un robot interviewé par des journalistes dans l'image choisie pour ouvrir le l'article de la journaliste indépendante (Figure 3.22).

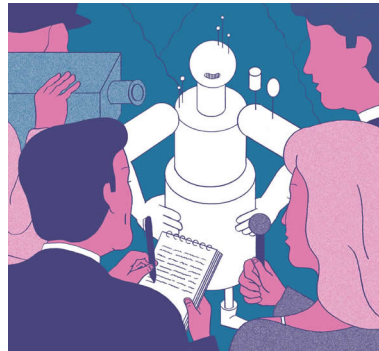


FIGURE 3.22 – Métaphore du robot interviewé par des journalistes

Le dispositif automatisé apparaît ainsi comme une source d'informations supplémentaire⁶¹. Comme dans ses précédents articles, la journaliste indépendante n'utilise que la seule métaphore du robot pour désigner l'application, un usage sans doute induit par son appellation "Bxl'air bot" ainsi que par l'activité journalistique de vulgarisation. La journaliste joue également sur cette métaphore pour présenter l'artefact d'une manière sympathique : *"Le Bxl'air bot ne prend pas beaucoup de place et ne sert pas le café. Il s'agit d'une simple application qui a fait son nid sur notre site internet"*. Dans le chapeau de l'article, elle indique qu'il s'agissait d'une première expérience "humano-robotique", ce qui peut être interprété, en ce qui la concerne tout au moins, comme un partenariat homme-machine qui serait ici réussi. Toutefois, elle précise que l'initiative n'émane pas de la rédaction, mais que celle-ci a été sollicitée pour participer à cette expérience : *"Il a mis le pied dans la porte de la rédaction en mars 2017 sans qu'on ne l'ait vraiment invité..."* Il est précisé, plus loin, que *"Du 1^{er} avril 2017 au 31 mars 2018, ce 'baby bot' a audité, chaque jour, la qualité de l'air dans la capitale"*. L'idée du partenariat homme-machine revient ensuite, dans une mise en relation avec l'avantage perçu du dispositif : *"L'intérêt du robot, c'est qu'il peut automatiser des calculs que les journalistes pourraient faire manuellement avec les données de CELINE mais qu'ils n'auraient – pour être honnêtes – jamais le temps et la patience de faire"*.

Le concept d'objectivité est également évoqué, considérant qu'il s'agit d'un idéal non réalisé dès lors que toute activité d'analyse de données implique une part humaine de subjectivité : *"Frondeur dans l'âme, notre petit robot nous fournit donc une moyenne annuelle de la concentration en dioxyde d'azote à Bruxelles en tenant compte des mesures d'Arts-Loi. (...) Si le dioxyde*

⁶⁰ URL : <https://www.alterechos.be/un-an-avec-un-robot/>

⁶¹ Voir annexe 16, "Article publié dans le dossier 'Pollution, l'air de rien', p.344

d'azote (NO₂) était un gaz hilarant, à Bruxelles, on serait déjà tous morts de rire". Dans le même temps, la métaphore est à nouveau utilisée pour donner une image sympathique du dispositif, qui serait celle d'un gant. La subjectivité de l'analyse des données est encore soulignée lorsqu'il s'agit d'établir des comparaisons avec les recommandations plus strictes de l'OMS. Si ces deux choix n'avaient pas été posés, les résultats des calculs des moyennes et du nombre de jours de dépassement auraient été différents. "Tout comme les autorités, qui interprètent les données d'une façon rassurante pour le public (...), le robot assume la part de subjectivité que comporte toute interprétation d'informations".

Son analyse se poursuit au fil de quatre graphiques : celui des moyennes du taux de dioxyde d'azote, qui illustrent une situation de dépassement des normes annuelles, y compris le jour de l'an; celui de la moyenne annuelle du taux de NO₂ avec et sans la situation Arts-Loi, qui témoigne d'une situation de dépassement dans tous les cas, mais qui est accentuée en tenant compte de l'ensemble des stations de mesure; une comparaison avec la communication des autorités, qui apparaît comme rassurante au regard des résultats enregistrés (*"Si l'on fait la moyenne par mois des indices de la qualité de l'air tels qu'ils sont communiqués par Bruxelles Environnement, on se dit que la pollution, il n'y a vraiment pas de quoi en faire un dossier spécial. Le pire du pire à Bruxelles, en moyenne mensuelle, sur la période d'enregistrement, c'est que l'air soit 'assez bon'"*); et la journée-type du respirant bruxellois, où il apparaît que les taux les plus bas sont enregistrés, en moyenne, le dimanche à 3 heures du matin. À propos de ce dernier graphique, la journaliste précise la méthode de collecte et le mode de calcul : *"Ce graphique est réalisé grâce aux moyennes de toutes les stations. Il ne reflète donc pas la situation des Bruxellois les plus pollués ni celle des moins pollués, mais celle du respirant 'moyen' dans un quartier 'moyen'. Le robot tient compte de deux enregistrements par heure (et non d'un enregistrement en continu) avec, parfois, des résultats inutilisables. Il s'agit donc d'une information précieuse, qui met en lien la pollution et l'activité humaine, mais qui mériterait d'être nuancée par d'autres types de mesures"*. Entre distance et appropriation, cet article, à la rhétorique nuancée, dresse un portrait sans complaisance de la qualité de l'air bruxellois. S'appuyant sur l'autorité du nombre et l'argument statistique, il assume indirectement la part de subjectivité du dispositif "Bxl'air bot" – qui résulte aussi de choix posés par la journaliste – tout en expliquant la méthode choisie, dans le souci d'une démarche de transparence.

Les entretiens d'évaluation avec l'ensemble de la rédaction ont été réalisés un peu plus d'un mois après la publication de ce dossier⁶². Ils ont permis de mieux cerner le positionnement de l'ensemble des journalistes. En premier lieu, ils n'ont pas eu le sentiment d'avoir passé "Un an avec un robot", pour paraphraser le titre de l'article de la journaliste indépendante : *"Il a occupé une petite place parce qu'on en a parlé dans des réunions et qu'il y a eu le numéro spécial qui est sorti mais pas beaucoup. Je ne m'en suis pas beaucoup saisi"* ; *"Une petite place car je le consultais de temps à autre. Il a joué un rôle dans le dossier, ça c'est clair. Dans ma pratique journalistique, je trouve que c'est à ce moment-là qu'il a vraiment pris sa place car on a pu voir les données qu'il mettait à notre disposition"* ; *"Je pense qu'on a pris le truc un peu au dernier mo-*

⁶² Soit le 26/06/2018, voir aussi annexe 9, "Guide d'entretien pour l'évaluation finale de l'expérience", p.334

ment, on avait des délais à respecter" ; "Il avait une place visuelle sur le site" ; "Il y a des moments où il était présent, ceux où on avait prévu de faire un article ou quand il y avait de l'actualité dans d'autres médias, et des moments où je l'ai oublié".

Les journalistes se sont rendus assez peu souvent sur le site web de l'application, par manque de temps, d'intérêt pour la thématique ou pour le fait chiffré. Ils ne considèrent pas que cela constitue une forme de journalisme, les données ne prenant sens qu'à partir du moment où elles sont traitées par un journaliste : *"C'est quelque chose qui peut alimenter du journalisme, c'est comme si un listing de notes serait du journalisme" ; "Pour moi ça sert de terreau pour le journalisme, une base de datajournalisme sur laquelle on peut construire du journalisme enrichi" ; "L'utilisation qui en a été faite, c'est du journalisme (...), le traitement journalistique l'a fait exister comme source" ; "C'est une forme de journalisme si on s'empare des données (...) et qu'on les traite de manière journalistique" ; "C'est un outil, c'est comme si on me disait que mon stylo est un journaliste. La météo, ce n'est pas du journalisme. Si un scientifique m'explique le réchauffement climatique, c'est du journalisme".* La coordinatrice de la rédaction précise que l'outil est utile pour remplacer une série de tâches répétitives, de manière à libérer du temps pour les journalistes. *"Mais ça, ce n'est pas gagné : la presse va très mal. Quand je travaillais chez Métro où je faisais du traitement de dépêches, j'avais un peu l'impression de faire du travail de robot (...) C'est un métier qui change vite, à la fois pour le meilleur et pour le pire : on peut perdre des emplois d'un côté et, de l'autre, voir de nouvelles possibilités qui s'ouvrent et (...) c'est fascinant".*

Le dispositif "Bxl'air bot" n'a pas eu d'impact sur les pratiques professionnelles de la rédaction : *"Je vois cette montée en puissance sur la donnée et je suis complètement à côté de ça (...) Je ne suis pas certain que ma génération néo-quarantenaire sera obligée d'y passer mais pour les suivantes, ce seront des outils de travail" ; "J'ai vu ça comme un outil parmi d'autres, encore marginal (...) mais je vois ça comme une opportunité et pas menace" ; "C'est un outil qui ne remplacera pas le journaliste car il n'a pas la même capacité d'analyse et d'esprit critique, mais c'est un outil intéressant pour compiler des données. En interne, on manque de compétences sur les chiffres" ; "Je trouve que c'est difficile de trouver quelque chose d'intéressant quand on se sent complètement paumé par rapport à l'outil. C'est un nouvel outil qu'il faut s'approprier, donc ça demande des formations et du temps (...) C'est compliqué, ardu".* Toutefois, pour la coordinatrice de la rédaction, cette expérience l'a amenée à s'ouvrir davantage aux pratiques du journalisme de données : *"Pour moi, un fichier Excel, ce n'était pas une source d'informations. J'ai travaillé sur une enquête relative à la pénurie d'enseignants et j'ai essayé de me construire un tableau pour voir si cela avait un impact sur les inégalités scolaires mais ça n'a rien donné. Maintenant, lorsque je reçois un PDF, je peux le convertir en Excel et le nettoyer. Le robot m'a sensibilisée à ça. Jusque-là, pour moi, ce n'était pas du tout évident".*

Malgré des réserves liées au domaine d'application et à une approche par données plus générale, la rédaction est unanime pour tirer un bilan globalement positif de l'expérience. Tous les journalistes se disent prêts à la renouveler mais, cette fois, sur un sujet qui serait plus proche de leurs centres d'intérêts personnels ou professionnels. *"Pour moi, l'expérience était claire-*

ment positive (...) S'il faudrait la refaire, peut-être que j'essaierais de m'impliquer un peu plus"; "L'évaluation est positive mais peut-être qu'on ne s'est pas assez emparés de l'outil. Pour moi, le challenge est de savoir comment ce genre d'outil peut être plus qu'un énorme adjuvant sur un dossier particulier, il peut aussi s'inscrire dans une pratique quotidienne (...) Ça peut mâcher le boulot et croiser des choses que tu n'aurais pas pu faire toi-même"; "Je trouve que c'est un bilan positif mais si le sujet avait été plus proche de mes matières, j'aurais essayé de farfouiller. Je ne suis pas hyper à l'aise avec les données (...) Je pense que je le referais sur un sujet que je connais mieux"; "Oui je serais partante (...) Je pense qu'ici, on en a fait quelque chose de bien"; "C'était très positif. Ce qui était difficile dans la démarche, c'est qu'on me donne un outil pour que l'on observe comment je l'utilise. Je suis partante pour recommencer, mais sur un sujet qu'on choisirait en rédaction. Pour moi, il y avait une barrière dans la thématique. Elle m'a permis de découvrir les exigences du journalisme scientifique (...) qui était la partie que j'ai trouvé la plus compliquée (...) Pour Alter, cela a donné une image dynamique et moderne".

L'évaluation finale de l'expérience s'est déroulée de manière séparée avec la journaliste indépendante. Dans le bilan qu'elle en dresse, elle estime que cela lui a ouvert de nouvelles perspectives professionnelles : "J'ai découvert une opportunité d'élargir complètement le champ du travail journalistique en y amenant d'autres types d'infos et d'interprétations qui m'étaient méconnues et, il me semble, globalement méconnues et sous exploitées dans le monde des médias (...) On n'a pas, autour de moi, le degré de conscience de ce que le journalisme de données peut apporter (...) mais on est encore à des balbutiements. En même temps, il faut rester prudent et veiller à ce que ça ne devienne pas un gadget ou quelque chose qu'on fait pour la forme. Il faut toujours se demander en quoi ça sert le journalisme. Il faut voir aussi si les moyens humains et financiers qui y sont consacrés sont à la hauteur de ce qui a été investi. Ici, on aurait peut-être pu en faire plus, mais je trouve que le résultat est super intéressant"⁶³.

Si cette expérience était à refaire? Elle se dit enthousiaste à l'idée, "mais il ne faut pas que la technologie devienne une fin en soi, plutôt une façon de concevoir un angle. Cela doit rester au service de l'information". Sur le plan de ses pratiques professionnelles, cela l'a amenée "à reconsidérer la manière de penser ou concevoir un sujet. Cela m'a ouvert au monde des données et quand je vois le résultat final de cette expérience, je trouve cela très chouette. Par définition, les données sont quelque chose de sec et d'ennuyeux. Ici, on est arrivé à les exploiter d'une manière qui les rende agréables à lire pour le lecteur". Elle estime également qu'une médiation humaine reste importante "parce que, malgré tout, le robot ne fait pas tout. Il faut bien interpréter et il y a une approche scientifique à avoir. Au-delà du développement, tu as acquis (s'adressant à la chercheuse, ndlr) une vraie connaissance du sujet qui était fondamentale et qui a permis de ne pas se cantonner au binôme technique-journaliste". Aussi, la journaliste indépendante avait-elle transmis préalablement chacun de ses articles pour une relecture : cela a donné lieu à plusieurs échanges autour de la thématique. Ceux-ci lui ont donc paru tout aussi important que l'application automatisée. Cet apport de la chercheuse, qui dispose d'une expérience professionnelle à la fois en développement et en journalisme, a permis à cette journaliste d'envisager

⁶³ Communication personnelle, Bruxelles, le 9/08/2018.

"Bxl'air bot" à la fois comme un outil et comme un objet du journalisme, "*dans la mesure où il y à la fois de la collecte de données mais dans une interprétation qui est le fruit d'une réflexion d'un être humain journaliste. Sa conceptrice était engagée dans une démarche journalistique. Tous les fondamentaux du journalisme sont donc garantis. Avoir un journaliste en amont et en aval du projet renforce le fait que ce robot fasse du journalisme*", conclut-elle.

L'interface de "Bxl'air bot" a été modifiée début mai 2018 puis remaniée en octobre 2018, de manière à clôturer l'expérience dans l'attente de la fin du contrat d'abonnement web qui s'est achevé début avril 2019. "Bxl'air bot" a trouvé une seconde vie, temporaire et dans une version simplifiée, dans le cadre des "États généraux de l'air de Bruxelles", qui se sont déroulés les 25 et 26 avril 2019⁶⁴.

3.3 Analyse des observations

L'expérience "Bxl'air bot" s'est déroulée dans le contexte d'une rédaction peu technophile et témoignant d'assez peu d'appétence pour le fait chiffré et une approche par données dans le journalisme, certains journalistes n'hésitant pas affirmer que cela leur fait peur. L'usage de technologies de l'information et de la communication y est perçu tantôt comme chronophage, tantôt comme mobilisant des compétences qui nécessitent du temps tant dans leur apprentissage que dans leur mise en pratique. Plusieurs journalistes justifient cette posture par un rapport à l'écrit qui relève, selon eux, d'une vision traditionnelle du journalisme où le travail de terrain constitue le cœur de leurs pratiques professionnelles. Bien que pouvant apparaître, *a priori*, comme un terrain de recherche peu favorable au déploiement d'un tel projet, celui-ci s'est toutefois révélé riche en enseignements tant sur le plan des usages, que sur celui de non-usages qui n'ont pas été réalisés par quatre journalistes de la rédaction.

La mise en perspective du cycle de vie du logiciel avec le modèle SCOT a permis de suivre le développement de l'expérience tant sur son plan social que technique, dans un processus itératif où la clôture de chaque étape a permis de nourrir la suivante, démontrant que le processus technique résulte de choix sociaux. Aussi, la construction sociale et la construction technique de l'artefact sont-elles intrinsèquement liées : en ce sens, le dispositif "Bxl'air bot" ne peut être considéré comme purement mécanique. Chacun des choix posés au cours de ce processus relève du domaine des intentions, lesquelles doivent être définies, comprises interprétées puis médiées. La boîte noire technologique ne pourrait donc être comprise autrement qu'à travers celle de la compréhension de la boîte noire journalistique : des choix journalistiques vont dépendre les choix techniques, s'agissant d'un façonnage mutuel. Dans un souci de transparence, ceux-ci seront expliqués et justifiés, que ce soit via l'application automatisée accessible en ligne ou via les articles signés par la journaliste indépendante spécialisée dans les questions relatives à la qualité de l'air. Ce souci de transparence était piloté par l'hypothèse selon laquelle l'usage d'un objet technique ne dépend pas seulement de sa représentation, mais aussi de sa compréhension. Celle-ci est parfois apparue comme ardue pour les utilisateurs, comme cela fut

⁶⁴ URL : <https://www.brusselsair.org/inex/bxlairbot/>

constaté lors de la formation organisée autour du traitement des données de l'application. La participation à cette demi-journée pourrait, en outre, être interprétée comme un moyen pour le journaliste de dominer la machine. Toutefois, celle-ci ne pouvait être que partiellement réalisée, étant donné la complexité perçue quant à l'acquisition de compétences qui nécessite un temps d'apprentissage et de mise en pratique.

Aux prémices de la construction sociotechnique de l'artefact "Bxl'air bot", se trouve un investissement humain important : au-delà d'une mobilisation de compétences techniques, l'expertise du domaine d'application s'est posée en préalable pour évaluer la pertinence des données ainsi que la manière dont elles seront traitées, à l'instar de n'importe quel autre type de projet s'appuyant sur un traitement automatique de la langue (Reiter & Dale 1997). Cela a nécessité de documenter la problématique, en ce compris en sollicitant des experts d'un domaine d'application complexe, en raison de son caractère scientifique. L'expérience de la technologie témoigne de la difficulté de se baser sur des données publiques, certes pouvant être utilisées librement, mais disponibles dans le contexte du format potentiellement mouvant de pages web. La gestion des anomalies a également nécessité un travail non négligeable, via des activités continues de vérification des données, qui ont pu donner lieu à des modifications du code informatique. Sans ces activités de maintenance, les objectifs journalistiques de fiabilité et de précision n'auraient pu être rencontrés. Aussi, le challenge s'est-il posé en termes techniques et journalistiques, l'automatisation ne pouvant être considérée comme la principale composante du système d'information.

L'étape de flexibilité interprétative, telle qu'elle est définie dans le modèle SCOT, montre la difficulté de se projeter dans des usages finaux lorsque la culture d'une approche par données n'est pas déjà ancrée dans les pratiques professionnelles. La principale pierre d'achoppement consistait à nourrir une démarche journalistique à partir de données susceptibles de constituer un angle, alors que les démarches habituelles sont celles de définir un angle avant d'entreprendre un travail engageant des sources d'information. Par la logique qu'il induit, le dispositif d'automatisation a donc potentiellement bousculé une manière de faire habituelle. Si cette expérience concernait six journalistes, deux seulement se sont impliquées dans le processus de construction sociale de "Bxl'air bot", le dispositif évoluant au rythme de la découverte de ses potentialités. Pour la coordinatrice de la rédaction, il s'agissait de découvrir les pratiques du journalisme de données, de réfléchir à l'avenir de la profession, mais aussi de s'investir dans un projet considéré comme innovant et susceptible de donner une image positive du média, auquel il est reconnu un manque de notoriété. Pour la journaliste indépendante, le besoin de l'outil n'existait pas avant que celui fasse son apparition. En ce sens, le bénéfice du dispositif a été considéré comme un moyen de disposer d'une source inédite. Son autre apport – qui n'avait pas été envisagé au début de cette expérience – fut de sensibiliser ces deux journalistes à une approche par données plus générale, les amenant à considérer cette option dans leurs pratiques professionnelles futures.

Les demandes exprimées par ces deux journalistes ont contribué à forger un outil qui apparaît comme incomplet, dès lors que la projection dans les usages finaux n'était pas particulièrement aisée à envisager dans le contexte d'un journalisme au long cours. Nos activités de médiation ont été jugées importantes pour le bon déroulement de l'expérience, s'agissant de conférer à l'artefact un caractère "humain" duquel il ne peut être détaché, et qui peut aussi avoir joué un rôle rassurant pour ces deux journalistes. Celles-ci furent également les deux seules à avoir fait usage du système d'automatisation, autrement que dans une perspective qui serait purement fonctionnelle. Nous pouvons également considérer que l'implication de ces deux journalistes dans le processus de construction sociotechnique de l'artefact les en a fait devenir des actrices de l'innovation (Akrich 2006), et cela dès le moment où s'est engagée l'étape de confrontation. Elles n'ont jamais perçu le dispositif avec réticence mais, au contraire, comme constituant une source d'opportunités.

Bien que les quatre autres membres de la rédaction d'*Alter Échos* étaient également concernés par le projet, dans la mesure où celui-ci allait faire l'objet d'un dossier les impliquant également, ceux-ci n'ont pas pris part à ce processus inclusif. Cette non-implication pourrait expliquer la distance qui s'est créée entre eux et le projet. Celle-ci sera notamment constatée au cours de la première réunion de rédaction, qui s'est tenue un mois après le lancement officiel du dispositif, au cours de laquelle une première forme de résistance explicite sera constatée lorsqu'un journaliste dira : "*À la limite, on n'a pas besoin du robot*". Cette réunion de rédaction marque pourtant le début de l'étape de stabilisation de l'artefact. Si ces quatre journalistes s'étaient impliqués dans le processus de construction sociotechnique de l'artefact, en auraient-ils fait un usage ultérieur pour autant? La question reste posée, bien que l'on doive également tenir compte de leur manque d'appétence dans les technologies numériques et dans une approche par données. De plus, ce n'est pas parce que des utilisateurs finaux sont associés au processus de conception d'un artefact technologique qu'ils en feront nécessairement usage ou qu'ils l'intégreront dans leurs pratiques professionnelles.

En sa qualité de porte-parole de la rédaction, la coordinatrice du mensuel a très vite témoigné un intérêt pour l'expérience. Ses usages sont de deux types. Le premier consiste en un usage utilitaire et fonctionnel de l'outil, le consultant à différents moments liés au lancement du projet, à la préparation de réunions de rédaction ou à une actualité spécifique relative à un pic de pollution ou à des problèmes de trafic automobile à Bruxelles. Le second est celui d'un usage d'adoption symbolique de l'objet, qui se réfère à l'acceptation de l'idée de l'innovation (Klonglan & Coward 1970). Elle fera également usage de la métaphore du robot dans un sens symbolique, faisant écho au "syndrome du robot hollywoodien" (Waddell 2018), en vertu duquel les robots sont des artefacts sympathiques et amicaux. Elle présente le dispositif comme une nouvelle recrue au sein de la rédaction, les tâches qui lui sont assignées sont définies et délimitées, ses potentialités en ce qui concerne la collecte et le traitement de données automatisés sont soulignées, mais il ne lui est pas reconnu un caractère "humain". Ceci apparaît comme paradoxal, compte tenu de la part humaine importante dans ce projet. D'ailleurs, écrira-t-elle, les qualités du dispositif ne sont pas à trouver dans la rédaction de textes. Cela

sous-entend une forme d'anthropomorphisme en vertu de laquelle seul un journaliste humain serait capable de rédiger un récit agréable à lire (Waddell 2018). Le "robot" est donc un objet qui devient un outil dès lors qu'il prend en charge des tâches ingrates. La manière dont elle aborde la métaphore du robot peut également constituer une forme de dédramatisation du phénomène, dans le contexte plus général d'une "quatrième révolution industrielle" dont les conséquences seraient celles d'une perte massive d'emplois humains, qu'ils soient manuels ou intellectuels.

Pour la journaliste indépendante impliquée dans cette expérience, l'avantage relatif du dispositif est bien celui de la prise en charge de tâches répétitives, difficiles à envisager de manière manuelle, qui ont donné lieu à des informations qu'elle n'aurait pas pu obtenir autrement, en raison de la nature chronophage que cela implique. S'il ne s'agissait pas d'un besoin qui existait préalablement, c'est l'outil qui l'a créé, selon elle. Tout comme la coordinatrice de la rédaction d'*Alter Échos*, son apport à la construction sociotechnique de l'outil est celui d'une extension de ses fonctionnalités. Toutefois, celui-ci est apparu comme limité dès lors qu'elle s'est retrouvée confrontée à la mise en récit des données : certains aspects n'avaient pas été prévus lors de la conception de l'artefact et nous ont donc engagée dans un travail supplémentaire, qui n'était pas automatisé, pour répondre à de nouvelles demandes spécifiques. Bien que cet appui humain n'ait pas été directement perçu – "*c'est à ça qu'il sert*", justifiera-t-elle, à propos du dispositif – cette journaliste soulignera l'importance d'une médiation humaine moins liée à notre posture de chercheuse qu'à celle de notre identité professionnelle antérieure. C'est d'ailleurs sur ce dernier plan que la journaliste indépendante attribuera la légitimité de l'expérience : si l'agent de médiation sociotechnique n'avait pas été identifié comme relevant du monde du journalisme, l'expérience n'aurait pas été perçue comme admissible ou valide. Ceci témoigne également de la difficulté de se projeter dans des usages finaux lorsqu'un artefact a pour objet de soutenir une démarche journalistique s'inscrivant dans le temps long de l'enquête.

Cette journaliste jouera également avec la métaphore du robot pour conférer un dispositif un caractère attachant – "*Quelle efficacité ce petit nono*"⁶⁵ – qui va donner lieu une forme de familiarité, dans un dialogue entre la journaliste et sa source – "*Robot, peux-tu comparer la moyenne des concentrations de dioxyde d'azote?*". Son usage de la métaphore va également donner une image amusante du dispositif, dans l'article qu'elle signe à l'issue de l'expérience – ("*Le Bxl'air bot ne prend pas beaucoup de place et ne sert pas le café*"), avant de se l'approprier en cours d'article ("*Notre petit robot*"). Malgré une prise de distance dans chacun de ses articles, où elle répète le caractère non sollicité du dispositif, elle en assume la part de subjectivité qui résulte de ses propres choix journalistiques. En ce sens, on ne peut pas parler d'un usage d'appropriation – l'acte de se constituer un soi (Jouët 2000) –, mais bien d'une intégration des productions automatisées dans ses pratiques professionnelles, étant donné une mise à distance (volontaire) entre la journaliste et l'outil. La complémentarité entre l'humain et la machine s'est donc jouée à la fois sur le plan des tâches prises en charge, et sur celui du résultat dès lors que l'information allait être mise en sens par son activité journalistique. L'étape de stabilisation décrite dans

⁶⁵ Communication personnelle, via courriel, le 25/04/2018.

le modèle SCOT sera réalisée en ce qui la concerne, dans la mesure où l'artefact est devenu un instrument de connaissance.

Tout au long de cette expérience, les imaginaires des journalistes ont été placés sous le signe de la dualité homme-machine, et ceux-ci ont largement contribué à forger tant les usages que les différentes formes de non-usages (désintérêt, non-participation, non-utilisation, non-adoption, non-appropriation). Les non-usages peuvent être interprétés comme autant de formes de résistance face à un dispositif qu'ils n'ont pas choisi ou sollicité (Boudokhane, 2006). Celles-ci ne sont pas toujours apparues manière explicite : par exemple, la non-participation aux enquêtes par formulaire peut être interprété comme une forme de résistance.

Vu sous le prisme d'une sociologie de l'innovation, relative à "*une idée, pratique ou objet que les gens voient comme différents*" (Zaltman & Wallendorf 1983, cités par Ram 1987), le phénomène de résistance peut être compris comme un refus du changement, que ce soit sur le plan des pratiques professionnelles ou sur celui de l'idéologie du journalisme. Aussi, la résistance est-elle ici intrinsèquement liée à la dimension représentative de l'artefact. Les imaginaires technologiques de ces quatre journalistes sont plutôt rétifs. Ils sont liés à des technologies de l'information et de la communication souvent perçues comme contraignantes et chronophages. Les journalistes ne rejettent pas les technologies en bloc, mais ils estiment qu'elles les concernent moins dans leur approche du journalisme, qu'ils qualifient eux-mêmes de "*traditionnelle*", voire "*old school*", alors que la moyenne d'âge de la rédaction est de moins de quarante ans. Cette appréhension peut s'expliquer par l'anxiété ou la peur que la technologie provoque chez l'individu (Selwyn 2003). La non-participation de ces quatre journalistes peut donc être interprétée comme une résistance dirigée contre une remise en question de pratiques privilégiant le terrain ou contre l'idée même de l'innovation, dont la symbolique menace l'emploi journalistique ("*La méga menace du robot qui piquera ma place est toujours bien présente*"⁶⁶). Par ailleurs, une résistance ne signifie pas nécessairement une non-adoption de l'innovation : la résistance et l'adoption peuvent co-exister tout au long du processus d'innovation et l'adoption peut se produire après avoir fait tomber les résistances (Ram 1987).

Sur le plan des pratiques professionnelles, les non-usages peuvent également être mis en relation avec une appréhension du journalisme de données qui ne soulève pas un enthousiasme particulier. En dehors d'un journaliste qui est amené à traiter régulièrement de sujets socio-économiques, impliquant le traitement de données chiffrées, les journalistes de la rédaction ne témoignent d'aucun attrait particulier pour une approche par données. Deux journalistes au moins diront ne pas être à l'aise avec les chiffres et les mathématiques... et que cela leur fait peur. Il s'agit d'une tendance souvent observée dans le monde du journalisme, en ce compris parmi les étudiants en journalisme (Schmitz-Weiss & Retis-Riva 2018). De plus, le dispositif d'automatisation a pu les renvoyer face à leurs propres compétences en matière de traitement de données, peu développées par manque d'appétence ou de temps pour les acquérir. S'ajoute à ces variables un manque d'intérêt pour le domaine d'application, une absence de besoin, un

⁶⁶ Évaluation par questionnaire, juillet 2017.

manque de motivation, voire une absence de sens (Kellner *et al.* 2010). Toutefois, les imaginaires de la rédaction ont évolué dans le temps, témoignant de leur instabilité (Musso 2009). L'ensemble des journalistes dit avoir finalement perçu les bénéfices du dispositif et être favorable pour renouveler une telle expérience, à condition que le domaine d'application soit plus en phase avec des intérêts personnels ou professionnels. Les journalistes ont également admis que l'artefact avait utilement nourri le journalisme, lui reconnaissant l'avantage relatif de constituer une nouvelle source d'information. La question de la compatibilité avec les valeurs journalistiques n'a pu être résolue qu'à partir du moment où une journaliste a mis les productions de l'artefact en récit et donc en sens. Le dispositif d'automatisation n'est pas considéré autrement que pour ce qu'il est : un outil au service d'un travail journalistique. Il ne peut être considéré comme une forme de journalisme à part entière, dès lors qu'il ne peut remplacer un travail journalistique de contextualisation et d'analyse. Cette vision partagée d'une complémentarité homme-machine est ce en quoi, pour l'ensemble des journalistes de la rédaction, le bilan de cette expérience peut être considéré comme positif.

4 | L'artefact "Quotebot" dans les usages de la rédaction de *L'Echo*

"Quotebot" est un projet initié par le responsable du développement et de la transformation numérique de *L'Echo*, quotidien belge spécialisé dans l'information économique et financière. Il a pour objet de soutenir les routines des journalistes attachés au service "Investir" du journal, dans le contexte de la couverture en temps réel de l'information boursière. Ses productions peuvent prendre une forme textuelle ou visuelle (graphiques ou tableaux), en fonction du type d'information traité. De Bruxelles à New York, sa zone de couverture est celle des principales places boursières mondiales. Le projet a bénéficié d'un financement du Google Digital News Initiative, dont l'objet est de soutenir l'innovation dans les médias européens, pour un budget total 211.099 euros (dont 147.000 euros sont financés par Google DNI). Son développement a été assuré par la start-up parisienne Syllabs, spécialisée dans le développement de technologies sémantiques, pour un montant s'élevant à environ 70% du budget. Son moteur de rédaction a fait ses premiers pas dans la presse en 2015, à l'occasion d'un partenariat avec *Le Monde* dans le cadre de la couverture des résultats du premier tour des élections régionales françaises¹. Les données nourrissant le système d'information sont fournies par le groupe VWD – partenaire historique du groupe Mediafin, propriétaire de *L'Echo*, qui détient également le quotidien néerlandophone *De Tijd* –, spécialisé dans la commercialisation de données financières.

Le soutien à des routines professionnelles constitue l'un des deux versants du projet "Quotebot", dont les productions automatisées sont aussi destinées à enrichir un service existant proposés par les sites web des deux quotidiens de Mediafin, dans la perspective d'une information servicielle : les portefeuilles virtuels d'actions, accessibles aux abonnés des quotidiens. Sortant du propos de cette recherche, ce volet n'y a pas été traité, bien qu'il illustre l'une des possibilités de déclinaison d'un projet d'automatisation au sein d'un média d'information, qui est celle de l'automatisation de contenus personnalisés. Par ailleurs, s'il est développé en langue française, le projet consiste également à proposer une version traduite de "Quotebot", pour soutenir les routines quotidiennes des journalistes du service boursier du *Tijd*. À l'origine, il était également prévu que "Quotebot" intègre un système d'alerte destiné à prévenir les journalistes de mouvements anormaux sur les marchés. Pour des raisons de planning, il a été abandonné en cours de route. Bien que chacune des étapes du projet ait été planifiée – c'est d'ailleurs une contrainte imposée par Google DNI –, les délais du développement de "Quotebot" n'ont pas

¹ Voir supra p.104

pu être respectés. Devant initialement s'échelonner sur une période d'un an, au moment de clôturer cette thèse, le projet accusait un retard important. Nous verrons quels sont les (nombreux) facteurs qui n'ont pas permis de tenir les délais prévus, lesquels permettent d'éclairer les difficultés liées au déploiement d'une innovation technologique au sein d'une rédaction.

Considérant qu'il s'agit d'un artefact piloté par des données dont la qualité est assurée par leur fournisseur, la problématique de la qualité des données paraissait moins importante que dans l'étude de cas précédente. Si le cadre théorique et la méthode de collecte des données empiriques sont comparables à cette dernière, il a été ici supposé que l'artefact d'automatisation serait utilisé à condition de rencontrer les exigences de sa conception, lesquelles furent posées par les journalistes eux-mêmes. La nature d'un outil destiné à soutenir des pratiques quotidiennes, via des productions automatisées s'intégrant dans le système de gestion de contenus habituellement utilisé par les journalistes, nous a amenée à envisager la problématique de l'usage sous le prisme de l'expérience utilisateur, qui peut être définie comme "*le degré d'émotions positives ou négatives que peut ressentir un utilisateur spécifique dans un contexte spécifique pendant et après l'utilisation du produit et qui motive son utilisation ultérieure*" (Schulze & Krömker 2010). Aussi, les évaluations des productions de "Quotebot" se sont appuyées sur des questionnaires conjuguant des méthodes issues du domaine de la linguistique computationnelle et des Interactions Homme-Machine (HCI, *Human-Computer Interaction*²). Dès lors qu'il s'agissait d'un projet de plus grande envergure où les parties prenantes étaient issues du monde du journalisme, de celui de l'informatique et de l'ingénierie linguistique, la médiation sociotechnique est apparue comme un maillon essentiel dans le processus de développement du projet. Son analyse s'est appuyée sur les sept principes de la norme ISO 9000, relative au management de la qualité³, de manière à s'inscrire dans la perspective d'une adéquation des productions automatisées à des usages définis par leurs utilisateurs finaux.

La collecte des données empiriques comprend :

- des méthodes d'enquête par questionnaire, destinées aux six journaliste du service "Investir" (trois formulaires accessibles en ligne, dont deux de manière anonyme) ;
- huit entretiens semi-dirigés (dont quatre avec le responsable du développement et de la transformation numérique de *LEcho*, qui pilote le projet, un avec l'un des journalistes ayant activement participé au projet, et trois avec des représentants de Syllabs) ;
- des observations participantes organisées dans un cadre formel (participation à six réunions de travail avec les journalistes du service "Investir", aux deux réunions de lancement du projet avec les responsables des deux quotidiens de Mediafin, et six appels passés en ligne entre des représentants de Mediafin et de Syllabs) ;
- une demi-journée d'observation au sein du service "Investir" de *LEcho* ;
- un accès à l'ensemble des échanges entre Mediafin et le prestataire de services (129 échanges de courriels), ainsi qu'à 62 documents de travail hébergés sur une plateforme collaborative en ligne.

² Voir "Modes d'évaluation", p.231

³ Voir "Médiation sociotechnique et norme ISO 9000", p.233

Cet accès privilégié a été permis en raison de notre activité de consultance dans ce projet, qui fut sollicitée par le responsable du développement et de la transformation numérique de *L'Echo* pour accompagner les journalistes du service "Investir" dans le processus de conception du moteur de rédaction. En vue de soutenir leurs routines quotidiennes de la manière la plus efficace possible, ceux-ci ont été invités à s'impliquer activement dans ce projet, que ce soit en définissant eux-mêmes leurs besoins ou en prenant en charge la rédaction, puis les corrections, des gabarits de texte devant servir de base à une automatisation. Au-delà de la mobilisation des compétences en informatique et en linguistique inhérente au développement d'un artefact d'automatisation de la production d'informations, l'expertise du domaine d'application est apparue comme un préalable, à l'instar de n'importe quel autre type de projet de traitement automatique de la langue (Reiter & Dale 1997). Cette expertise, doublée d'une expertise éditoriale, relève du domaine des journalistes; tandis que celle du développement du moteur de rédaction relève de celle du prestataire de services. La fiche de présentation du projet, accessible depuis le site de Google DNI, le résume d'ailleurs comme suit : "Editorial led, robot delivered" ("Piloté par l'éditorial, délivré par un robot")⁴.

4.1 Contexte de l'expérience

Si le traitement de l'information boursière fait partie de l'ADN du site web de *L'Echo*, celui-ci consistait en des rendez-vous quotidiens jusqu'en 2012, époque à laquelle fut lancé le "Market Live", pour assurer la couverture, en temps réel, des principaux marchés boursiers mondiaux. Il s'agit d'un format journalistique relevant du "live blogging", caractérisé par des textes courts qui n'excèdent généralement pas une dizaine de lignes. Cette couverture mobilise deux journalistes du service "Investir" – qui en compte six –, lesquels se succèdent en deux *shifts*, de 7 heures du matin à 15 heures et de 15 à 19 heures. Elle est ensuite assurée par les journalistes du service web jusqu'à 23 heures, soit une heure après la clôture de Wall Street (bourse de New York). Pour les journalistes attachés au "Market Live", cela se traduit par un travail routinier et intense, qui leur laisse peu de répit.

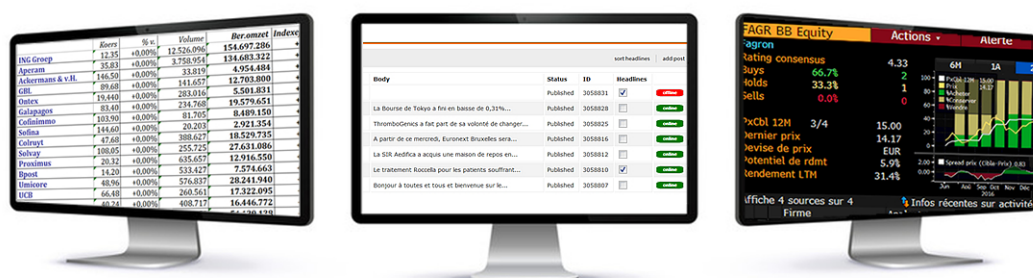


FIGURE 4.1 – Postes de travail utilisés par un journaliste du service "Investir", de gauche à droite : un tableur présentant des valeurs boursières évoluant en temps réel, le système de gestion de contenus interne et les informations de l'agence Bloomberg

⁴ Source : "Editorial-led, robot-delivered", Google DNI, consulté le 07/07/2018, URL : <https://newsinitiative.withgoogle.com/dnifund/insights/quotebot-editorial-led-robot-delivered-insight/>

Les journalistes jonglent avec trois écrans d'ordinateur (Figure 4.1) – celui d'un poste de travail pour la rédaction des contenus, celui d'un poste de travail permettant de suivre l'évolution des cours du marché via des tableurs connectés à un système interne de diffusion des données, et celui d'un poste de travail dédié aux informations de l'agence de presse Bloomberg, spécialisée en informations financières – pour rendre compte des principaux mouvements enregistrés sur les marchés boursiers. Lors d'une journée d'observation *in situ*, un journaliste confiera : "*Il est difficile de s'arrêter pour prendre le temps de lire correctement l'analyse du marché. Il est délicat de prendre une demi-heure de pause à midi, car il faut répercuter l'info*"⁵. Dans ce contexte, le propos de "Quotebot" est de fournir aux journalistes des brouillons automatisés devant nécessairement faire l'objet d'une validation humaine avant leur publication : à leur charge de les publier tels quels, de les enrichir d'informations de contexte ou de ne pas les publier du tout. "Quotebot" est envisagé comme un complément à un arsenal existant, qui inclut notamment des dépêches d'agences de presse, des e-mails provenant d'acteurs du secteur économique, et une activité de veille de comptes spécialisés via le réseau social Twitter (qui peuvent également jouer un rôle d'alerte). En amont de la conception du projet, le gain de temps a été estimé à une demi-heure par jour et par journaliste, ce qui n'est pas négligeable compte tenu d'un travail s'organisant en flux tendu. Ce temps pourra être mis à profit pour les journalistes se concentrer sur des activités d'analyse ou des tâches à plus forte valeur ajoutée : "*Quand on suit l'actualité boursière en temps réel, il faut être constamment aux aguets (...) S'il prend en charge certaines tâches, je ne devrai donc pas m'en occuper et gagnerai en productivité. Cela va permettre d'offrir aux lecteurs un contenu plus riche et plus intéressant*", indique un journaliste⁶.

Bien qu'il soit caractérisé par des données chiffrées, le domaine d'application ne peut être abordé à travers le seul prisme de valeurs fluctuant sur les marchés boursiers. Ces derniers sont également pilotés par un sentiment de confiance, irrationnel et diffus, qui ne peut être quantifié ou déduit de manière automatique. Aussi, l'explication des phénomènes observés n'est pas toujours à trouver dans le seul fait économique. "*Une actualité peut influencer le cours des bourses (...) Actuellement, la politique américaine rend les marchés volatiles. Il y a beaucoup d'incertitudes et ils n'aiment pas ça*", souligne un journaliste⁷. Par ailleurs, un communiqué peut informer un journaliste du dépassement du seuil d'actionnaires dans une société donnée : "*C'est le genre d'information qui n'a pas beaucoup de valeur, mais qui témoigne de la sensibilité du marché*"⁸. Cette sensibilité peut également s'exprimer en réaction à des faits d'actualité. Par exemple, la crise terroriste du début des années 2000 aux États-Unis avait entraîné des "*effets psychologiques comme les mouvements de panique et de défiance vis-à-vis des marchés bancaires et financiers (...) l'inquiétude des Américains et du reste du monde au sujet d'une multiplicité des attaques terroristes sur d'autres cibles a eu des effets négatifs spécialement sur les secteurs du transport aérien et du tourisme*" (Bensafta & Gervasio 2011). C'est aussi pourquoi un article publié sur le "Market Live" est susceptible d'aller au-delà d'une présentation factuelle, en proposant une mise en contexte de l'information. Il faut également tenir compte du fait que

⁵ Communication personnelle, Bruxelles, le 27/06/2018.

⁶ Communication personnelle, par téléphone, le 10/09/2019.

⁷ Communication personnelle, Bruxelles, le 27/06/2018.

⁸ Communication personnelle, Bruxelles, le 27/06/2018.

le domaine du journalisme boursier couvre une variété de secteurs économiques (biotechnologies, immobilier, ...). Dans la rédaction, cela implique un partage d'expertise entre journalistes spécialisés : *"On doit (régulièrement) échanger car on ne sait pas tout faire tous seuls"*, souligne un journaliste⁹.

Pour le responsable du développement et de la transformation numérique de *L'Echo*, l'originalité du projet "Quotebot" est celui de constituer *"une solution éditoriale, pas quelque chose qui a été imposé à la salle de presse par les informaticiens. Le journalisme de qualité est la raison pour laquelle les investisseurs souscrivent à Mediafin. 'Quotebot' nous aide à offrir cela mieux, plus rapidement et à moindre coût"*, indique-t-il sur la page web que Google DNI a dédié au projet¹⁰. À l'époque de la constitution du dossier visant à répondre à l'appel lancé par Google DNI, c'est un partenariat homme-machine potentiellement fructueux qu'il vise, le challenge consistant à réussir à mobiliser les journalistes et à les fédérer autour de ce projet. *"Cette automatisation a été pensée de différentes manières : d'une part, car je pense que cela va être un phénomène majeur et que c'est un premier pas pour se dire que la machine a peut-être quelque chose à nous offrir. C'est aussi un challenge journalistique, qui correspond à un segment de notre audience intéressé par les résultats financiers. Après, aux journalistes de voir s'ils veulent utiliser le texte tel quel ou en tant que source"*¹¹.

S'il est destiné, à terme, à deux services d'information boursière distincts, celui de *L'Echo* et celui du *Tijd*, le projet "Quotebot" a été conçu en français, à partir d'un cahier des charges établi en collaboration avec les journalistes francophones et le responsable du développement et de la transformation numérique de *L'Echo*. La rédaction du quotidien néerlandophone n'a donc pas été activement impliquée dans la conception du moteur de rédaction, qui fut donc d'abord développé en langue française avant de faire l'objet d'une traduction. Ce parti pris se justifie essentiellement pour des raisons de moyens : *"On n'avait ni les moyens ni le temps pour réaliser un projet avec des spécificités différentes, en fonction de la langue. Par contre, il pouvait y avoir des petits changements. Par exemple, les journalistes néerlandophones souhaitent des titres beaucoup plus factuels et moins éditorialisés"*, commente le responsable du développement et de la transition numérique de *L'Echo*¹². "Quotebot" a été présenté à l'équipe managériale du *Tijd* le 22 février 2019, soit près d'un an après le lancement des travaux relatifs à sa conception. Sa réaction peut être qualifiée de très positive : *Très chouette et intéressant*, *"Super, yes"*, *"Ça facilite la vie pour ajouter de la valeur journalistique (...) ça prend en charge le côté 'dom' (bête, ndlr) du job"*. Des inquiétudes ont toutefois été exprimées à propos de la correction de la langue (*"Les textes doivent être bien construits, en néerlandais courant"*), posant ici la question de la satisfaction des journalistes de *L'Echo* quant à la qualité des textes générés en langue française. Nous verrons que cette problématique n'est pas qu'une question du traitement des données ou de la langue : elle relève, avant tout, de l'expertise du domaine d'application.

⁹ Communication personnelle, Bruxelles, le 27/06/2018.

¹⁰ Source : "Editorial-led, robot-delivered", Google DNI, consulté le 07/07/2018, URL : <https://newsinitiative.withgoogle.com/dnifund/insights/quotebot-editorial-led-robot-delivered-insight/>

¹¹ Communication personnelle, Bruxelles, le 14/11/2017.

¹² Communication personnelle, Bruxelles, le 21/03/2019.

4.1.1 Chaîne sociotechnique et cadre technologique

La chaîne sociotechnique qui se déploie autour de "Quotebot" compte quatre groupes d'acteurs (Figure 4.2) : les rédactions néerlandophone et francophone des quotidiens ainsi que les responsables du groupe Mediafin; la société VWD, qui commercialise les données boursières alimentant le moteur de rédaction; la start-up Syllabs en charge du développement opérationnel de l'artefact; et les sociétés externes au groupe de presse qui prennent en charge la sous-traitance d'une partie de ses activités IT. Ces dernières ne font pas l'objet d'un traitement spécifique dans cette thèse, dans la mesure où leur rôle est considéré comme secondaire, leur implication dans le projet consistant à implémenter les flux d'informations transmis par Syllabs dans le système de gestion de contenus habituellement utilisé par les journalistes.

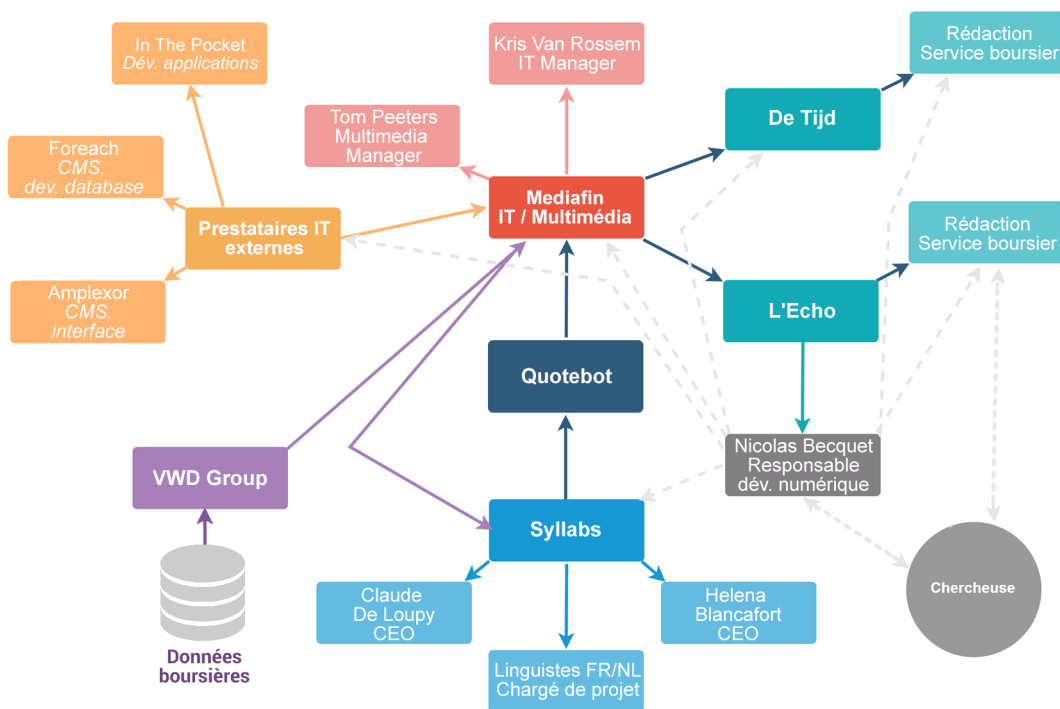


FIGURE 4.2 – Schéma conceptuel de la chaîne sociotechnique formée autour de "Quotebot"

La diversité de parties prenantes a supposé la mise en place et la gestion d'un planning commun, qui ne fut pas toujours facilité en raison d'échelles de priorités pouvant être différemment perçue par les prestataires de services IT externes à Mediafin. Pour ceux-là, "Quotebot" s'inscrivait dans le cadre de la gestion d'un ensemble d'autres projets. Cela explique, en partie, le retard dans la mise en œuvre de celui-ci. Nous verrons qu'une accumulation d'autres facteurs, organisationnels et opérationnels, ont également affecté la planification du projet, lequel devait initialement s'échelonner sur un an¹³.

¹³ Voir "Construction sociale de l'artefact technologique", p.235

4.1.1.1 Le journal *L'Echo*

L'Echo est né à la fin du dix-neuvième siècle, époque à laquelle règne un certain optimisme industriel. Son premier numéro est publié le 22 mai 1881 (Figure 4.3), à l'initiative de trois agents de change, Alfred Van Der Elst, André Mélot et Hyacinthe Peemans. Il est d'abord publié deux fois puis trois fois par semaine. Né *L'Echo de la Bourse de Bruxelles*, la mention "Journal financier, industriel et commercial" figure dans sa manchette. Il devient *L'Echo de la Bourse* en 1890. La même année, il adopte un rythme de publication quotidien (De Bens 1973, 2001). Le journal va connaître une période difficile au cours de la Première Guerre mondiale, puisque la Bourse de Bruxelles a gardé ses portes closes jusqu'en 1919. Le journal a également cessé de paraître au cours de la Seconde Guerre mondiale, également en raison de la fermeture temporaire de la Bourse de Bruxelles (De Bens 1973).



FIGURE 4.3 – Premier numéro de *L'Echo de la Bourse de Bruxelles*, le 22/05/1881

À l'aube des années 1960, le journal est édité par la société anonyme Publica, à l'imprimerie Imifi (Imprimerie industrielle et financière), juridiquement autonome et disposant d'une communauté d'actionnaires. Son tirage dépasse légèrement les 15.000 exemplaires, ce qui représente plus de la moitié du tirage de la presse financière belge, qui compte alors quatre titres. Les ressources de ces journaux dépendent à la fois du marché publicitaire et des relations publiques des sociétés, dont principalement la Société Générale. De manière à élargir son audience et à s'ouvrir à davantage de ressources en matière de publicité commerciale, *L'Echo de la Bourse* a développé des rubriques d'intérêt général (cinéma, théâtre, politique intérieure et internationale, etc.), "mais le résultat est relativement mince en ce qui concerne l'ouverture du public" (CRISP 1959). Aussi, malgré ces efforts, le tirage du journal atteint moins de 30% de celui enregistré avant 1930.

Dans les années 1960, la politique d'expansion de l'imprimerie Imifi, nécessitant de lourds investissements, n'est pas sans conséquences sur la santé financière du journal. En 1967, le groupe d'assurances Josi prend une participation de 58,7% au capital du journal. Cette solution est trouvée à l'initiative du Premier ministre Paul Vanden Boyenants, qui voit dans *L'Echo de la Bourse* "une source d'informations mais aussi (...) un soutien au développement économique et défendant le libre marché" (De Bens 2001 :145).

La fin des années 1970 est marquée par l'arrivée de nouveaux actionnaires. À l'été 1977, la société Editions Echos de la Bourse prend le relais de la SA Publica. Le groupe Josi en reste actionnaire, mais sa participation descend à 17%. Ses autres actionnaires sont notamment la Royale Belge (société d'assurances, 5%), la Sofina (groupe d'assurances AG, 10%), la Société d'Édition et de Publicité (fondée par Maurice Brébart, editrice du quotidien *La Dernière Heure*), Electrabel (5%), la Fédération des entreprises de Belgique (FEB, 7,5%), la Générale de Banque (2%), le Crédit Communal (sur les cendres de laquelle est née la banque Belfius, 1,5%) et la Banque Bruxelles Lambert (intégrée dans le groupe ING en 1998, 1%). L'année suivante, des collaborations sont lancées avec la société de Maurice Brébart, tant sur les plans administratifs que commerciaux. Celles-ci s'intensifient lorsque le journal est imprimé par la Sodimeco (Société d'imprimerie commune), laquelle imprime également les quotidiens *La Dernière Heure* et *La Libre Belgique*. La situation financière fragile d'Agefi (Agence économique et financière de Belgique, née en 1910) aboutira à sa fusion avec *L'Echo de la Bourse* en 1980, cédée pour un franc belge symbolique par la Société Générale. La même année, *Le Courrier de la Bourse*, un autre titre historique de la presse économique belge, fondé en 1885, entre dans le giron de la société editrice du journal, via le rachat du groupe Sobeli (De Bens 2001 :416).

Un nouveau mouvement de capital survient en 1989, année au cours de laquelle le groupe français Eurexpansion de l'homme d'affaires Jean-Louis Servan-Schreiber entre, à hauteur de 40%, dans le capital de la société editrice de *L'Echo* (devenue Editeco). L'ambition de Servan-Schreiber est de se positionner en tant que groupe de presse européen spécialisé dans l'information financière – il détient également les titres allemand *Handelsblatte*, italien *Class Editori*, et américain *Dow Jones*. Toutefois, il vendra la moitié de ses parts en 1994. Celles-ci seront alors partagées entre le groupe Tractebel, le Groupe Bruxelles Lambert (via l'homme d'affaires Albert Frère), la Société Mosane, la Sofina, le groupe IPM (éditeur des quotidiens *La Libre Belgique* et de *La Dernière Heure*), et le groupe Josi (De Bens 2001 :417). Les 20% restant sont alors convoités par la société editrice du quotidien économique et financier *De Financieel Economische Tijd* mais, pour éviter la situation de blocage que cette participation pourrait entraîner, ce sont deux autres groupes de presse francophone qui font leur entrée dans le capital de la société : Rossel, à hauteur de 14,5%, et Vers L'Avenir, à celle de 4,5% (De Bens 2001 :418).

D'une édition du soir, le quotidien passe à une édition de l'après-midi en 1987. En vue de mieux définir sa marque, qui n'est pas seulement celle d'un journal économique et financier dès lors qu'il traite aussi de politique belge et internationale, le titre adopte un nouveau logo et devient *L'Echo*. Fin des années 1990, son tirage s'élève à 34.476 exemplaires (contre 30.000 en 1970). Ses recettes proviennent à la fois de la publicité (66,4%) et de ses ventes (33,6%). Il compte 58% d'abonnés et emploie une quarantaine de journalistes à temps plein (De Bens 2001 :419). "*Politiquement neutre, il se considère cependant comme un journal d'opinion favorable à l'économie de marché et à la liberté d'entreprise*" (Dumont et al. 1998 :71).

Le site internet de *L'Echo*, d'abord accessible via l'adresse echonet.be, est lancé en 1996. S'adressant aux managers et investisseurs, il propose plusieurs services tels que l'accès à des bases

de données et des caleuses de change (Dumont *et al.* 1998 :79). À l'époque, il est considéré comme un " *site de référence, prouvant que l'information économique et financière permet une rentabilité à court terme*". Il est accessible gratuitement mais est " *réservé aux utilisateurs enregistrés*" (Dumont *et al.* 1998 :79). Conçu comme une déclinaison du journal papier tirant parti de la valeur ajoutée que permet un site d'informations en ligne, il proposera, très rapidement, des rendez-vous ponctuels en matière d'information boursière, sous l'onglet "Les marchés". Au fil du temps et des versions du site ¹⁴, il se présentera sous la forme d'un espace spécifique, dédié au "live-blogging", où figure la mention "Live" pour souligner qu'il s'agit là d'une couverture de l'information en temps réel. Nous n'avons pas pu déterminer le moment exact du moment du lancement de cette section du site mais, d'après une source interne, cela remonterait à 2012. Le compte Twitter du "Market Live" a été créé en 2016, s'inscrivant dans cette même logique de flux d'information en continu. Ses mises à jour s'opèrent entre 8 heures du matin et 18 heures, à un rythme moyen de 151 tweets mensuels.

En 2003, la société éditrice de *L'Echo*, Editeco, est rachetée par le groupe francophone Rossel, éditeur du quotidien *Le Soir* et des journaux de l'actuel groupe Sud Presse, ainsi que par le groupe de presse néerlandophone De Persgroep, qui édite notamment les quotidiens *Het Laatste Nieuws* et *De Standaard* ¹⁵. En 2005, les éditeurs du *Tijd* et de *L'Echo*, Uitgeversbedrijf Tijd et Editeco, s'associent au sein de la joint-venture Mediafin. Chacun d'entre eux en détient 50%. Le 20 juin de l'année suivante, une nouvelle maquette du journal est présentée, caractérisée par un passage au format belge et à la quadrichromie ¹⁶. La même année, tous les départements de *L'Echo* et du *Tijd* fusionnent, à l'exception des rédactions qui conservent leur indépendance. " *Un changement actionnarial qui ne modifiera en rien la ligne éditoriale du journal, à savoir une information indépendante de qualité destinée à ceux qui entreprennent tant professionnellement que personnellement*", souligne Joan Condijs, alors rédacteur en chef de *L'Echo* ¹⁷. En 2008, les deux quotidiens commencent à partager des projets rédactionnels communs, à travers le lancement des suppléments *Mon Argent* et *Netto*, consacrés à l'économie à la finance, et *Sabato*, un hebdomadaire dédié aux loisirs de luxe. Les pages des deux quotidiens adoptent la couleur saumon en 2009, de manière à affirmer leur attachement au monde du business : dans le monde entier, il s'agit de la couleur des grands journaux financiers. Un nouveau site web est lancé en 2010, dont l'accès illimité est réservé aux abonnés payants. En 2013, ceux-ci représentent 25% du total des lecteurs. En 2010 également, les deux quotidiens lancent leur application pour smartphones et iPad. Deux ans plus tard, le format belge (520 x 365 mm) sera abandonné au profit du format berlinois (470 x 320 mm), plus compact (Vanderhaeghe 2011).

¹⁴ Voir annexe 17, "Évolution de l'interface du site web de *L'Echo*, de 2002 à 2019", p.350

¹⁵ "'L'Echo' entre dans la famille Rossel", Philippe Servaty, *Le Soir*, 28/08/2003, consulté le 20/05/2019, URL :

¹⁶ Sources : "Les mues de *L'Echo*", Laurence Dierickx, *Journalistes*, revue éditée par l'Association des journalistes professionnels, n°72, juin 2006; et "Historique – Mediafin", Mediafin, consulté le 20/05/2019, URL : <https://www.mediafin.be/fr/historique/>

¹⁷ "Indépendance et qualité", Joan Condijs, *L'Echo*, 03/10/2017, consulté le 20/05/2019, URL : <https://www.lecho.be/opinions/edito/independance-et-qualite/9938472.html>

Tout au long des années 2000, *L'Echo* a vu diminuer la diffusion de son produit imprimé de manière constante, une tendance observée dans l'ensemble des quotidiens belges francophones. Celle-ci est ainsi passée de 28.131 exemplaires vendus quotidiennement en 2000 à 16.847 exemplaires en 2010¹⁸, et elle fut seulement compensée par les abonnements numériques à partir de 2010 (Figure 4.4). En 2012, alors que l'érosion de ses ventes se poursuivait (-11,53%), le journal enregistrait une augmentation de 1.548 ventes supplémentaires par jour pour sa version digitale¹⁹. Cinq ans plus tard, la reprise est amorcée. Dans le même temps, les audiences du journal progressent de manière significative, avec un gain de 28,60% en 2018, par rapport à 2016-2017. *L'Echo* est le seul quotidien francophone à enregistrer une aussi nette progression avec une moyenne quotidienne de 84.500 lecteurs. "Ces chiffres sont en ligne avec la solide croissance des ventes du titre au cours de ces derniers mois. Entre janvier et août 2018, elles ont augmenté de 9,3% à 15.101 exemplaires quotidiens, portées par le digital (... qui ...) représente près de la moitié des ventes du titre", indique *L'Echo*²⁰.

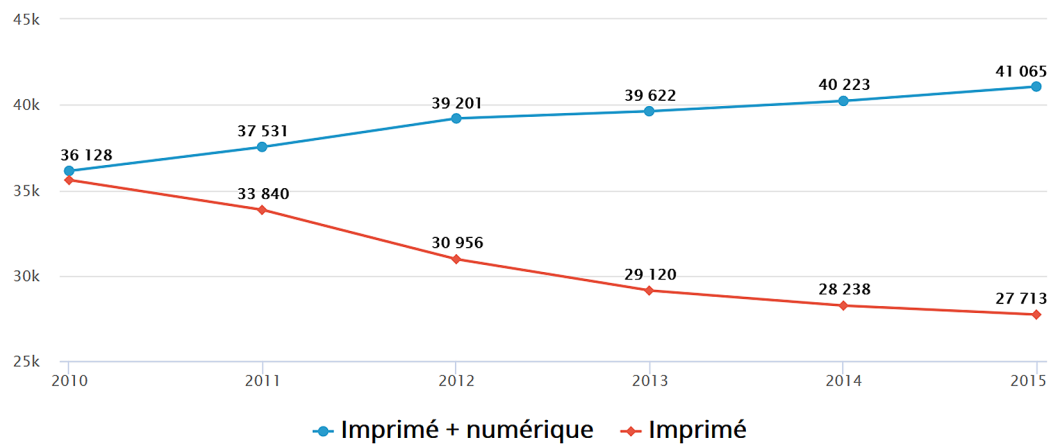


FIGURE 4.4 – Évolution de la diffusion payante de *L'Echo*, de 2010 à 2015 (Centre d'information sur les médias, CIM)

Le début des années 2010 est marqué par la diversification des activités de Mediafin, qui acquiert Content Republic, fournisseur de services sur mesure pour les pouvoirs publics et les entreprises dans les services business et financiers, Comfi, un bureau de consultance spécialisé dans la communication commerciale et financière, ainsi que Bepublic & Bereal, deux agences de relations publiques spécialisées dans le journalisme d'entreprise et l'immobilier. En 2017, De Persgroep vend sa participation de 50% à Roularta Media Group, un groupe de presse mul-

¹⁸ Sources : "Echtverklaring resultaten", CIM, consulté le 22/05/2019, URL : <https://www.cim.be/nl/pers/echtverklaring-resultaten>; et "Comprendre les États généraux des médias d'information", Association des journalistes professionnels, 2014, consulté le 20/05/2019, URL : <http://www.ajp.be/egmi/appli>

¹⁹ "Baisse généralisée des ventes papier pour les journaux francophones en 2012", Belga, *Le Vif/L'Express*, le 01/02/2013, consulté le 20/05/2019, URL : <https://www.levif.be/actualite/baisse-generalisee-des-ventes-papier-pour-les-journaux-francophones-en-2012/article-normal-135683.html>

²⁰ "Spectaculaire croissance du lectorat de *L'Echo*", le 18/09/2018, consulté le 20/05/2019, URL : <https://www.lecho.be/entreprises/media-marketing/spectaculaire-croissance-du-lectorat-de-l-echo/10050592.html>

timédia spécialisé dans l'édition de périodiques en néerlandais et en français (*Le Vif/L'Express, Trends Tendances, Femmes d'Aujourd'hui, Sport magazine, Knack, ...*)²¹. En 2018, Mediafin annonce une croissance de 10% avec un chiffre d'affaires de 64 millions d'euros, ainsi que des investissements dans l'innovation à hauteur de 2,5 millions d'euros. Les produits digitaux sont visés par cette mesure car c'est là où se trouve la majorité des audiences : 55% des lecteurs de *L'Echo* et du *Tijd* consultent leurs versions numériques²².

4.1.1.2 Les membres de la rédaction de *L'Echo* associés au projet

Sur la cinquantaine de journalistes que compte la rédaction, six sont attachés au service "Investir" de *L'Echo*. Ils sont habitués à jongler tant avec les chiffres qu'avec les écrans et les tableaux. S'ils ont participé à la conception de "Quotebot", leur implication n'a pas été la même pour tous, dès lors que l'organisation des réunions de travail ne correspondait pas toujours à leurs moments de présence à la rédaction. Trois générations de journalistes sont représentées dans cette rédaction. Ils sont formés en journalisme, en sciences économiques ou en droit. Un de ces journalistes a participé à la conception de l'artefact alors qu'il n'en sera pas utilisateur final, parti en retraite avant le lancement de "Quotebot". Si le responsable du développement et de la transition numérique du journal a initié le projet, il en a assuré le pilotage tout au long de ses phases de conception et de production²³.

Nicolas Becquet, responsable du développement et de la transformation numérique

Nicolas Becquet est titulaire d'une maîtrise en littérature générale et comparée et d'un master en sciences de l'information et de la communication. Après une expérience en presse radio à l'issue de ses études (dont trois ans à Bel RTL), il est recruté à *L'Echo* en 2010 où il est d'abord journaliste et webmaster éditorial, avant de prendre en charge la gestion du développement des plateformes numériques du journal. Ouvert à l'expérimentation, en termes de nouveaux formats et de nouvelles formes de narration, il s'est également formé aux langages numériques. S'il s'agit de la première fois qu'il pilote un projet d'automatisation de la production d'informations, son rôle s'inscrit dans la continuité de ses activités professionnelles où, régulièrement, il va assurer un rôle de médiateur entre les équipes rédactionnelles et techniques. Il se décrit lui-même comme un "facilitateur", dont la mission est de créer des processus destinés à "*faciliter le travail des journalistes*". Cela implique, pour lui, de rester à l'écoute des besoins et contraintes inhérents au projet, qu'ils soient administratifs, techniques ou journalistiques. Il envisage les technologies numériques comme des appuis au travail journalistique, lesquelles sont susceptibles d'apporter de la valeur ajoutée à destination des lecteurs. Il dispense régulièrement des formations relatives à l'usage du téléphone mobile dans le reportage, et est profes-

²¹ "Roularta acquiert la moitié de Mediafin (L'Echo et De Tijd)", *L'Echo*, 22/10/2017, consulté le 10/10/2019, URL : <https://www.lecho.be/entreprises/media-marketing/roularta-acquiert-la-moitie-de-mediafin-l-echo-et-de-tijd/9938203.html>

²² "Mediafin enregistre une croissance de 10% grâce à la hausse du nombre de lecteurs et d'annonceurs", Mediafin, consulté le 10/10/2019, URL : <https://www.mediafin.be/fr/chiffres-cles/>

²³ La méthode de collecte des informations de cette section s'appuie sur des contacts personnels (par courriel ou par téléphone), d'un formulaire diffusé en ligne ("Présentation du service d'informations boursières"), les pages LinkedIn des membres de la rédaction, et des pages de présentation de chaque journaliste sur le site web de *L'Echo*.

seur invité à l'Université Catholique de Louvain (UCL). Il est également l'auteur d'un blog de réflexion sur les pratiques numériques et les évolutions technologiques, où il partage régulièrement ses expériences²⁴.

Carine Mathieu, responsable du service "Investir"

Cette journaliste "*décrypte au quotidien l'actualité des marchés et des placements, les enjeux de la politique monétaire et de la finance. Suit aussi l'évolution des finances personnelles : crédits, immobilier, pension, fiscalité...*"²⁵. Fin 2019, elle a quitté sa fonction de responsable du service "Investir", pour devenir rédactrice en chef adjointe.

David Collin

Titulaire d'un master en sciences économiques, David Collin est entré au service web de *L'Echo* en 1997. Cette première expérience l'a ouvert sur les outils du "webmastering", et c'est dans ce cadre qu'il a appris les bases des codes HTML et CSS, utilisés pour la mise en forme de pages web. Très vite, il a pris en charge la rédaction de chroniques boursières pour le journal, d'abord à cheval entre deux services, pour ne plus se consacrer qu'à la couverture des marchés boursiers (actions, devises, matières premières, obligations, ...) ²⁶.

Marc Collet

Il est l'aîné de la rédaction du service "Investir", et est parti en retraite au cours du développement du projet "Quotebot". Bien qu'ayant suivi une candidature en droit et des cours de graduat de "Banques, Bourse, Finances", il se considère comme un autodidacte. "*J'ai commencé à lire le journal à 12 ans sans avoir imaginé qu'un jour je travaillerais comme journaliste. C'est à 16 ans que j'ai commencé à lire des articles (journaux et analyses, conférences, et autres documents) se rapportant à la finance, sans pour autant toujours comprendre ce que je lisais. Mais j'ai persévéré et travaillé sans cesse cette matière. Et aujourd'hui encore, je continue à beaucoup m'informer et à accumuler des infos*". Il a débuté sa carrière dans le secteur bancaire à la fin des années 1970, et est rentré à *L'Echo* en 1988, au service des cotations pour progressivement assurer des commentaires sur les marchés boursiers. Il a ensuite intégré le service "Investir", où il livrait régulièrement des analyses, se considérant davantage comme un analyste que comme un journaliste. Il aborde sa relation avec les technologies informatiques sous l'angle d'un travail qu'elles facilitent et sous celui de la valeur ajoutée qu'elles permettent dans le traitement de l'information boursière : "*Il serait difficile, aujourd'hui, de s'en passer*". Marc Collet est également co-auteur de plusieurs ouvrages de vulgarisation sur l'économie : "*100 clés pour comprendre la nouvelle économie*", "*La banqueroute de Fortis*" et "*130 ans de L'Echo*" ²⁷.

²⁴ "Mediatype : journalisme en mutation". URL : <https://mediatype.be/>

²⁵ Source : <https://www.lecho.be/auteur/Carine-Mathieu.883.html>

²⁶ Source : <https://www.lecho.be/auteur/David-Collin.886.html>

²⁷ Communication personnelle, via courriel, le 26/05/2019.

Jehan Goffin

Jehan Goffin est le cadet du service "Investir", qu'il a rejoint en mai 2014, à l'issue de ses études en journalisme à l'École de journalisme de Louvain (EJL). Ce trentenaire y "*couvre l'actualité des marchés boursiers en Europe et à Wall Street en direct dans le Market Live*"²⁸.

Philippe Galloy

Titulaire d'une licence en droit, qu'il a décrochée en 2002, et d'un diplôme d'études complémentaire en journalisme, qu'il a obtenu en 2003, Philippe Galloy (41 ans) a travaillé comme indépendant pour *La Libre Belgique* et *Le Vif/L'Express*, avant de rejoindre *L'Echo* en 2010. Parallèlement, il a également été journaliste à la RTBF pendant dix ans. "*Observateur des marchés boursiers depuis 2004, je suis de près l'actualité financière (actions, obligations, matières premières, ...). La politique monétaire est l'un de ses sujets de prédilection. Je couvre aussi régulièrement les matières fiscales. Grand défenseur des jeux de mots dans les titres des articles*", indique-t-il sur sa page de présentation en ligne²⁹.

Jennifer Nille

Jennifer Nille a rejoint *L'Echo* en 2004. Spécialiste de l'actualité des marchés financiers, elle porte "*un intérêt particulier pour la structure des marchés, le trading à haute fréquence, les courtiers en ligne et les placements*"³⁰. Elle est également auteure de "*La Bourse, une machine infernale : histoire de la Bourse du XIIe au XXIe siècle*", un ouvrage publié en 2014 à la Renaissance du Livre.

4.1.1.3 Syllabs

À l'échelle mondiale, on dénombre moins d'une quinzaine de sociétés spécialisées dans la génération automatique de textes en langage naturel. Les solutions proposées vont du simple code qui extrait des données qui seront utilisées pour remplir des "trous" dans des gabarits pré-écrits en suivant une série de règles prédéfinies, à des approches plus sophistiquées qui vont également fournir une analyse des données (Graefe 2016). En pratique, ces sociétés vont développer soit des services sur mesure via un moteur de rédaction qui sera paramétré par leurs soins, soit un logiciel indépendant de la langue qui pourra être paramétré directement par le client. La start-up française Syllabs appartient à cette première catégorie de prestataires³¹.

Née en 2006, Syllabs s'est d'abord spécialisée dans l'analyse sémantique et l'extraction d'informations (valorisation d'archives, aide à la revue de presse...), avant de développer, dès 2011, un moteur de rédaction, temporairement baptisé "Data2Content". Dès ses débuts, celui-ci prend trois langues en charge : le français, l'anglais et l'espagnol³². L'allemand et le néerlandais s'y

²⁸ Source : <https://www.lecho.be/auteur/Jehan-Goffin.1731.html>

²⁹ Source : <https://www.lecho.be/auteur/Philippe-Galloy.73.html>

³⁰ Source : <https://www.lecho.be/auteur/Jennifer-Nille.69.html>

³¹ Cette section s'appuie sur un entretien exploratoire mené avec les responsables de Syllabs en 2016, trois entretiens avec des représentants de Syllabs dans le contexte du projet "Quotebot", et un corpus d'articles de presse.

³² Source : "Future of journalism : The rise of the robot journalists", Content Insights, 12/07/2016, consulté le 09/07/2018, URL : <https://contentinsights.com/the-rise-of-the-robot-journalists/>

ajouteront en 2018. D'abord active dans les secteurs de l'e-commerce et de l'e-tourisme, la société a progressivement étendu ses activités à ceux de l'immobilier et des médias. Elle y compte parmi ses premiers clients *Les Échos*, *Slate.fr*, *La Tribune*, France Télévisions et *Le Monde*. Les deux fondateurs de Syllabs sont issus du monde de la recherche. Claude de Loupy est docteur en informatique. Avant la création de la start-up, il a dirigé le laboratoire de recherche et développement de Sinequa, une société proposant des moteurs de recherche sémantiques aux entreprises³³. Helena Blancafort est linguiste et traductrice (huit langues parlées)³⁴. Depuis le lancement de ses activités, Syllabs connaît une croissance constante. En 2018, la société annonçait avoir levé deux millions d'euros pour internationaliser son moteur de rédaction. Cet investissement se partage entre Web Valley (basé à Brest, accélérateur de start-up et fonds d'investissement), Sipa Ouest France (basé à Rennes, édite le quotidien *Ouest France*), et BNP Paribas Développement (fonds de développement de la banque BNP Paribas)³⁵. En mars 2018, elle employait dix-huit collaborateurs (contre une dizaine quatre ans plus tôt). Leurs profils professionnels sont ceux d'informaticiens, de linguistes et de spécialistes des données.

Chez Syllabs, chaque projet est déployé en deux phases. La première est relative à la configuration du moteur de rédaction, qui consiste en un système à base de règles. Des scénarios de rédaction sont alors définis manuellement, en fonction des données disponibles. Il s'agit d'une garantie pour que "à la sortie, les informations sont cent pour cent correctes parce que ce n'est pas le moteur qui invente une façon de raconter. (...) Le système choisit de manière aléatoire un chemin, parmi les millions de choix possibles, et produit un texte qui est difficilement lisible. Puis il va vérifier (les paramètres linguistiques, comme les accords et conjugaisons), c'est un lissage des textes. On est capable de mettre des conditionnels à chaque instant de la production (...), en fonction des données ou des répétitions dans le texte. À partir de ça, on peut produire différents types de textes : des contenus froids, des contenus chauds, des contenus tièdes qui seront mis à jour ou du contenu à la demande. On garantit l'unicité des textes. Ils peuvent être publiés tels quels, sans relecture"³⁶. Cette première étape peut prendre "d'une à dix semaines, selon le projet"³⁷ (Figure 4.5).

La seconde étape concerne les tests effectués par le client. À chaque projet, son moteur de rédaction dédié. Lors des élections régionales françaises, quatre médias d'information avaient sollicité les services de Syllabs : quatre moteurs de rédaction différents avaient donc été développés à partir des mêmes données fournies par le ministère de l'Intérieur³⁸. La difficulté, dans un projet destiné à un média d'information, repose souvent dans le manque d'indications

³³ Source : lenouvelobs.fr, URL : <http://www.nouvelobs.com/journaliste/364805/claude-de-loupy.html>, non daté, consulté le 09/07/2018.

³⁴ Source : Production automatique de textes : l'IA au service des journalistes", INA Global, 12/02/2018, consulté le 10/07/2018, URL : <https://larevuedesmedias.ina.fr/production-automatique-de-textes-lia-au-service-des-journalistes>

³⁵ Source : "Syllabs lève 2 millions pour internationaliser son robot journaliste", LesEchos.fr, 10/03/2018, consulté le 09/07/2018, URL : <https://business.lesechos.fr/entrepreneurs/financer-sa-creation/0301387184813-syllabs-leve-2-millions-pour-internationaliser-son-robot-journaliste-319405.php>

³⁶ Claude de Loupy, réunion "kick off", Mediafin, Bruxelles, le 31/01/2018.

³⁷ Source : Content Insights, ibidem.

³⁸ Helena Blancafort et Claude de Loupy, entretien exploratoire, Paris, le 29/02/2016.

fournies par le client : " Très étonnamment, les médias ne font presque rien. C'est plutôt nous qui donnons les exemples de textes, avec plusieurs variantes. Eux, ils vont nous aiguiller par rapport à leurs choix. Le souci est de pouvoir s'adapter à chaque média car ils ne définissent pas précisément ce qu'ils veulent. Ça se fait au fur et à mesure, parce c'est quelque chose de nouveau pour eux. Cela prend du temps pour faire comprendre comment ça marche"³⁹.



FIGURE 4.5 – Processus conceptuel du moteur de rédaction développé par Syllabs, infographie promotionnelle diffusée en ligne en 2019

Les générations automatiques en langage naturel parviennent aux clients sous la forme de flux, auxquels ils souscrivent via un abonnement annuel. Les tarifs pratiqués vont varier en fonction de la taille du projet : lors de la Conférence Nationale des Métiers du Journalisme (CNMJ), qui s'est tenue à Paris en janvier 2019, Claude de Loupy a indiqué qu'un plan tarifaire débutait à 7.000 euros par an. Selon *Les Echos*, ils peuvent s'élever jusque 200.000 euros, selon la taille du projet⁴⁰. Pour Claude de Loupy, faire appel aux services de Syllabs n'a de sens qu'à condition de

³⁹ Helena Blancafort et Claude de Loupy, entretien exploratoire, Paris, le 29/02/2016.

⁴⁰ Source : " Syllabs lève 2 millions pour internationaliser son robot journaliste", idem.

générer un volume de textes conséquent : "*Si vous voulez créer 10 ou 50 articles ou textes, c'est généralement moins cher de les écrire manuellement*"⁴¹. La rentabilité du moteur de rédaction de Syllabs sera donc fonction de la quantité de textes produits : "*Lorsqu'on nous contacte pour produire 90 textes alors que cela va coûter 10.000 euros, on conseille au client de les écrire à la main. Il faut un nombre important de textes et des textes de même nature, au moins un millier de textes. Mais on peut aussi demander 20 ou 30 euros pour un texte qualitatif. C'est pour cela que l'on s'oriente vers une offre sur étagère où l'on prend sur nos fonds pour paramétrer le moteur et où les clients peuvent acheter 50 ou 100 textes*"⁴².

En 2014, Syllabs revendiquait une capacité de traitement allant de 350 mégas à 4 gigas de texte par heure et par processeur, sans contrainte de source, de thème ou de langue⁴³. Pour Claude de Loupy, en termes de vitesse et de quantité d'articles produits, "*la robot-rédaction permet d'aller beaucoup plus loin que ce qu'un journaliste peut faire. Mais en termes d'analyse, celle-ci ne pourra être que succincte puisque la langue porte la complexité humaine, et rien ne nous permet de penser que les machines arriveront un jour à un niveau d'intelligence proche des humains*"⁴⁴.

À l'examen de leurs discours publics, les représentants de Syllabs peuvent être considérés comme des "montreurs de communication" (Latour, cité par Neveu 1994), contribuant "*à diffuser et à légitimer une grille de perception du monde qui en fait un univers d'échanges d'information et de messages (...) S'il peut refléter les représentations vécues par certains professionnels, aider parfois à l'expression de malaises sociaux, ce discours de la communication est fondamentalement discours de dépolitisation des problèmes sociaux*". En l'occurrence, il s'agit ici de rassurer à propos d'une concurrence de la machine sur l'homme... voire celle de la start-up en elle-même sur l'emploi journalistique. Il s'agit également de montrer un moteur de rédaction, dans une activité de communication "*visant à accréditer l'idée du caractère nécessaire, inévitable, de son savoir-faire*" (Neveu 1994).

Rassurer les journalistes

Le positionnement de Syllabs est celui d'un fournisseur de services technologiques, et l'on n'y prétend pas y faire acte de journalisme : il s'agit là d'une constante dans le discours de ses représentants. Lors de la dixième édition des Assises Internationales du Journalisme, qui s'est déroulée à Tours du 15 au 17 mars 2017, Claude de Loupy indiquait : "*Nous ne faisons pas du journalisme (...) Nous produisons juste de l'information sous la forme de texte. Notre objectif n'est pas de remplacer les humains*"⁴⁵. Dans ses différentes interviews ou interventions lors d'événe-

⁴¹ Source : Content Insights, idem.

⁴² Claude de Loupy, réunion "Kick off", ibidem.

⁴³ Source : "Syllabs veut faire parler les gros volumes de données textuelles", Thomas Pontiroli, CLUBIC, 02/04/2014, modifié le 01/06/2018, consulté le 15/07/2018, URL : <https://www.clubic.com/pro/it-business/actualite-693596-syllabs-solution-analyse-semantic-big-data.html>

⁴⁴ Source : "Une IA dans ma rédac", Delphine Soulas-Gesson, Strategies.fr, 30/05/2018, consulté le 24/04/2019, URL : <http://www.strategies.fr/actualites/medias/4012754W/une-ia-dans-ma-redac.html>

⁴⁵ "Robots : amis ou ennemis", 15/03/2017, consulté le 10/07/2018, URL :

ments ou de colloque où le "robot journalisme" est mis en débat, Claude de Loupy ne manque pas de rassurer à ce propos. *"Je me bats pour réfuter cette terminologie de robots journalistes. Ce ne sont en aucun cas des journalistes, mais des robots rédacteurs!"*⁴⁶ (...) *"Le robot de rédaction ne prend pas la place d'un journaliste, mais effectue des opérations souvent répétitives et publiables en temps réel"*⁴⁷. La responsabilité de Syllabs est aussi celle du maintien de l'emploi, indique son fondateur : *"Nous avons clairement une responsabilité en tant qu'acteur sur ce secteur. Notre but est non pas de détruire, mais bien de valoriser l'activité humaine. Notre contribution consiste à soutenir la visibilité des médias sur le web, à fidéliser sur de l'info de service, à appâter plus de lecteurs. Les informations que nous couvrons n'entrent en concurrence ni avec le travail éditorial, ni le reportage de terrain. Nous nous positionnons au contraire comme des libérateurs de potentialité journalistique, en permettant aux journalistes de se concentrer sur ce qui fait leur valeur ajoutée et leur apport à la société"*⁴⁸.

Les responsables de la start-up sont conscients de la sensibilité du sujet dans le monde du journalisme. Dans une carte blanche publiée en mai 2015, "Non, messieurs les journalistes, les robots ne voleront pas votre emploi", Claude de Loupy détaillait les bénéfices de l'automatisation : *"Les 'robots' apportent des points différents par rapport aux journalistes et peuvent même contribuer à l'effort journalistique dans son sens le plus noble en permettant de couvrir un maximum d'informations, même celles qui ne sont pas rentables avec les méthodes traditionnelles (...) Un 'robot' n'est pas capable de corrélérer des événements de manière intelligente en ressortant les raisons de cette corrélation (...) (Il) ne dispose pas d'une vision conceptuelle globale autour d'un événement. (...) Le journaliste (humain, professionnel) apporte un savoir-faire et une connaissance qui n'est pas encore à la portée des robots et qui ne le sera pas avant longtemps, voire jamais totalement"*⁴⁹. Cette limite reconnue au moteur de rédaction est également évoquée dans le contexte du projet "Quotebot" : *"On peut produire automatiquement des textes sur les tops et les flops du jour à la Bourse (...) Mais (...) seul le journaliste est capable d'analyser les causes et les conséquences de ces tops et de ces flops. On ne peut pas laisser une intelligence artificielle décider s'il y a un lien de cause à effet, c'est beaucoup trop complexe"*⁵⁰.

Lorsqu'il compare un texte rédigé par un journaliste à un compte-rendu rédigé par un moteur de rédaction, Claude de Loupy indique que *"contrairement aux journalistes, on est vraiment sur des faits et il manque quelques informations pour colorer le texte. On sait qu'il y a eu un but de tel joueur mais on ne sait pas si le but était magnifique. Là, il manque de la matière palpable et entraînante"*⁵¹. Ce qui manque également, admet-il, ce sont des informations de contexte :

<http://assises.journalisme.epjt.fr/enquete-robots-amis-ou-ennemis>

⁴⁶ Source : "Les robots à l'assaut du journalisme?", idem.

⁴⁷ Source : "Syllabs lève 2 millions pour internationaliser son robot journaliste", ibidem.

⁴⁸ INA Global, idem.

⁴⁹ Source : "Non messieurs les journalistes, les robots ne voleront pas votre emploi", L'Obs/Rue 89, 09/05/2014, consulté le 10/07/2018, URL : <https://www.nouvelobs.com/rue89/rue89-tech/20140509.RUE3752/non-messieurs-les-journalistes-les-robots-ne-voleront-pas-vos-emplois.html>

⁵⁰ Source : "L'intelligence artificielle, futur turbo des rédactions?", Agence France Presse in Le Point, 13/03/2018, URL : https://www.lepoint.fr/high-tech-internet/l-intelligence-artificielle-futur-turbo-des-redactions-13-03-2018-2202005_47.php, consulté le 24/04/2019.

⁵¹ Claude de Loupy, réunion "kick off", Mediafin, Bruxelles, le 31/01/2018.

par exemple, le nombre de matches joué par tel joueur ou son salaire annuel. "*Ces informations ne sont pas facilement récupérables et, parfois, elles ne sont pas factuelles*"⁵². Pour autant, la limite n'est pas seulement technique dès lors que Syllabs déploie ses activités dans un contexte journalistique : "*Les personnes qui travaillent ici sont des ingénieurs et des linguistes. On se rend compte qu'il leur manque une formation éditoriale en rédaction. Les personnes qui ont étudié les sciences du langage ne sont pas forcément des rédacteurs. Ils ont une approche de la langue très technique*"⁵³. C'est pourquoi, en 2018, la nécessité d'expertise du domaine d'application a conduit la start-up à s'adjoindre les services d'un journaliste sportif, dans le cadre d'un projet d'automatisation des résultats de la Coupe du monde de football. Il s'agissait de la première fois que Syllabs faisait appel à un journaliste dans le cadre de ses activités commerciales.

Aujourd'hui, la question d'une menace sur l'emploi journalistique n'est plus éludée : "*Toute arrivée technologique provoque un impact sur l'emploi. (...) Lorsque l'on parle des correspondants et des pigistes, ils auront peut-être moins de choses à rédiger. D'un autre côté, quand on regarde les correspondants, on voit dans les rédactions, ils écrivent un petit billet sur un événement qui s'est déroulé, puis qu'il n'est absolument pas publié. Il est repris par un journaliste parce qu'on n'a pas confiance et qu'il faut revoir la façon dont cela a été écrit et qu'il faut le refaire. (...) Cela représente du travail en plus pour les journalistes. Si on demandait au localier ou aux correspondants d'entrer de la data, on produirait des textes en lesquels on aurait totalement confiance parce qu'on vérifierait éventuellement la data, mais au moins, le texte serait bien rédigé et on n'aurait pas à le réécrire. Ce sont donc des métiers qui vont évoluer, avec de nouveaux métiers et des gens qui vont aussi récupérer de la data (...) Selon moi, le robot-journalisme n'existe pas. Notre service automatisé de création d'articles est en réalité un 'bouche-trou' qui rend service aux journalistes, mais on ne pourra pas remplacer ce métier*"⁵⁴. Pour Claude de Loupy, il existe "*des dizaines de façons de faire collaborer les journalistes et les moteurs de rédactions*"⁵⁵ : qu'il s'agisse d'enrichissements proposés de manière automatique (une biographie ou le descriptif d'une entreprise, par exemple) ou de génération de milliers de textes dans le cadre de projets d'envergure. "*Tout le monde y gagne : les rédactions peuvent imaginer des projets d'envergure avec la création de dizaines de milliers d'articles, tout en se concentrant sur la valeur éditoriale journalistique, les médias voient leur fréquentation et leur audience augmenter, et nous augmentons les possibilités de notre moteur de rédaction*"⁵⁶.

Chez Syllabs, on admet donc désormais que certains emplois pourraient être menacés, tout en envisageant une transformation du métier de journaliste : de producteur de contenus, celui-ci deviendrait producteur de données. Pour autant, cette idée d'un partenariat journaliste-machine, où le premier assurerait une part de récolte et d'encodage des données pour alimenter le moteur de rédaction, n'est pas neuve. Elle avait déjà été émise lors d'un entretien exploratoire, où Claude de Loupy indiquait "*Un rêve. Ce qu'on imagine, ce sont des gens qui remplissent*

⁵² Claude de Loupy, réunion "*kick off*", Mediafin, Bruxelles, le 31/01/2018.

⁵³ Helena Blancafort et Claude de Loupy, entretien exploratoire, Paris, le 29/02/2016.

⁵⁴ Conférence Nationale des Métiers du Journalisme, verbatim, La Sorbonne, Paris, 24/01/2019.

⁵⁵ Source : INA Global, *ibidem*.

⁵⁶ Source : INA Global, *op.cit*.

des formulaires"⁵⁷. Mais s'il s'agit de décharger les journalistes de tâches répétitives et chronophages, de manière à ce qu'ils puissent se consacrer à des activités plus valorisantes, peut-on considérer qu'encoder des données dans un tableur fait partie de ces tâches plus valorisantes et que celles-ci ne sont ni répétitives ni chronophages ?

Un positionnement ambigu

Si elle compte bien parmi ses clients plusieurs médias nationaux et régionaux français (l'Agence France Presse, France Télévisions, *Ouest France*, *L'Express*, *Le Monde*, Radio France, *Le Parisien*, ...), le nombre de projets impliquant Syllabs dans ce secteur reste limité. Dans une autre interview, Claude de Loupy disait regretter le fait que les médias d'information ne considèrent pas l'intelligence artificielle comme une priorité : "*Les responsables innovation nous disent qu'un robot permettrait aux journalistes de mieux faire leur travail et de développer des opportunités commerciales. Mais ce n'est pas dans leurs priorités. En tant que gérant de start-up, je n'avance pas. De vendeur de technologie, je me transforme en vendeur de contenus. Les médias sont l'industrie la plus innovante qui existe, mais une révolution dans la production de contenus est à l'œuvre, et ils se font dépasser par les entreprises sur ces sujets*"⁵⁸. "*On a beaucoup de mal à percer dans les médias. On est leaders en France et on commence à avoir des clients en Belgique, mais il y a toujours des résistances*"⁵⁹. Dans le même temps, il reconnaît que les clients média de Syllabs ne sont pas ceux qui lui rapportent le plus : "*Les médias sont notre marché le plus porteur, mais ce n'est pas forcément celui qui nous rapporte le plus proportionnellement. Vous connaissez l'état des médias. À côté, on a d'autres clients, dans l'immobilier, le marketing, le tourisme et l'e-commerce*"⁶⁰.

Pour séduire le secteur des médias d'information (et les journalistes, pouvant être la source de résistance au sein de certaines rédactions), les arguments de Syllabs ne manquent pas : allègement des tâches, complémentarité, évolution inévitable mais profitable, partenariat gagnant-gagnant, ... Les discours de ses représentants surfent également sur les tendances du moment. En 2015, ses responsables parlaient de systèmes à base de règles impliquant des activités humaines de paramétrage : "*Ce n'est pas vraiment de la robotisation, parce que nous ce sont des ingénieurs linguistes qui paramètrent le système. Le travail est donc un peu manuel. Tout seul, le robot ne sait rien faire. Ce n'est pas la robotisation telle qu'on l'imagine, mais il est vrai que nos interlocuteurs pensent souvent que tout est automatique, que l'on fournit un fichier, en entrée, qui donne un texte en sortie. Ce n'est pas vraiment comme ça que ça se passe*" (Dierickx 2015). Quatre ans plus tard, ce sont les technologies de l'intelligence artificielle qui sont mises en avant. Toutefois, un système à base de règles est à distinguer des technologies caractérisant

⁵⁷ Helena Blancafort et Claude de Loupy, entretien exploratoire, Paris, le 29/02/2016.

⁵⁸ "L'Intelligence Artificielle dans les Médias", ZDNET.fr, Social Media Club, 11/04/2017, consulté le 12/07/2018, URL : <https://www.zdnet.fr/blogs/social-media-club/l-intelligence-artificielle-dans-les-medias-39851112.htm>

⁵⁹ Source : "Le journaliste du futur serait-il un robot?", podcast "Antibrouillard", Usbek & Rica, 23/04/2019, consulté le 25/04/2019, URL : <https://podcasts.usbeketrica.com/article/le-journaliste-du-futur-serait-il-un-robot>

⁶⁰ Claude de Loupy, réunion "kick off", Mediafin, Bruxelles, le 31/01/2018.

l'évolution de l'intelligence artificielle actuelle, telles que le *machine learning* – apprentissage par la machine – ou le *deep learning* – apprentissage profond. L'explication de ce glissement discursif pourrait être trouvée dans l'attractivité d'un secteur où les start-ups attirent entre 15 et 50% de financement de plus que les autres start-ups technologiques⁶¹. Aussi, l'intelligence artificielle résonne-t-elle aujourd'hui comme un argument de marketing surfant sur un effet de mode (Wang 2010).

Une autre évolution du discours du Syllabs est celle de reconnaître exercer une responsabilité "éthique, culturelle et sociétale (...) traiter l'information et la synthétiser sous forme d'articles (...) n'est pas un acte anodin. Il peut avoir des conséquences importantes sur la société s'il n'est pas correctement maîtrisé et encadré"⁶². "Dans les médias, beaucoup d'articles ne sont pas associés à un auteur. Le fait de ne pas dire que cela ne soit pas rédigé automatiquement n'est pas une tromperie. On donne l'info. Quel est l'intérêt, en soi, si le lecteur sait que cela a été rédigé par une machine? On ne cherche pas à tromper. Par exemple, on a été sollicités pour rédiger des faux avis d'internautes. Nous, on ne sait pas comment notre technologie peut être utilisée. On ne veut pas tomber dans n'importe quelles mains. On garde la maîtrise via notre éthique"⁶³. C'est là où se situe toute l'ambiguïté du discours de Syllabs, entre l'affirmation répétée de ne pas appartenir au monde du journalisme (et donc, de ne pas "faire" du journalisme) et celle de vouloir être reconnu comme nouvel acteur d'un secteur pour lequel Claude de Loupy reconnaît volontiers de l'appétence. Il a même affirmé avoir pensé à fonder un média "plus par frustration, on est persuadés que ces technologies peuvent apporter beaucoup aux médias"⁶⁴.

En juillet 2017, l'Agence France Presse (AFP) annonçait la mise en œuvre d'un nouvel "écosystème éditorial" consistant à créer des jeux de données avec une "haute valeur éditoriale ajoutée", à partir d'*open data*, en vue de produire automatiquement ou semi-automatiquement des contenus en temps réel⁶⁵. Ce projet, qui bénéficie d'un financement du Google Digital News Initiative, porte notamment sur le traitement de résultats sportifs ou de prévisions météorologiques, qui seront par ailleurs revendus par Syllabs de manière autonome : "On fait notre propre ligne éditoriale (...) Grosso modo, c'est un AFP light automatisé sur certaines thématiques bien précises"⁶⁶. En mai 2019, à l'occasion des élections européennes, Syllabs lançait un fil d'information automatisé. "En quatre ans, nous avons acquis une expérience technologique et éditoriale, qui nous a permis de définir un standard, à même de répondre aux besoins de tout type

⁶¹ Sources : "The State of AI : Divergence", MMC Ventures, 2019, consulté le 28/04/2019, URL : <https://www.mmcventures.com/wp-content/uploads/2019/02/The-State-of-AI-2019-Divergence.pdf>; "Nearly Half Of All 'AI Startups' Are Cashing In On Hype", Parmy Olson, Forbes, 04/03/2019, consulté le 28/04/2019, URL : <https://www.forbes.com/sites/parmyolson/2019/03/04/nearly-half-of-all-ai-startups-are-cashing-in-on-hype/>

⁶² Source : "Claude de Loupy (Syllabs) : 'Il faut sortir l'intelligence artificielle de sa tour d'ivoire'", MindMedia, 06/06/2018, consulté le 15/08/2018, <https://www.mindnews.fr/article/12284/claude-de-loupy-syllabs-il-faut-sortir-l-intelligence-artificielle-de-sa-tour-d-ivoire/>

⁶³ Source : "Le journaliste du futur serait-il un robot?", idem.

⁶⁴ Source : "Le journaliste du futur serait-il un robot?", ibidem.

⁶⁵ Source : "Press release - AFP Data Project, Funding Google DNI, 06/07/2017, URL : <https://www.afp.com/en/agency/press-releases-newsletter/afpdata-project-featuring-afp-and-three-partners-selected-google-dni-funding>, consulté le 15/07/2018.

⁶⁶ Claude de Loupy, réunion "kick off", Mediafin, Bruxelles, le 31/01/2018.

de média. Nous sommes passés d'un modèle de moteurs de rédaction spécifiques à celui d'un fil d'information automatisé pour tous, nettement plus accessible (...) Elles sont un complément de contenus hyper locaux, trop volumineux et fastidieux pour être traités par une rédaction, et par ailleurs sans pertinence par rapport au savoir-faire journalistique" ⁶⁷.

Dans le cadre du projet "Quotebot", le budget alloué à Syllabs est de 120.000 euros. Il ne comprend pas les frais d'infrastructure, de fonctionnement et de maintenance corrective, qui font l'objet d'un contrat distinct, plaçant ainsi Mediafin dans une situation de dépendance envers le prestataire de services, qui conserve la main sur les flux d'information. Claude de Loupy a assuré la préparation du dossier de financement avec le responsable du développement et de la transformation numérique de *LEcho*. Il a représenté Syllabs lors de la réunion de lancement du "Quotebot", le seul et unique déplacement d'un représentant de la société à Bruxelles en deux ans. En interne, Helena Blancafort a assuré les relations avec Mediafin. Un *project manager* ainsi qu'une linguiste "junior" ont été attachés à ce projet, de même qu'un linguiste néerlandophone engagé quelques mois après le démarrage du chantier.

4.1.1.4 VWD Group

Les données qui nourrissent "Quotebot" sont distribuées, moyennant un abonnement annuel payant, via la société VWD Group, anciennement Time Stock Market Media, dont le siège social est établi à Francfort. Ces données sont donc déjà utilisées par les services bousiers de *LEcho* et *De Tijd*. Active depuis une vingtaine d'années en tant que fournisseur de données financières, cette société trouve son ancrage dans les médias d'information économiques et financiers. Son lien avec les quotidiens de Mediafin est historique. En effet, sa filiale VWD Group Belgium-Netherland s'est construite, en Belgique, sur les cendres de Tijd Electronic Services, née en 1987 à l'initiative du *Tijd*. En 1998, celle-ci fusionna avec la société néerlandaise Beursmedia pour devenir TBM. La société s'intégra dans les activités de Mediafin en 2005, les éditeurs Rossel et De Persgroep voyant là une opportunité pour le développement de l'information financière. Deux ans plus tard, ils se mirent à la recherche d'un partenaire stratégique, considérant que l'extension des activités de TBM s'écartaient trop des leurs. Ils le trouvèrent avec VWD, dont TBM devint une filiale à 100% ⁶⁸. Aujourd'hui, le groupe VWD est implanté dans six pays et compte 2.400 clients business ⁶⁹.

⁶⁷ "Elections : Syllabs déploie son fil info", Thomas Moysan, CB News, le 22/05/2019, URL : <https://www.cbnews.fr/digital/elections-syllabs-deploie-son-fil-info-43659>, consulté le 22/05/2019

⁶⁸ Source : "VWD Group History", via VWD Group Internet Archive, 13/05/2014, consulté le 21/05/2019, URL : <https://web.archive.org/web/20140513025504/http://www.vwd.eu/history>

⁶⁹ Source : "About VWD", consulté le 21/05/2019, URL : <https://www.vwd.com/en/about-vwd/>

4.1.2 Modèle d'analyse et d'évaluation du dispositif sociotechnique

Si ce quatrième et dernier chapitre suit le même cheminement méthodologique que le précédent, des méthodes particulières de recherche ont été mises en œuvre, de manière à proposer une analyse approfondie de l'évaluation d'un artefact technologique dont le développement a été assuré par une start-up spécialisée dans les technologies sémantiques. Il s'agit donc ici de se concentrer, plus particulièrement, sur la seconde hypothèse posée dans cette thèse, relative à l'adéquation des productions automatisées avec les savoir-faire et exigences du journalisme, tant sur le plan de la médiation sociotechnique que sur celui de la perception qu'ont les journalistes des productions automatisées. Cette démarche se justifie par le contexte d'une expérience où les journalistes-utilisateurs finaux se sont impliqués activement dans la conception de "Quotebot". Cette section présente les fondements théoriques des méthodes d'analyse mises en œuvre dans la section relative à la clôture et à l'évaluation de l'expérience⁷⁰.

L'analyse du dispositif sociotechnique se base sur des évaluations de productions générées par le moteur de rédaction de Syllabs, et des évaluations de l'expérience utilisateur des journalistes du service "Investir" de *L'Echo*. Les activités de médiation sociotechnique sont, quant à elle, analysées sous l'angle des principes de la norme ISO 9000, relatifs au management de la qualité, considérant que ces activités vont impliquer des agents sociaux issus de deux mondes professionnels aux savoirs et savoir-faire distincts. Les textes générés de manière automatique par Syllabs ont pour objectif de s'intégrer dans les routines des journalistes du service "Investir", en leur fournissant des brouillons automatisés qu'ils publieront tels quels, qu'ils enrichiront avant leur publication, ou qu'ils ne publieront pas. Ces activités participent à l'expérience que les utilisateurs feront des productions automatisées. Le modèle d'évaluation de l'expérience trouve ses fondements théoriques dans les domaines de l'interaction homme-machine (*Human-Computer Interaction*, HCI) – plus spécifiquement, celui de l'expérience utilisateur (UX) – et du traitement automatique de la langue (TAL), en matière d'évaluations orientées utilisateur.

4.1.2.1 Fondements théoriques

Plusieurs normes ISO permettent de définir, de manière complémentaire, le concept multidimensionnel qu'est celui de l'expérience utilisateur, dont le propos est de rencontrer la satisfaction de l'utilisateur (Abran *et al.* 2003). La norme ISO 9241-11, relative à l'ergonomie des interactions homme-système définit l'utilisabilité d'un système comme "*la mesure dans laquelle un produit (ou service ou un environnement) peut être utilisé par des utilisateurs pour atteindre des objectifs spécifiés avec efficacité, efficience et satisfaction dans un contexte d'utilisation spécifié (particulier)*" (Abran *et al.* 2003). Ce concept vise donc un objectif de qualité (Bevan 1995), un concept défini par la norme ISO 9000 comme "*la totalité des caractéristiques d'un produit, d'un processus ou d'un service qui influe sur sa capacité à satisfaire des besoins explicites ou implicites*" (Boydens & van Hooland 2011). L'ISO/IEC 14598-1, relative aux technologies de l'information et à l'évaluation des produits logiciels distingue le concept de qualité,

⁷⁰ Voir infra p.247

en tant que caractéristique inhérente du produit, de la qualité d'utilisation, définie comme la "*mesure dans laquelle une entité satisfait aux besoins déclarés et implicites lorsqu'elle est utilisée dans des conditions définies*". L'objectif de l'évaluation de la qualité logicielle est donc de garantir que le produit présente une certaine qualité d'utilisation et réponde aux besoins des utilisateurs.

L'expérience utilisateur peut être définie comme "*le degré d'émotions positives ou négatives que peut ressentir un utilisateur spécifique dans un contexte spécifique pendant et après l'utilisation du produit et qui motive son utilisation ultérieure*" (Schulze & Krömker 2010). Elle fait appel à un état émotionnel des utilisateurs, considéré comme crucial pour développer "*des produits satisfaisants et riches en expérience*" (Ganglbauer *et al.* 2009). Par opposition à la perception de l'efficacité, de l'efficience et de la satisfaction en tant que représentations de réponses collectives d'un groupe d'utilisateurs, l'expérience utilisateur est centrée sur l'expérience d'un individu, ses préférences, ses perceptions et ses émotions (Bevan *et al.* 2015). Les facteurs individuels peuvent comprendre des aspects liés aux motivations personnelles (Bevan *et al.* 2015). Cette expérience est susceptible d'évoluer au fil des utilisations (Petrie & Bevan 2009).

Les méthodes traditionnelles d'évaluation de l'expérience utilisateur (UX) comprennent des questionnaires, des entretiens, des techniques narratives, des techniques d'observation et des enquêtes contextuelles. Il existe quantité de cadres de travail afin de mener à bien une évaluation de l'expérience utilisateur. Toutefois, les chercheurs s'accordent, de longue date, sur le fait qu'aucune méthode ne se suffit à elle-même : chaque méthode possède ses avantages et ses inconvénients. Les combiner permet d'aboutir à une meilleure qualité scientifique (Vermeeren *et al.* 2010).

Dans la perspective où l'objectif de convivialité est d'aboutir à une qualité d'utilisation, Bevan (1995) préconise des évaluations fréquentes des maquettes et prototypes. Et "*s'il est pertinent d'évaluer les expériences à court terme, étant donné les changements dynamiques des objectifs des utilisateurs et des besoins liés aux facteurs contextuels, il est également important de savoir comment (et pourquoi) les expériences évoluent avec le temps. De plus, les valeurs des utilisateurs affectent leurs expériences avec les produits et services, et cette relation doit donc être prise en compte dès le début du processus de conception*" (Vermeeren *et al.* 2010). De même que les usages se construisent dans le temps, l'expérience utilisateur consiste en "*une construction dynamique dont la manifestation varie même après l'utilisation du produit*" (Jung *et al.* 2017). Aussi, la prise en compte de la temporalité sur laquelle s'échelonne l'expérience vise-t-elle à envisager "*la manière dont les expériences sont formées, modifiées et stockées*" (Karapanos 2013). En cela, l'expérience utilisateur et la formation des usages peuvent être considérées comme partageant des perspectives communes.

En ce qui concerne le domaine spécifique de la génération automatique de textes en langue naturelle (GAT), celui-ci compte une longue tradition de recherche portant sur des évaluations orientées utilisateurs. Si celles-ci sont importantes pour faire progresser ce domaine (Dale &

White 2007), elles sont aussi compliquées en raison des critères permettant de définir un bon générateur : "*il ne suffit pas de statuer sur la qualité linguistique finale de la sortie, il faut aussi que le texte produit satisfasse au mieux les objectifs qui motivent le processus de génération*" (Danlos & Roussarie 2000).

Dans cette perspective, trois modes d'évaluations peuvent être envisagés : (1) les évaluations centrées sur les tâches, consistant à mesurer l'impact et les performances d'un système en faisant appel à des techniques de psychologie (Belz & Reiter 2006) ; (2) les évaluations basées sur des jugements humains, où des juges évaluent un corpus de textes générés par un ordinateur et de textes rédigés par des humains (ou de textes de référence) – voire deux variantes d'un même texte – en leur attribuant une cote sur une échelle de notation (Belz & Reiter 2009) ; (3) les évaluations automatiques, reposant sur des systèmes d'évaluation métrique qui sont traditionnellement utilisés dans le domaine de la traduction automatique (TA). Ce type d'évaluation est indépendant de la langue (Habash 2004) et ne donne pas lieu à des mesures utiles sur le plan de la qualité des contenus (Belz & Reiter 2006). Qu'il s'agisse d'évaluations humaines ou automatiques, les mesures d'analyse des résultats doivent être significatives, bien décrites et appropriées (Dale & White 2012).

4.1.2.2 Méthodes d'évaluation

Compte tenu des bénéfices scientifiques d'une approche croisée en matière d'évaluation et du contexte particulier de cette recherche, les dispositifs développés ci-dessous ont été envisagés dans leur complémentarité. D'une part, ils visent à évaluer en quoi les textes générés de manière automatiques correspondent aux gabarits de textes rédigés par les journalistes (méthode quantitative). D'autre part, ils ont pour objet d'évaluer la manière dont les journalistes perçoivent la qualité des productions automatisées, dans la perspective de leurs usages (méthode qualitative).

À. Évaluations automatiques

Ce mode d'évaluation vise à répondre aux questions de recherche suivantes : (1) En quoi les productions de "Quotebot" rencontrent-elles les besoins définis par les utilisateurs, lors de la conception du dispositif? (2) Existe-t-il des différences entre les textes sources et les textes cibles? (3) Présentent-ils le même degré de lisibilité? Il se fonde sur l'hypothèse selon laquelle des textes ne rencontrant pas les exigences des journalistes ne seront pas ou moins utilisés. La dimension cognitive du journalisme, qui se rapporte à la connaissance des techniques utilisées par les professionnels et la formation pour les maîtriser, est donc ici mobilisée (Singer 2003). Considérant que la qualité consiste à satisfaire les besoins et exigences des utilisateurs, qu'ils soient exprimés de manière implicite ou explicite, au regard de la définition proposée par la norme ISO 9000, la méthode mise en œuvre consiste à comparer un échantillon de textes sources (fournis par les journalistes lors des travaux préparatoires à la conception de Quotebot) à un échantillon de textes cibles générés de manière automatique.

Ces échantillons ont été évalués selon six indicateurs⁷¹ :

- Le score de Flesch-Kincaid (formule dérivée du score Flesch utilisé en littérature) calcule le degré de lisibilité d'un texte. Ce score varie généralement de 0 à 100 (Flesch 1948, Kincaid *et al.* 1975).
- L'index (indice) Coleman-Liau est un test de lisibilité conçu par Meri Coleman et TL Liau pour évaluer l'intelligibilité d'un texte. Plus le score est bas, moins le texte est lisible (McCallum & Peterson 1982).
- Le grade Flesch-Kincaid se rapporte au système d'éducation américain. Ce score équivaut au nombre d'années d'éducation dont une personne doit disposer pour comprendre le texte. Un score de 8 signifie généralement que le texte est lisible pour le grand public. Un score de 10-12 correspond à un niveau d'éducation supérieur.
- L'index (indice) de lisibilité automatisé (ARI) est une formule destinée à évaluer le degré d'éducation nécessaire pour comprendre un texte. À titre indicatif, un niveau de 8 correspond à la capacité de lecture d'un adolescent de 14 ans, et un niveau de 12 à celle d'un adolescent de 17 ans (Senter & Smith 1967).
- La distance de Levenshtein, ou distance d'édition, mesure la similarité entre deux chaînes de caractères (Yujian & Bo 2007). La formule de calcul suppose de disposer d'un texte source (rédigé par un humain) et d'un texte cible (rédigé par un logiciel).
- Le taux de similarité vise à établir le pourcentage de caractères semblables entre un texte source et un texte-cible (Gomaa et Fahmy 2013).

B. Évaluations basées sur des jugements humains

Plusieurs recherches se sont appuyées sur des dispositifs basés sur des jugements humains pour évaluer la perception, par les audiences, de contenus générés de manière automatique (Clerwall 2014, Kraemer & Van der Kaa 2014, Graefe *et al.* 2015, Haim & Graefe 2017, Wölker & Powell 2018, Melin *et al.* 2018). La première recherche de ce type a été menée par Clerwall (2014). Il s'appuyait sur une grille d'évaluation tenant en douze variables auxquelles il fallait attribuer une note de 1 à 5 sur une échelle de Likert : cohérence du texte, valeur descriptive, texte utilisable, bien écrit, informatif, clair, plaisant à lire, intéressant, ennuyeux, précis, digne de confiance et objectif. Dans l'étude de Kraemer & Van der Kaa (2014), le dispositif consistait également à évaluer un article en utilisant l'échelle de Likert – de 1 (très faible) à 5 (très élevé) – pour apprécier des critères d'intelligence, d'éducation, de fiabilité, d'autorité, de partialité, d'exactitude, d'exhaustivité, de factualité, de qualité et d'honnêteté⁷².

La méthode d'évaluation retenue dans cette recherche s'inspire de ces modèles. Elle consiste en deux évaluations par questionnaires soumises aux utilisateurs finaux de "Quotebot". Les types de réponses sont soit basés sur une échelle de Likert, numéroté de 1 à 5 (de pas du tout d'accord à tout à fait d'accord), soit un mode booléen "oui/non", soit sur des questions

⁷¹ Voir annexe 18, "Modes de calcul des indicateurs automatiques", p.351

⁷² Voir "Dimension évaluative", p.145

à choix multiples. Dans certains cas, les répondants avaient la possibilité de commenter leurs réponses. Ils ont également eu celle de déposer un commentaire libre à l'issue du questionnaire, auquel ils ont répondu de manière anonyme. L'objet de cette méthode est d'évaluer la perception qu'ont les journalistes de la qualité des productions de "Quotebot" et de la manière dont ils envisagent de les intégrer dans leurs routines professionnelles⁷³. Les questionnaires sont un "*un outil couramment utilisé pour l'évaluation de la qualité et de la convivialité des logiciels, en fonction des utilisateurs. Ils permettent une mesure quantitative efficace des caractéristiques du produit*", mais pour obtenir des résultats interprétables, ils doivent être combinés avec d'autres méthodes d'évaluation qualitatives (Laugwitz *et al.* 2008). S'il n'était pas possible d'organiser une évaluation finale du projet avec chacun des journalistes du service "Investir", compte tenu des retards importants accusés dans son déroulement, une évaluation intermédiaire a été organisée avec l'un des journalistes qui s'est le plus impliqué dans le processus de conception de "Quotebot". Il s'agissait d'un entretien, par téléphone, dont le caractère ouvert avait pour objectif de mieux cerner la perception du dispositif, dans une visée qualitative.

4.1.2.3 Médiation sociotechnique et norme ISO 9000

Les activités de médiation sociotechnique s'inscrivent dans le cadre d'interactions entre les agents sociaux impliqués dans le processus de médiation, et elles vont participer à la construction du sens. Elles sont à la fois techniques, considérant que la médiation d'un artefact technologique va structurer les pratiques; et sociales, car elles vont donner du sens à la pratique (Jouët 1993, Flichy 2008). Elles sont également considérées comme un moyen de favoriser l'expérience utilisateur, tout en soulignant le caractère éminemment social de l'innovation technologique. Compte tenu des particularités du contexte organisationnel ainsi que de la multiplicité des acteurs engagés dans le projet "Quotebot", les activités de médiation sociotechnique ont été analysées à travers la grille de lecture proposée par la norme ISO 9000. Celle-ci va permettre d'éclairer le rôle central joué par le responsable du développement et de la transformation numérique tout au long de la conception de l'artefact d'automatisation, s'agissant de s'assurer de la bonne transmission des intentions formulées par les journalistes, en vue de faciliter leur traduction sociotechnique.

L'ISO 9000 est née en 1987 et a fait, depuis, l'objet de quatre modifications. Sa cinquième version révisée date de 2015. Elle fait référence à une famille de trois normes liées au management de la qualité : l'ISO 9000, relative aux principes fondamentaux et à la théorie de la gestion de la qualité et des processus; l'ISO 9001, relative aux exigences d'obtention de la certification; et l'ISO 9004, qui fournit des directives pour l'amélioration de la qualité. Selon la définition de l'ISO 9000, les normes fournissent des orientations et des outils permettant de garantir que les produits ou services répondent aux exigences de l'utilisateur, ce qui signifie que la qualité doit être constamment améliorée. L'application de ces normes est toujours sujette à interprétations, lesquelles dépendront du contexte et des conditions dans lesquelles elles sont appliquées (Terziovski *et al.* 2003).

⁷³ Voir annexe 19, "Formulaires d'évaluation diffusés via Google Forms dans le cadre de l'expérience 'Quotebot'", p.352

Les normes ISO 9000 ont pour objet de garantir la satisfaction du client en impliquant toutes les personnes au sein de l'organisation, en adoptant une approche processus pour gérer les activités et les ressources, et en appliquant une approche système en vue d'une amélioration constante. Les processus sont définis "*comme des activités reliées utilisant des ressources pour permettre la transformation d'entrées en sorties*", impliquant "*une forte concentration*" sur les activités de contrôle de qualité au sein de l'organisation. Cela suppose l'identification et la connaissance des exigences tant des utilisateurs finaux que de toutes les parties prenantes, de même qu'une clarification des processus de gestion des modifications (Singh *et al.* 2011). Les bénéfices de l'ISO 9000 sont donc de fournir une technique au sens de la méthode d'organisation et non pas une technique au sens de la méthode de production. Son application implique des choix collectifs (Maurand-Valet 2007) ainsi que le développement d'une "*culture de la qualité*" au sein de l'organisation (Terziovski *et al.* 2003).

La norme ISO 9000 s'articule en sept principes : orientation client, leadership, engagement du personnel, approche processus, amélioration, prise de décision basée sur des preuves, et gestion des relations (Anttila & Jussila 2017)⁷⁴. L'ISO précise qu'un principe est "*une croyance fondamentale, théorie ou règle qui a une influence majeure sur la manière dont quelque chose est fait. Les principes de gestion de la qualité sont un ensemble de croyances, normes, règles et valeurs fondamentales acceptés comme vrais et peuvent être utilisés comme base pour la gestion de la qualité*". Vu dans la perspective de l'automatisation de la production d'informations, les avantages potentiels de ces principes consisteraient dans la sensibilisation à la qualité au sein de l'organisation, ainsi que dans la perception de la qualité et de la satisfaction de l'utilisateur (Stevenson & Barnes 2001).

L'une des principales critiques à l'encontre de la norme porte sur son incapacité à garantir la qualité du produit ou du service développé (Guler *et al.* 2002), bien que plusieurs recherches aient démontré qu'elle contribuait à améliorer les performances d'une organisation sur le plan du business et des processus (Lo *et al.* 2013, Terziovski & Guerrero 2014). De plus, la mise en œuvre de ces principes ne peut garantir, à elle seule, un usage final, dès lors qu'il s'agit d'un processus qui s'installe dans le temps et qui est susceptible d'être influencé par des facteurs externes et internes au système (Guler & Guillén 2002). Toutefois, l'analyse relative à la manière dont ces principes sont abordés dans une organisation offre une grille de lecture riche et pertinente, considérant qu'aucun processus technique ne peut être envisagé en dehors de ses aspects sociaux.

⁷⁴ Ces principes sont également disponibles via le site de l'ISO. Toutes les citations concernant ces principes sont issues de ce document introductif : "Principes du management de la qualité, consulté le 01/10/2018, URL : <https://www.iso.org/files/live/sites/isoorg/files/archive/pdf/fr/pub100080.pdf>

4.2 Construction sociale de l'artefact technologique



La construction sociale de l'artefact "Quotebot" met en lumière des interactions entre deux types d'agents sociaux : les journalistes, qui vont participer activement à la conception du dispositif d'automatisation en fournissant des gabarits de textes définis en fonction de leurs besoins; et les employés de Syllabs, la start-up en charge du développement du moteur de rédaction. Les premiers jalons de ce projet furent posés début octobre 2017, époque à laquelle le responsable du développement et de la transition numérique de *LEcho* répond à un appel à projets lancé par le Google Digital News Initiative. Les objectifs de "Quotebot" sont doubles : renforcer le positionnement du groupe de presse en tant que "*news as service*", en enrichissant de productions automatisées les portefeuilles virtuels d'actions accessibles aux abonnés des quotidiens; et en tant que soutien aux journalistes, dans le contexte de l'information sur les marchés boursiers fournie en continu dans la rubrique "Market Live" du site⁷⁵.

La réunion de lancement du projet ("*kick off*") a été organisée le 31 janvier 2018, dans les locaux de Mediafin, à Bruxelles. Elle avait pour objet d'en présenter les grandes lignes, mais aussi de mettre autour de la table toutes les parties prenantes au projet. Elle visait également à définir un planning, qui devait s'échelonner sur un an. Celui-ci devait être soumis à Google DNI qui, sur le plan administratif, posait l'exigence de définir trois phases de travail ("*milestones*") devant chacune rencontrer des objectifs particuliers. Deux représentants de Mediafin (Tom Peeters, multimedia manager, et Chris Van Rossem, IT manager) ont participé à cette réunion, de même que Claude de Loupy, qui représentait Syllabs, et Nicolas Becquet, responsable du développement et de la transformation numérique de *LEcho*. Un calendrier théorique a été alors défini comme suit :

- **du 01/04 au 30/06/2018** : structuration des données, identification des besoins, définition éditoriale, adaptation du moteur de génération au néerlandais;
- **du 01/07/2018 au 31/01/2019** : fin de l'adaptation au néerlandais, élaboration et rédaction des textes, tests avec les journalistes et évaluations;
- **du 01/02 au 31/03/2019** : évaluation de l'expérience utilisateur et des usages quotidiens, définition des formats et de la distribution des contenus, implémentation.

Le projet a été présenté aux journalistes du service "Investir" de *LEcho* lors d'une première réunion de travail, le 19 avril 2018. Le responsable du développement et de la transition numérique explique alors que le principe de "Quotebot" est d'automatiser certaines tâches répétitives, dans le cadre des informations diffusées via le "Market Live" : "*L'objectif est de vous alléger le travail (...) pour vous concentrer sur l'analyse, en gagnant en efficacité*". Il précise qu'il s'agit

⁷⁵ D'après une communication personnelle, via courriel, le 12/10/2017.

d'un essai, d'une tentative : "*l'objectif est qu'il vous soit utile*". Le partenaire Syllabs sera brièvement présenté, à propos duquel il dira : "*Nos intentions, il faut qu'ils les traduisent*". "Quotebot" apparaît alors comme une source d'informations supplémentaire pour les journalistes. Sa capacité à pouvoir les alerter en cas de mouvements inhabituels des valeurs boursières (sous-performances ou sur-performances) y est encore soulignée. Cette réunion a posé les jalons d'un *workflow* où les journalistes vont garder la main quant à la définition des contenus et fonctionnalités du dispositif, au regard de leurs besoins. L'accueil des journalistes est enthousiaste, ceux-ci percevant d'emblée les bénéfices de l'outil : à savoir, celui de leur faire gagner du temps dans le contexte d'activités routinières qui leur laisse peu de répit.

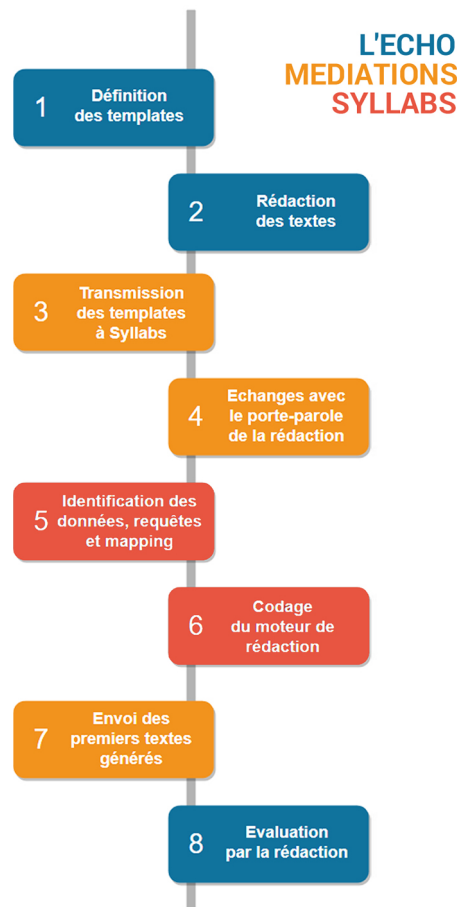


FIGURE 4.6 – Workflow de la première phase de la conception de "Quotebot"

La construction sociotechnique de l'artefact s'est appuyée sur un mouvement d'allers-retours entre les journalistes et Syllabs, via des documents partagés sur une plateforme collaborative en ligne, dans le cadre d'un processus itératif (Figure 4.6). Cinq réunions de travail ont mobilisé les journalistes du service "Investir" de *LEcho*, de la mi-avril à la mi-juillet 2018, en vue de définir leurs besoins. Ils ont été invités à participer de manière active à la conception de "Quotebot", en rédigeant les gabarits de textes (ou "templates") qui furent transmis à Syllabs en vue d'intégrer ces modèles dans son moteur de rédaction. Plusieurs allers-retours entre Mediafin et Syllabs se sont déroulées dans l'intervalle, sous la houlette du responsable du développement et de la transition numérique de *LEcho*. Pour ce dernier, il s'agissait des premières activités de

médiation entre les mondes de la technique et du journalisme dans le cadre de ce projet. Très vite, celles-ci vont apparaître comme fondamentales, dès lors qu'il s'agissait de préciser ou de clarifier les exigences des journalistes, qui détiennent l'expertise du domaine d'application.

4.2.1 Dans le monde du journalisme



L'identification des besoins des journalistes s'inscrit dans le cadre d'un processus inclusif, pouvant être considéré comme une première forme d'usage, dès lors que la figure de l'utilisateur est abordée sous l'angle du rapport entre concepteur et utilisateur dans le cadre d'une co-participation ou d'une co-création (Akrich 2006). Trois journalistes, en particulier, se sont impliqués dans la définition de la structure des contenus, leurs collègues n'étant pas toujours disponibles au moment où étaient organisées les réunions de travail. Dans un premier temps, ils ont été invités à définir les types de contenus susceptibles de soutenir leurs routines, à partir de deux axes principaux : les bilans quotidiens des activités observées sur les marchés boursiers et la mise en exergue des valeurs connaissant les plus fortes variations au cours d'une journée. Au fil des réunions, leurs besoins se sont précisés autour de trois types de contenus – textes, graphiques et tableaux –, envisagés de manière complémentaire. Dans certains cas, l'information est plus compréhensible lorsqu'elle est présentée graphiquement plutôt que textuellement, la génération automatique de textes ne se justifiant pas dans tous les cas (Reiter & Dale 1997). Un journaliste a d'ailleurs souligné que "*le visuel est (...) parfois, (...) mieux que les mots*"⁷⁶. Deux autres demandes seront également exprimées et, cela, de manière répétée : la première porte sur un dispositif d'alerte à propos des valeurs boursières se comportant de manière inhabituelle, tandis que la seconde concerne la possibilité d'établir un historique des valeurs boursières.

Les productions de "Quotebot" sont destinées à couvrir les trois temps forts d'une journée boursière : ouverture de séance, mi-séance et clôture de séance. Mais pas seulement. Elles ont aussi pour objectif de mettre en avant les valeurs ayant enregistré les meilleures et les plus faibles performances. Les différentes demandes formulées par les journalistes vont être classées sur une échelle de priorités éditoriales, la plus haute étant attribuée aux informations relatives à l'ouverture et à la clôture des marchés et la plus basse, aux bilans de mi-séance. Les journalistes ont défini les places boursières devant être intégrées dans le flux d'informations (Bruxelles et New York), ainsi que les indices boursiers devant faire l'objet d'un suivi quotidien (BEL20, DAX, CAC 40, AEX, Footsie, Euro Stoxx 600, S&P 500 et Nasdaq). Les journalistes ont également été invités à s'exprimer sur la manière dont ils souhaitaient recevoir les contenus automatisés : via e-mail en ce qui concerne les alertes relatives à des comportements inhabituels de valeurs boursières, via le système de gestion de contenus interne, en ce qui concerne

⁷⁶ Réunion 2, Bruxelles, le 26/04/2018 et Réunion 3, Bruxelles, le 09/05/2017.

les productions générées de manière automatique. Il est convenu que chacun de ces contenus fera l'objet d'une validation, à charge du journaliste de décider s'il le publiera tel quel, s'il l'enrichira d'éléments contextuels ou s'il ne le publiera pas. Dans un souci de transparence, les audiences seront averties du caractère automatisé de la publication, une mention spécifique ayant été prévue à leur attention : "*Rédigé par Quotebot*".

Ces aspects conceptuels ont été traduits, sur le plan opérationnel, par l'élaboration de gabarits de texte qui tiennent compte, sur le plan technique, d'une réalisation s'appuyant sur un système à base de règles ("si la valeur est de X alors il s'agit d'une hausse/baisse de X"). Une vingtaine de modèles de textes ont ainsi été travaillés, tout prenant en considération les possibles variations d'indices ou de valeurs⁷⁷. Par exemple, l'ouverture d'un marché boursier peut présenter une tendance à la hausse ou à la baisse, une légère baisse ou une légère hausse, un emballement ou une chute. Dans un souci de précision et considérant que les journalistes sont les experts de leur domaine d'application, ces derniers ont défini des fourchettes de variation qui quantifient ces phénomènes. En raison d'un vocabulaire spécifique inhérent à l'information boursière, et de manière à pouvoir proposer une variation des générations de textes, les journalistes ont travaillé sur une taxonomie comportant plusieurs termes spécifiques et génériques, ainsi que des expressions de référence⁷⁸. L'approche par *template* qui fut ici privilégiée porte moins sur la réalisation syntaxique et la détermination des processus : elle trouve sens lorsque la variabilité des textes générés est limitée, ce qui est le cas dans le contexte de "Quotebot", dès lors que les textes à automatiser ne dépassent pas une dizaine de lignes. Son autre avantage réside dans une meilleure compréhension du processus de génération par les experts du domaine (Reiter & Dale 1997). Les spécifications prévoient que chaque gabarit de texte (ou "template") comporte un titre, et suggèrent que les textes générés de manière automatique intègrent l'heure de leur production.

L'objectif de ce travail préparatoire était de fournir à Syllabs un maximum d'indications, en vue de répondre de la manière la plus conforme aux demandes exprimées par les journalistes. Toutefois, le responsable du développement et de la transition numérique de *L'Echo* fera part de la déception de Syllabs à propos des formats de textes fournis : "*Ils étaient un peu déçus parce que ce sont des petits textes, très factuels. Ils ne peuvent pas laisser exploser leur style en termes de structure syntaxique. Par exemple, ils travaillent sur des descriptions de produits qui sont beaucoup plus longues, avec plusieurs paragraphes. Avec les informations de base qu'on leur a données, il n'y a pas 36.000 constructions possibles de la phrase. Je vais leur demander des propositions et on dira si ça marche ou pas. (...) Je ne veux pas broder pour broder mais ils ont proposé d'aller plus loin si on le veut*". Un journaliste dira qu'il a compris le projet comme portant sur des textes courts, tandis qu'un autre précisera que le "Market Live", qui est présenté sous la forme d'un blog, suppose des textes suivis, et donc courts⁷⁹.

⁷⁷ Voir annexe 20, "Échantillon de textes rédigés par les journalistes du service "Investir" de *L'Echo*", p.355

⁷⁸ Par exemple, "augmenter"/"progresser", "à la hausse"/"en augmentation", "dans le rouge"/"recule"/"rétrograde", "emballement"/"monte en flèche", "À l'autre bout du spectre"/ "Les meilleures progressions"/"Au sein de l'indicateur".

⁷⁹ Réunion 3, idem.

En ce qui concerne les représentations de l'objet technique, les journalistes impliqués dans le projet n'ont pas témoigné d'*a priori* négatifs tout au long de cette phase de conception. Leur perception de "Quotebot" est celle d'un outil qui va soutenir des routines professionnelles répétitives et, partant, leur faire gagner du temps : *"On va la faire travailler, cette machine!" "C'est une source d'informations supplémentaire qui va nous faciliter la tâche"*⁸⁰. *"Si ça peut nous aider à ne pas recopier des cours de bourse" (...)* *"Si ça peut nous aider à automatiser certaines tâches pour nous concentrer sur autre chose" (...)* *"C'est chouette comme projet, c'est bien"*⁸¹. L'aspect de la compétition entre l'homme et la machine a été abordé, de manière indirecte et sur le ton de la boutade, lors de la quatrième réunion de travail. Il s'agissait, plus précisément, de la concurrence jouée par Syllabs sur le terrain de l'information : *"C'est eux qui vont nous remplacer? On les aide à nous remplacer"*⁸². Les limites du système seront également pointées : *"il y a une limite d'appréciation que la machine ne peut avoir"*⁸³. La possibilité que l'artefact contienne une part d'eux-mêmes a également été évoquée, dans le contexte du départ à la retraite d'un journaliste du service. Cela a fait émerger un éventuel rapport "affectif" à certains contenus textuels, voire à l'artefact lui-même. Toutefois, lors d'une précédente réunion de travail⁸⁴, la question de rebaptiser "Quotebot" avec un prénom qui soit plus "humain" fut posée aux journalistes, mais ils préférèrent que cette appellation ne soit pas modifiée.

De la fin avril à la mi-juin 2018, les différents gabarits de textes ont été finalisés et validés par les journalistes, chacun d'entre eux comportant une ou plusieurs variations. Une cinquième réunion de travail a été organisée le 16 août 2018, en vue de faire le point sur l'avancement du projet, sur les indices boursiers qui y sont intégrés – l'Euro Stoxx 600 n'étant pas disponible via VWD, il est décidé de le remplacer par l'Euro Stoxx 50 –, sur l'heure de récupération des données et sur le moment de transmission des textes générés.

À cette occasion, il a été question d'un décalage de quinze minutes dans la diffusion des données relatives à la Bourse de Londres et à l'indice Nasdaq. Ceci s'explique par la nature du contrat avec VWD, qui ne prévoit pas une transmission en temps réel dans tous les cas. Dès lors que la fraîcheur des données a un coût, une modification des termes de contrat aurait engendré, pour Mediafin, un surcoût conséquent. *"On ne pouvait pas faire confiance aux data, soit en raison de ruptures de services liées gros bugs chez le prestataire, qui ont d'ailleurs affecté beaucoup de leurs clients, soit parce que la nature de notre contrat impliquait la fourniture de données en temps réel pour certaines et, pour d'autres, avec quinze minutes de délai. Cela nous a contraint à remettre à plat les conditions de ce contrat. Une fois que l'on a compris tout cela, c'était beaucoup plus clair. On a donc dû détricoter certains formats que les journalistes avaient prévu, en raison du manque de fraîcheur des données"*⁸⁵.

⁸⁰ Réunion 1, Bruxelles le 29/04/2018.

⁸¹ Réunion 4, Bruxelles, le 16/05/2018.

⁸² Réunion 4, Bruxelles, le 16/05/2018.

⁸³ Réunion 2, idem.

⁸⁴ Réunion 3, ibidem.

⁸⁵ Nicolas Becquet, communication personnelle, Bruxelles, le 19/12/2019.

Si le développement de "Quotebot" semble bien engagé, mais les retards vont s'enchaîner. Les raisons ne seront pas seulement à trouver dans certaines limites d'accès aux données, mais aussi dans un calendrier difficile à harmoniser entre Mediafin et le prestataire IT en charge de l'implémentation du flux d'information transmis par Syllabs dans le système interne de gestion des contenus. À cela, va s'ajouter un processus de correction qui va s'avérer plus long que prévu, non seulement parce que les activités liées à "Quotebot" s'ajoutaient aux charges de travail habituelles des journalistes et qu'il s'agissait de pouvoir dégager du temps pour s'y consacrer, mais aussi parce que la société en charge du développement de "Quotebot" manquait d'expertise dans le domaine de l'information boursière. Le projet sera une nouvelle fois freiné à l'été 2019, alors qu'il entamait la phase finale de son développement. Cette fois, il s'agissait d'une résistance exprimée par l'un des journalistes du service "Investir". Celle-ci n'était pas dirigée contre le projet en tant que tel : son caractère, qualifié de violent et agressif, remettait globalement en cause l'organisation du média⁸⁶. Ce fut l'élément déclencheur qui contribua à geler le projet "Quotebot" jusqu'à la fin 2019, soit deux ans après l'octroi de la bourse du Google DNI. Nous reviendrons sur cet épisode lors de l'analyse des observations.

4.2.2 Dans le monde de la technique



Les travaux préparatoires à la conception de "Quotebot" ont fait l'objet de comptes-rendus réguliers auprès de Syllabs par le responsable du développement et de la transition numérique de *L'Echo*, via des appels passés par des services en ligne⁸⁷ ainsi que via des échanges de courriels. Ces communications ont permis d'aborder des aspects plus techniques, essentiellement liés à la transmission des données de Mediafin vers Syllabs, dans la visée opérationnelle de collecter "la bonne donnée au bon moment". Pour Syllabs, cet aspect est particulièrement important, dès lors qu'il influe sur la nature des requêtes effectuées via l'API (interface de programmation) de Mediafin : *"Dans votre contrat fournisseur des données apparemment les données en temps réel ne sont pas prévues pour Londres. Comment peut-on avancer sur ce point pour savoir exactement quelle info est fournie à quel moment? Il est important que ce point soit clarifié afin d'éviter des modifications dans les robots et notre infrastructure (call des données)"*⁸⁸. Plusieurs éclaircissements relatifs aux gabarits de textes fournis par les journalistes ont également été traités, que ce soit à propos de leur structure et de possibles variations lexicales ou syntaxiques, des caractéristiques d'un indice boursier, ou de la définition de seuils permettant, par exemple, de définir un effet de stagnation ou de forte hausse.

⁸⁶ Communication personnelle, Bruxelles, le 19/12/2019.

⁸⁷ Via Skype ou Google Hangout.

⁸⁸ Courriel envoyé par Syllabs, le 30/08/2018.

Il s'agissait aussi, pour Syllabs, d'expliquer la manière dont un processus de génération automatique se déploie. *"La première étape c'est de comprendre les données qu'on va recevoir. On demande à nos clients (...) un échantillon de données (...) On a besoin de savoir quelles sont les données qu'on va recevoir. On va s'assurer qu'il ne manque pas d'informations dans les textes. (...) Il faut savoir si je requête les données directement via leur API (VWD) à eux ou si je dois passer chez vous (...) Par exemple, pour faire un compte-rendu de match de football, on requête cinq API différentes (équipes, actions de jeux, classification...) car on n'obtient pas toujours toutes les informations avec une seule requête. (...) Si vous avez commencé à travailler les templates, on peut regarder de notre côté et confirmer qu'on peut le paramétrer de notre côté pour avoir le wording (...) Les personnes qui paramètrent le système écrivent de manière manuelle, puis on valide avec les clients. Ensuite, on paramètre le moteur et on partage une version des textes générés de manière automatique. On fait une deuxième génération qui va tenir compte de vos retours et on itère. Dès qu'on a une première version, on peut créer un web service, de manière basique, pour le mettre à votre disposition et vous voyez comment vous les traitez. (...) La priorité est de définir un format d'échange. Avec les médias, on a deux modes de fonctionnement : soit on push (les textes), soit le média requête notre service pour récupérer les textes"*⁸⁹. Pour cette représentante de Syllabs, le fait que le projet soit appuyé par la rédaction est "génial" et "change tout"⁹⁰. Elle se dit également "impressionnée de voir tout l'avancement et tout le travail qui a été fait. La bonne nouvelle est que tout me semble très bien. La data me semble bien structurée, propre et utilisable. J'ai regardé les templates et tout me semble très clair. Il n'y a rien de complexe pour nous, ce sont des choses que l'on sait faire"⁹¹. Elle demandera néanmoins des précisions relatives à la nature des textes – "S'agit-il d'un texte final ou juste d'un bout de texte qui doit être contextualisé?" –, à leur contexte de publication et à la marge laissée à Syllabs sur le plan de la structure des textes – "Il s'agit d'une structure figée, presque comme des textes à trous. Doit-on strictement suivre les phrases telles quelles ou des variations sont-elles possibles?".

Habituee à travailler dans la perspective d'une optimisation pour les moteurs de recherche, elle indique que *"en relisant votre dossier, je n'ai pas l'impression que ce soit votre priorité. (...) Nous nous adaptons à la demande du client (...) Pour L'Express, la demande portait sur la variabilité (...) Pour Le Monde, les templates étaient bien figés et il n'y avait pas beaucoup de possibilités. À l'AFP, on considérait que certaines de nos variantes étaient un sacrilège"*. Le responsable du développement et de la transformation numérique de L'Echo lui explique alors qu'il y a peu de marge dans la rédaction des textes : *"L'information est courte et très codifiée. C'est pour cela que l'on automatise : c'est tellement récurrent"*⁹². Par ailleurs, des questions subsistent à propos du vocabulaire employé et de la différence entre certains mouvements de valeurs⁹³. La représentante de Syllabs précise également que s'agissant de la génération automatisée de graphiques, qui s'appuie sur la bibliothèque JavaScript "Highcharts", une assistance sera fournie par la start-up WeDoData – sa voisine de bureaux à Paris, spécialisée dans la réalisation de vi-

⁸⁹ Appel passé via Skype, le 02/05/2018.

⁹⁰ Appel passé via Skype, le 02/05/2018.

⁹¹ Appel passé via Skype, le 04/05/2018.

⁹² Appel passé via Skype, le 04/05/2018.

⁹³ Appel passé via Skype, le 07/06/2018.

sualisation de données pour les médias français – pour leur intégration. " *On commence à faire des graphiques plus évolués que ce que l'on faisait jusqu'à présent. Il s'agira soit d'images fixes, soit de dataviz dynamiques*"⁹⁴.

Chez Syllabs, trois personnes sont attachées au projet "Quotebot" : la co-fondatrice de la société, linguiste de formation, qui est en charge de la relation avec le client Mediafin; un responsable de projet (*project manager*), qui gère les aspects relatifs à la collecte et au stockage des données ainsi que ceux relatifs à l'envoi des productions générées de manière automatique; et une linguiste junior, qui travaille sur le codage des textes, lesquels seront validés par une linguiste senior. Des entretiens avec ces trois agents sociaux ont été organisés en cours de réalisation du projet, dans les bureaux parisiens de Syllabs. Ils ont permis de mieux comprendre la manière dont ils appréhendent leur collaboration avec le monde du journalisme.

Habituellement, un processus de génération automatique de textes débute, chez Syllabs, par une analyse des données " *pour comprendre ce que l'on va recevoir*". Un échantillon de textes est ensuite rédigé à la main, avant d'être soumis au client qui le validera ou l'amendera dans le cadre d'un processus itératif. Cette part du travail éditorial a été, ici, prise en charge par les journalistes-utilisateurs finaux de "Quotebot". Cette manière de faire a facilité le travail de Syllabs, indique la co-fondatrice de Syllabs " *mais ce qui a manqué, pour que cela soit parfait, dans le cas où nous travaillons nous-mêmes sur un échantillon manuel, ce sont les données. C'est une question de savoir ce que l'on a entrée et de savoir ce que l'on peut avoir en sortie*"⁹⁵.

Cet aspect relatif aux données ainsi qu'à leur qualité avait été évoqué lors d'un entretien exploratoire avec Syllabs : sans "bonne data", pas de bonne génération : " *il y a des contrôles sur la cohérence des data mais si la data est fausse, on ne peut pas faire grand-chose. Dans certains projets, on a dû faire ça manuellement (...). Par exemple, on avait dû valider des données sur le web à propos du tourisme et c'est une catastrophe : sur Trip Advisor, il y a 50% d'erreurs. Les sites web des hôtels, c'est aussi une catastrophe, avec 25% d'erreurs. Il y a une vraie problématique de fiabilité des données et c'est clair que, dans un contexte journalistique, il faut faire attention. Il y a plusieurs façons de voir le fact-checking, tel qu'on l'entend habituellement : on va vérifier si l'information d'un texte est correcte de manière automatique et pour faire ça, il faut une donnée fiable; l'autre partie est de comparer différentes sources, mais ça demande beaucoup de moyens si les données ne sont pas structurées*"⁹⁶.

Dans le projet "Quotebot", il s'agissait de la première fois qu'une rédaction s'impliquait à ce point dans la définition d'un projet, nouant ainsi une collaboration étroite avec les journalistes. " *C'est un point extrêmement positif car lorsque l'on travaille avec d'autres médias, c'est nous la force de proposition au niveau de la ligne éditoriale. Finalement, pour tous nos projets, c'est à nous de nous adapter et à devenir un peu spécialistes du sujet, comme pour les élections ou*

⁹⁴ Appel passé via Skype, le 04/05/2018.

⁹⁵ Helena Blancafort, Paris, le 24 juillet 2018

⁹⁶ Helena Blancafort et Claude de Loupy, op.cit.

pour le football, où on a travaillé avec un journaliste externe pour comprendre les règles. Ce qui est nouveau, c'est d'avoir un client qui définit les 'templates' de manière détaillée (...) Ce que les journalistes appellent templates, nous, on les appelle des règles pour la génération de textes (...) Ça nous a facilité la vie mais ce qui a manqué, pour que ce soit parfait, c'est que nos échantillons manuels partent toujours de la donnée. On part toujours de la data pour imaginer le texte. Là, le souci qu'on a eu, c'est qu'on ne sait pas toujours exactement ce que l'on a en entrée. Ce qui manque, c'est la correspondance" ⁹⁷.

Un autre problème soulevé porte sur l'expertise du domaine d'application. Aussi, la linguiste en charge du paramétrage du moteur de rédaction s'est-elle retrouvée face à des difficultés inattendues lors de la préparation des textes, indique la responsable de Syllabs : "*Je ne m'y attendais pas. Elle ne connaissait pas du tout la bourse. Je savais que la bourse est un domaine particulier mais le niveau des templates qu'on a ne sont pas très spécifiques, parce que ça fait partie des choses qu'on fait. Après, je me suis rendue compte que, par exemple, elle ne savait pas ce que c'était un graphique intraday*⁹⁸. *Ça ne m'apparaissait pas aussi complexe, mais je viens aussi de la traduction économique*" ⁹⁹.

Étant donné la structure fixe et le format court des textes à générer de manière automatique dans le projet "Quotebot", les manières de faire varier ces textes relève davantage du niveau lexical que syntaxique. L'information boursière est caractérisée par un jargon particulier, nécessitant une certaine expertise. Si celle-ci relève du domaine des journalistes, la linguiste en charge du codage de la première version des textes indique qu'elle a dû se documenter sur un sujet avec lequel elle était peu familière : "*Je ne suis pas une experte de la bourse. Cela m'a poussée à me documenter, à faire des recherches (...) Je mange, je bois, je dors bourse pour l'instant (...) grâce aux vrais journalistes qui travaillent dessus*" ¹⁰⁰.

Sur le plan des données, la responsable de Syllabs indique que les principales questions posées sont celles des flux et de la granularité du web service de Mediafin (à partir duquel les données sont récupérées par Syllabs) : "*Au début, on avait imaginé un web service par template proposé (et non un global) : un pour l'ouverture, un pour la fermeture, etc. Imaginons que l'on veuille vendre ça ensuite à des clients et qu'il y en a un qui est seulement intéressé par l'ouverture de la bourse aux États-Unis, ça ne tient pas la route. Il y a donc aussi cette question qui se pose sur le plan de l'infrastructure, en fonction des abonnements que les clients prendraient plus tard. Comment on va commercialiser ça*". Pour le *project manager* de Syllabs, "Quotebot" ne revêt pas des spécificités particulières, en comparaison avec d'autres projets menés pour d'autres médias. Toutefois, il souligne que chaque jeu de données va toujours disposer de ses propres particularités, ce qui nécessite une vue claire sur l'API fournie par Mediafin : "*La difficulté est*

⁹⁷ Helena Blancafort, idem.

⁹⁸ Graphique "au cours de la journée", qui permet de suivre l'évolution d'un titre durant une séance de Bourse. Source : Zone Bourse, consulté le 06/05/2019, URL : <https://www.zonebourse.com/formation/lexique/Intraday-225/>

⁹⁹ Helena Blancafort, ibidem.

¹⁰⁰ Communication personnelle, Paris, le 24/07/2018.

de comprendre où trouver les informations qui nous intéressent, mais c'est classique et normal. Il faut comprendre les champs, leur étiquetage et leur signification" ¹⁰¹.

Bien qu'elle agisse en tant qu'actrice au sein d'un processus éditorial, Syllabs ne considère pas que la génération automatique de contenus journalistiques relève d'une activité journalistique. *"Nous avons banni la locution 'robot journaliste' car elle peut susciter le fantasme du robot qui remplace le journaliste. Il s'agit de transformer des données en textes et ces textes sont assez mécaniques. Ce n'est pas du journalisme car il n'y a pas d'angle ou d'analyse en profondeur. On donne des résultats, comme on l'a fait pour les élections. Un journaliste va toujours essayer de donner un contexte. Ici, c'est juste du 'data to content', avec des valeurs qui montent et des valeurs qui descendent"*, souligne la co-fondatrice de Syllabs. Cette position est partagée par l'ensemble de nos interlocuteurs. Pour la linguiste, il ne faudrait pas sous-estimer la part humaine qu'implique le développement d'un tel projet : *"C'est moi qui code, ce sont des heures de travail"*. Pour le responsable du projet, l'image du "robot" évoque celle d'un personnage de science-fiction. Il souligne la part importante de l'humain derrière le moteur de rédaction, tout en reconnaissant les limites *"de la génération elle-même ou de ce qu'un outil peut analyser. Dans le journalisme, il y a une certaine part de subjectivité. Nous ne sommes pas capables de l'avoir car nous sommes dans la rationalité"* ¹⁰².

De septembre à décembre 2018, une cinquantaine d'e-mails ont été échangés à propos d'aspects moins journalistiques que techniques qui impliquaient des prestataires IT externes à Mediafin. Des lenteurs ont pu être observées dans les échanges d'informations en ce qui concerne les formats et modes de transmission des contenus. Deux appels ont également été passés via Skype les 21 septembre et le 26 novembre, en vue de préciser ces aspects, de même que ceux relatifs à la planification du projet, à la disponibilité des données et aux priorités dans le développement de "Quotebot".

4.2.3 Processus de correction



Bien que les premières productions de "Quotebot" étaient disponibles depuis le début janvier 2019 via une interface proposée par Syllabs (en accès protégé), en vue de tester les échantillons de textes générés, celle-ci n'avait pas été jugée suffisamment accessible par les journalistes ¹⁰³. Aussi, ces derniers ont-ils pris connaissance des premières générations de textes en février 2019, ce qui correspond au moment où les productions automatisées ont été intégrées dans le système de gestion de contenus (CMS) interne de Mediafin, c'est à dire dans le contexte

¹⁰¹ Communication personnelle, idem.

¹⁰² Communication personnelle, ibidem.

¹⁰³ Voir annexe 21 "Interface de démonstration de Syllabs", p.358

d'utilisation finale des productions de "Quotebot". Chaque nouvelle génération de texte y apparaît dans une liste où figurent également les articles rédigés par les journalistes. Ces productions automatisées s'affichent dans une version non publiée, signalée en rouge (Figure 4.7).

	Time	Title	Body	Status	ID	Headlines
1	20/02/2019 22:20	A suivre ce jeudi 21 février	ECONOMIE / FINANCEAllemagne* 9h30 Indices PMI IHS...	Published	3156778	online
2	20/02/2019 22:18	La Bourse de New York en hausse malgré la Fed	Wall Street a clôturé en hausse mercredi, après...	Published	3156776	online
3	20/02/2019 22:02	Le Dow Jones	Des 30 actions qui composent l'indice Dow Jones,...	Draft	3156762	offline
4	20/02/2019 22:02	New York en ordre dispersé	À la clôture de la séance, l'hésitation dominait à...	Draft	3156761	offline
5	20/02/2019 22:02	Le Nasdaq 100	Sur les 39 actions qui composent l'indice Nasdaq,...	Draft	3156760	offline
6	20/02/2019 22:00	Wall Street clôture en hausse	Le Dow Jones a pris 0,25% à 25.955,73 points.Le...	Published	3156750	online
7	20/02/2019 21:09	Le pétrole termine en hausse à New York	Les cours du pétrole ont terminé en hausse sur le...	Published	3156714	online
8	20/02/2019 20:30	La Fed adopte la "patience" face à un ralentissement de la croissance	La Banque centrale américaine (Fed) a estimé que...	Published	3156720	online
9	20/02/2019 19:30	QF réduit substantiellement le dividende	La société immobilière réglementée QRF va verser...	Published	3156718	online
10	20/02/2019 18:21	Leasinvest augmente son dividende de 2% à 5,10 euros	La société immobilière réglementée (SIR)...	Published	3156610	online
11	20/02/2019 18:13	Cofinimmo annonce un contrat pour son immeuble Quartz	La société immobilière réglementée (SIR) Cofinimmo...	Published	3156603	online
12	20/02/2019 18:05	Résumé de la séance en Europe	Le secteur automobile a bénéficié de l'optimisme...	Published	3156597	online
13	20/02/2019 18:02	Les principaux indicateurs de la séance en Europe	Les chiffres clésIndicesVar.%Euro Stoxx...	Draft	3156593	offline
14	20/02/2019 18:02	La Bourse de Wall Street gagne du terrain ce midi	Les indices actions de New York ont continué la...	Draft	3156592	offline
15	20/02/2019 18:02	L'essentiel de la séance boursière européenne	À la Bourse de Bruxelles, l'indice Bel20 est monté...	Draft	3156591	offline
16	20/02/2019 18:02	Les tops et les flops de la journée en Europe	Hit Parade de l'indice Euro Stoxx 50 TDP...	Draft	3156590	offline
17	20/02/2019 17:42	La Bourse de Bruxelles finit la séance du bon pied	L'indice Bel20 a clôturé en hausse ce mercredi...	Published	3156570	online
18	20/02/2019 17:42	Les principaux marchés du continent européen finissent la séance du bon pied	Les principaux marchés d'Europe étaient bien...	Draft	3156569	offline
19	20/02/2019 17:42	Les principaux marchés européens gagnent du terrain à la clôture de la séance	Les principaux marchés d'Europe étaient en hausse...	Draft	3156568	offline

FIGURE 4.7 – Intégration des productions de "Quotebot" dans le CMS de Mediafin

Dès que les journalistes ont pu prendre connaissance des productions du moteur de rédaction, celles-ci ont été compilées dans un document déposé sur une plateforme de partage, de manière à permettre un travail collaboratif entre les journalistes en charge des corrections. Cette formule facilitait également les échanges avec Syllabs, en vue d'une amélioration des textes générés. Une sixième réunion de travail a été organisée pour baliser ce processus de correction. Ce fut l'occasion, pour le responsable du développement et de la transformation numérique de *LEcho* d'apporter des précisions relatives à la fraîcheur des données, "*sachant qu'il y a un problème structurel lié à notre contrat : il y a un décalage de quinze minutes avec tout ce qui sort d'Euronext. C'est quelque chose que l'on ne peut pas résoudre car cela coûte des milliers d'euros pour avoir les données en temps réel. On va décaler de quinze minutes pour être sûrs d'avoir toutes les données.*". Concernant le moteur de rédaction, il a également épinglé "*un problème sur les formulations, comme 'ce jour', qui n'est pas utile dans le contexte du live. Pour les titres, je pense qu'il faudra encore affiner, comme il n'y a pas énormément de variations et que les textes ne viennent qu'une fois dans la journée. Il n'y a pas besoin d'avoir 10.000 synonymes ou titres différents, il faut juste que ça change de jour en jour. On pourrait donc tourner avec trois titres.*".

Parmi les aspects qui devaient encore être traités, figuraient la question du hit-parade, qui supposait une mise en relation avec l'historique des données "*pour les mettre en contexte*", ainsi que la vérification du nom des valeurs boursières. Il a également été indiqué que, compte tenu des délais, les alertes ne seraient pas traitées dans l'immédiat, la priorité étant accordée au chantier des portefeuilles personnalisés. En ce qui concerne les journalistes, deux problèmes avaient encore été soulevés : "*des textes qui sont encore rédigés de manière un peu bizarre ou alambiquée*", et des inexactitudes dans les titres des textes relatifs aux clôtures de séance¹⁰⁴.

¹⁰⁴ Réunion 6, Bruxelles, le 26/02/2019.

Dans le premier document enregistré le 25 février 2019 sur la plateforme collaborative Google Drive, qui comptait trente-deux pages, pas moins de 150 corrections étaient recensées pour un échantillon de 46 textes (dont certains étaient accompagnés par un tableau compilant des valeurs boursières). Dans un commentaire, un journaliste s'est étonné de ce taux d'erreur élevé, à ce stade de développement du projet. Il lui a été précisé, via courriel, que celui-ci est imputable à un seul réglage : *"Une même erreur du programme (ou formulation qui ne convient pas) a été corrigée à chaque fois, ce qui n'était pas vraiment nécessaire et qui donne effectivement l'impression qu'il reste beaucoup de choses à corriger. Je comprends que cela soit plus naturel pour vous de tout corriger, mais il est plus simple pour nous de repérer les corrections à faire s'il n'y a pas de répétitions"*¹⁰⁵.

Les erreurs relevées par les journalistes portaient tant sur la forme (mots superflus – "ce jour", "globalement" –, répétitions, doublons, redondances, formulations incorrectes ou considérées comme "alambiquées" ou "bizarres"), que sur le fond ("*Je me pose la question de l'utilité de cette colonne*", *La différence entre indices en hausse et ceux en baisse n'est pas assez marquée*", *Comment se fait-il qu'il y ait un second post pour la Bourse de Bruxelles et les marchés européens à la même heure?*", *Wall Street était fermée ce lundi donc les données sont complètement fausses*", *Des 39 actions qui composent l'indice Nasdaq*" alors que *Il y a beaucoup plus d'actions reprises dans le Nasdaq*", *On parle d'indices, non de taux*", *L'utilisation de l'imparfait est logique en journée car les données évoluent chaque seconde mais à la clôture, comme le nom l'indique, les marchés se ferment. Les données sont donc fixes en théorie*", *S'il faut ajouter des données (et, ici, un morceau de phrase) manuellement, on a aussi vite fait de faire le post complètement nous-mêmes, donc il n'est plus pertinent*"). Deux mois plus tard, la plupart de ces commentaires avaient été supprimés de ce document, les problèmes soulevés ayant potentiellement été résolus.

De nouvelles versions des contenus générés par le moteur de rédaction de Syllabs ont été présentées aux journalistes pour une deuxième vague de corrections, du 13 au 17 mai 2019. À l'époque, le lancement officiel de "Quotebot" devait avoir lieu trois mois plus tard. Les problèmes épinglés par les journalistes portaient essentiellement sur la forme (retours à la ligne, espaces manquants, formatage des tableaux, formulations, correction orthographique de certaines valeurs boursières), bien que quelques corrections de fond aient également été demandées ("*Tous les marchés sont en baisse, de manière assez marquée. On ne peut donc pas titrer sur 'en panne tendance'. Solution : 'Les principaux marchés européens orientés à la baisse ce midi par exemple'*", *Le post de clôture fait le bilan de l'ensemble de la séance. Supprimer donc 'à la clôture'*", *Contradiction : le titre parle d'une ouverture en hausse de Wall Street et dans le texte l'indice Dow Jones est parfaitement stable. Comme le Dow Jones est stable et le Nasdaq en hausse, on peut parler d'ouverture en ordre dispersé des marchés américains*", *Comme ces actions sont en hausse, on ne peut pas dire 'limitaient les pertes'*"). Plusieurs problèmes relatifs à la pertinence des titres ont également été soulignés. Vingt-deux productions étaient concernées par ces commentaires. Celles-ci avaient été produites en l'espace d'une semaine, ce qui représente

¹⁰⁵ Communication reçue via courriel, le 04/03/2019.

environ un cinquième du volume total des textes transmis, puisqu'une vingtaine de contenus automatisés doivent normalement être produits chaque jour¹⁰⁶.

À l'occasion d'une nouvelle phase de tests, qui s'est échelonnée de la fin mai à la fin juin 2019, les journalistes ont annoté à plus de 90 reprises un document comportant 25 pages. Leurs remarques portaient tant sur la forme (présentation des tableaux de valeurs ou dans la formulation des titres et des textes) que sur le fond ("*Le cours et la variation correspondent aux données de clôture de la séance du 1^{er} juillet. L'action n'a jamais gagné 2,56% durant cette matinée du 2 juillet 2019*", "*Toutes les données sont à revoir de fond en comble : les variations sont incorrectes (...) On a plus vite fait de citer les deux seules variations correctes*"). Ces allers-retours entre la rédaction et Syllabs témoignent de la nécessaire expertise que requiert le traitement de l'information boursière, à la fois en termes éditoriaux (formulation, vocabulaire, présentation des résultats) et de pertinence de l'analyse. Si des données de qualité sont indispensables pour qu'un tel projet fonctionne, elles ne peuvent être traitées sans une connaissance correcte du domaine d'application auquel elles se rapportent. Pour les journalistes, cet important travail de correction a entraîné une charge de travail supplémentaire. "*Je me disais que ça allait aller assez vite puisque que ce sont des chiffres, que la machine allait les avaler et ensuite ressortir un texte bien fait. Mais au fil du temps, je me suis rendu compte que le projet demandait plus de travail que prévu au niveau de la correction, puisque les personnes chez Syllabs ne connaissent pas vraiment bien les marchés boursiers. Il y avait certaines choses qui nous paraissaient évidentes et qu'ils ne comprenaient pas, ou qu'ils traitaient de manière erronée*", indique un journaliste¹⁰⁷.

4.2.4 Clôture et évaluation de l'expérience

Au moment de clôturer cette thèse, "Quotebot" se trouvait en *stand by*, une dernière phase de tests devant encore être organisée en vue d'y apporter la touche finale qui le rendrait opérationnel. Deux raisons expliquent cet état d'inertie depuis plusieurs mois : d'une part, une forte résistance fut exprimée par un journaliste à l'été 2019, et celle-ci ne fut pas sans conséquences sur le déroulement du projet; d'autre part, le quotidien se trouvait dans une phase de transition, le contraignant à une modification complète de son système informatique, en ce compris en ce qui concerne le système de gestion de contenus utilisé par la rédaction. Dans le premier cas, compte tenu de la communication qui nous a été rapportée à ce propos, on peut raisonnablement parler d'une attitude verbalement violente, en raison de la force des propos tenus, qui était d'abord dirigée contre l'organisation en elle-même et son management. "Quotebot" fut donc l'élément déclencheur d'une crise interne, ce journaliste estimant que le processus de correction entraînait une charge de travail supplémentaire, alors qu'il constatait encore plusieurs erreurs liées aux formulations ainsi qu'à la correction des données et à leur analyse. Sans entrer dans les détails de ce conflit, qui ne fut pas sans conséquences également sur le plan humain, il apparaît que celui-ci aurait pu être évité ou atténué en engageant un processus de dialogue qui ne fut, pourtant, pas possible. Dans le second cas, il s'agit d'un facteur externe au projet qui a eu pour effet de ne pas le considérer comme prioritaire dans cette importante

¹⁰⁶ Voir annexe 23 "Liste des templates de 'Quotebot' et indications sur la récupération des données", p.365

¹⁰⁷ Communication personnelle, par téléphone, le 10/09/2019.

phase de transition IT. En décembre 2019, il était prévu de reporter le projet "Quotebot" au deuxième trimestre de 2020. Dès lors, l'étape de fermeture de la construction sociale de l'artefact n'a pas pu être constatée.

L'évaluation des usages journalistiques de "Quotebot" porte sur ceux des journalistes en tant qu'acteurs de l'innovation sociotechnique. Celle-ci comporte trois volets distincts mais complémentaires : (1) une analyse du rôle de la médiation sociotechnique, qui est apparue comme un élément fondamental dans le déroulement du projet, en termes de transmission d'informations ; (2) une évaluation automatique des échantillons de textes générés, en vue de quantifier leur taux de correspondance avec les exigences posées par les journalistes ; (3) deux évaluations par questionnaire, réalisées en mars et en décembre 2019, visant à comprendre la manière dont les journalistes du service "Investir" perçoivent les productions de "Quotebot" ainsi que celle dont ils en envisagent (ou non) des usages futurs.

4.2.4.1 Analyse de la médiation sociotechnique

La construction sociale de l'artefact "Quotebot" s'est opérée dans le cadre d'une double médiation : de la technique vers le social, et du social vers le social. Ces activités de médiation avaient pour objet de faciliter la transmission des intentions formulées par les journalistes, dans le cadre d'une traduction sociotechnique s'opérant autour de l'objet "Quotebot". Dans la mesure où il s'agissait de le façonner en fonction des exigences posées par ses utilisateurs finaux, elles ont mis en perspective une première forme d'usage où les journalistes sont apparus en tant qu'acteurs de l'innovation (Akrich 2009). Dans ce projet, un agent social a joué un rôle charnière dans la dynamique créée autour de l'objet technique : le responsable du développement et de la transformation numérique de *L'Echo*, dans sa fonction de gestionnaire de projet.

Ce rôle, qu'il décrit comme celui d'un "facilitateur", n'est pas très différent de celui qu'il assure habituellement : celui de faire la jonction entre le monde de la technique et celui du journalisme. *"Dans le cadre de mon travail, je vois les deux côtés : le technique et le journalistique. De temps en temps, je pose des gestes techniques mais, surtout, j'essaie de créer un processus de production qui facilite le travail des journalistes. J'ai appris à gérer les erreurs et à manipuler du code pour la production et la diffusion. Je ne suis pas développeur ou datajournaliste mais je sais comment cela fonctionne. (...) Surtout, je vais donner aux journalistes une méthode, un processus à suivre. Lorsqu'il s'agit d'un projet spécifique, je vais mettre toutes les parties prenantes autour de la table. Je vais essayer d'anticiper les problèmes techniques qui pourraient se poser, et veiller à ce que ce soit utilisable pour les journalistes. (...) Ma maîtrise technique permet (...) de faire le lien avec la rédaction et de dialoguer avec les techniciens. Il y a déjà une forme d'automatisation au sein de la rédaction, dans le sens où les objets technologiques sont conçus pour être réutilisés"* ¹⁰⁸.

Dans le cadre du projet "Quotebot", il a joué le rôle d'un agent de médiation entre les rédactions des services boursiers de Mediafin et leurs hiérarchies respectives, les prestataires IT externes

¹⁰⁸ Nicolas Becquet, communication personnelle, Bruxelles, le 02/02/2018.

collaborant avec Mediafin, la start-up technologique sollicitée pour développer le système d'information, les responsables du Google DNI, et l'artefact technologique en devenir. Il a envisagé ce rôle comme celui d'un "chef d'orchestre" à l'écoute des besoins et contraintes de toutes natures (administratives, techniques et journalistiques). Il a donc assisté le processus de traduction d'un scénario défini par les journalistes du service "Investir", chaque décision technique engageant une distribution des compétences entre les agents sociaux membres du réseau sociotechnique (Akrich 1993) : celle des journalistes, détenteurs de l'expertise de l'information boursière; celle des agents de Syllabs, par la prise en charge du développement technique de l'artefact; et celles des agents des partenaires IT de Mediafin, qui ont assuré l'implémentation du dispositif dans le système de gestion des contenus (CMS) utilisé par les journalistes.

Ses objectifs étaient de s'assurer que le scénario défini en amont puisse être réalisé en aval, et qu'il soit aussi rendu opérationnel sur le plan des usages, sans que cela n'entraîne, pour les journalistes-usagers finaux, le développement d'un savoir-faire supplémentaire, s'agissant de les outiller en vue de leur faire gagner du temps dans leurs routines quotidiennes. Ses activités de médiation ont pris leur sens dans une mise en relation entre des aspects journalistiques et techniques, dans des environnements sociaux de prime abord indépendants, à la fois par la nature de leurs activités distinctes et par leur localisation géographique (Bruxelles pour la rédaction, Paris pour la start-up en charge du développement de "Quotebot", Louvain pour le service IT en charge de l'implémentation de "Quotebot" dans le réseau informatique de Mediafin). L'angle des sept principes posés par la norme ISO 9000 permet de mieux comprendre la portée opérationnelle de cette médiation sociotechnique, à la fois en termes de finalités et de processus.

I. Orientation client : viser l'expérience utilisateur

Le principe d'orientation client de la norme peut être compris dans la définition des besoins des utilisateurs finaux de "Quotebot". D'une manière plus générale, ce principe fait référence à l'expérience utilisateur. Si ces besoins ont été identifiés au cours de cinq réunions de travail, celles-ci ont été organisées, coordonnées et animées par l'agent de médiation. L'objectif de ces réunions était de mobiliser la satisfaction des journalistes et, partant, leurs usages. Le responsable du développement et de la transformation numérique de *L'Echo* se dit très satisfait de ce processus car celui-ci a impliqué un "*vrai dialogue, même si ce n'est pas toujours évident pour eux car ils ont une grosse charge de travail. Au fur et à mesure, des équilibres se sont trouvés et certains journalistes se sont impliqués plus que d'autres pour assurer un relai au sein de la rédaction*"¹⁰⁹. Le postulat de cette approche inclusive était donc celle de favoriser les usages finaux, sans pour autant que ceux-ci puissent être garantis.

¹⁰⁹ Communication personnelle, Bruxelles, le 21/03/2019.

II. *Leadership* : l'importance de la médiation sociotechnique

Le principe de *leadership* de l'ISO 9000 vise à ce que le management "*crée des conditions dans lesquelles les gens sont engagés dans la réalisation des objectifs de qualité de l'organisation*". Dans une application stricte de l'ISO 9000, cette fonction est assurée par un "responsable qualité", qui agit en tant que traducteur de la norme. Il a en charge la réalisation d'audits internes, en vue d'améliorer le système organisationnel selon le principe d'amélioration continue. Un profil professionnel hybride est pertinent pour assurer cette fonction (Maurand-Valet 2007). Ce caractère hybride a été rencontré dans la personne de l'agent de médiation, qui ne considère pas avoir joué un rôle différent dans ce projet en particulier : "*C'était d'ailleurs le principal défi. Comme je me trouvais au cœur du projet, c'était à moi de comprendre chacun de ses paramètres de manière précise, que ce soit en termes de connaissance des marchés boursiers, en termes d'enjeux rencontrés par Syllabs, ou en termes de planning et de ressources relatives au fonctionnement du site web de L'Echo*" ¹¹⁰.

III. Engagement du personnel : la participation, une première forme d'usage

La norme ISO 9000 met l'accent sur la nécessité de s'appuyer sur "*des personnes compétentes, responsabilisées et engagées à tous les niveaux de l'organisation (...) pour renforcer sa capacité à créer et offrir de la valeur*". Lorsqu'un système d'automatisation est développé en tant qu'outil du journalisme, ce principe est mis en pratique dès lors que l'on associe les utilisateurs finaux à la conception de l'artefact. Il s'agit d'une forme d'engagement de l'acteur-utilisateur, mais aussi d'une première forme d'usage (Akrich 2006) qui vise l'appropriation du dispositif socio-technique. Les associations avec les utilisateurs d'une innovation sont essentielles en raison de sa double nature technique et sociale (Akrich 1990). Dans le projet "Quotebot", l'approche participative et itérative peut aussi être comprise comme une manière de démystifier la complexité d'un processus de génération automatique de textes, en laissant la main aux journalistes sur la définition des contenus, supposant qu'ils déconstruisent leur propre manière de travailler en vue de la standardiser. Le responsable du développement et de la transformation numérique de *L'Echo* indique que les journalistes ont témoigné de "*la réceptivité ou de l'intérêt pour le projet*", alors que ce dernier n'est pas une émanation directe du service "Investir". Pour autant, alors que la phase de correction des productions automatisées touchait à sa fin, il est apparu que cette implication avait donné lieu à un travail supplémentaire qui n'avait pas été envisagé, conséquence du manque d'expertise du prestataire de services en ce qui concerne l'information boursière. Pour Syllabs, l'inclusion des utilisateurs finaux dans le processus de conception de "Quotebot" a bousculé une manière de faire habituelle. Aussi, a-t-elle été amenée à s'adapter à une méthode de travail où elle partait de textes pré-écrits par les journalistes pour configurer un moteur de rédaction alors que, généralement, les données constituent le point de départ d'échantillons rédigés manuellement et qui sont soumis au client pour validation, avant de faire l'objet d'un traitement automatisé. De part et d'autre de l'artefact "Quotebot", il y a donc eu reconfiguration des pratiques.

¹¹⁰ Communication personnelle, Bruxelles, le 21/03/2019.

IV. Approche orientée processus : identifier et prévenir les risques

La norme ISO 9000 dispose que "*des résultats cohérents et prévisibles sont obtenus plus efficacement lorsque les activités sont comprises et gérées comme étant inter-reliées dans des processus qui fonctionnent comme un système cohérent*". Cela suppose de penser en termes de processus et de séquences, et donc de développer une forme de "pensée computationnelle". Pour plusieurs chercheurs et professionnels engagés dans une vision prospective, il s'agirait même d'une condition *sine qua non* pour que les journalistes s'adaptent aux innovations technologiques (Gynnild 2013, Linden 2017). De la même manière, il est apparu qu'en raison du contexte journalistique dans lequel s'est déployé le projet "Quotebot", cela supposait également de développer une "pensée journalistique", qui permette aux agents sociaux du monde de la technique de mieux comprendre les exigences d'un processus éditorial où l'expertise du domaine d'application constitue un préalable. "*La connaissance de la matière est très très importante. J'ai appris beaucoup en ce sens (...) mais je ne pouvais pas, malgré mon implication dans le projet, remplacer le savoir-faire de ces journalistes. Concernant Syllabs, le manque de connaissance de la matière est un problème, parce qu'il a fallu faire de la pédagogie sur ce que sont les marchés financiers et comment ils fonctionnent. On l'a fait au fur et à mesure mais si je devais refaire le projet, on aurait dû organiser une session de deux jours sur le fonctionnement et les spécificités de la matière traitée, pour que l'on ait une base de connaissance commune qui permette aussi à Syllabs d'appliquer les choses de façon moins littérale. J'ai été sans doute aveuglé par le fait que le projet était basé sur le domaine d'application et ceux qui l'appliquent au quotidien, les journalistes experts. On aurait dû transmettre davantage cette expertise. Mais on aurait aussi dû se voir car la distance Paris-Bruxelles a créé des forces d'inertie*" ¹¹¹.

Le quatrième principe posé par la norme suppose également de prioriser les améliorations à apporter au système d'information et de prendre en compte la gestion des risques, un élément considéré comme essentiel dans la gestion de la qualité (Antilla & Jussila 2017). Dans un projet d'automatisation de la production d'informations visant les usages des journalistes, les risques peuvent se situer en amont et en aval du système d'information. Dans le projet "Quotebot", ceux-ci étaient notamment liés à des retards de planning, à une échelle de priorité qui n'était pas la même pour tous les agents sociaux impliqués (notamment en ce qui concerne un prestataire IT externe à Mediafin), et à la dépendance envers Syllabs en ce qui concerne d'éventuelles modifications et/ou pour assurer la pérennité du système dans le temps. Aux débuts du projet, les deux principales craintes du responsable du développement et de la transformation numérique de *L'Echo* portaient sur la capacité d'intégration des contenus automatisés au système de gestions de contenus interne; ainsi que sur des demandes spécifiques qui émaneraient des journalistes néerlandophone et qui pourraient difficilement être rencontrées en raison de variables à la fois temporelles et budgétaires. Compte tenu des retards importants accusés dans le développement de "Quotebot" – un risque qui ne fut pas ici possible d'évaluer autrement qu'en faisant l'expérience du projet –, cet aspect n'a pas été traité. En effet, l'exigence avait été

¹¹¹ Communication personnelle, Bruxelles, le 19/12/2019.

posée d'entrée de jeu : la version néerlandaise de "Quotebot" devait nécessairement s'appuyer sur une version aboutie en français.

V. Améliorations : implémenter les demandes des utilisateurs

"L'amélioration est essentielle pour qu'une organisation maintienne les niveaux actuels de performance, pour réagir aux changements dans ses conditions internes et externes et créer de nouvelles opportunités", indique le cinquième principe de la norme. Il s'agit donc de créer les conditions de l'amélioration continue du système d'information. Cela sous-entend que ces améliorations puissent s'opérer tant sur un plan organisationnel qu'opérationnel. Dans le projet "Quotebot", elles ont eu lieu avant la mise en disponibilité du système. Les échantillons de textes réalisés par les journalistes ont d'abord été paramétrés dans le moteur de rédaction de Syllabs, avant de faire l'objet de plusieurs versions qui furent soumises à la validation des journalistes dans le cadre d'un processus itératif. Dans celui-ci, la start-up n'est pas apparue en mesure d'intégrer l'ensemble des exigences journalistiques dans le paramétrage de son moteur de rédaction¹¹², malgré l'assurance de ses responsables ("*Il n'y a rien de complexe pour nous, ce sont des choses que l'on sait faire*", "*Nous nous adaptons à la demande du client*"). Toutefois, l'agent de médiation estime que l'ensemble du processus de développement de "Quotebot" s'est déroulé de manière fluide, malgré la distance physique des différents prestataires de service impliqués dans le projet, et malgré les retards accumulés. Aujourd'hui, *L'Echo* se trouve en situation de dépendance envers Syllabs, non seulement en ce qui concerne la maintenance de "Quotebot" et d'éventuelles améliorations futures, mais aussi dans la transmission du flux des informations générées automatiquement. Il s'agit d'une situation dont on est conscient au sein de l'entreprise, et cela fait l'objet d'un traitement distinct du contrat signé dans le cadre du financement octroyé par Google DNI.

VI. Prises de décision basée sur des preuves : la part subjective de l'analyse

Le sixième principe de l'ISO 9000 dispose que des "*décisions basées sur l'analyse et l'évaluation des données et des informations sont davantage susceptibles de produire les résultats souhaités*". Sur le plan opérationnel, ce principe peut être compris comme la capacité à améliorer le dispositif sociotechnique en fonction des résultats d'évaluations pouvant être réalisées à différentes étapes du processus de conception. Il s'agit là d'un des aspects les plus subjectifs de la norme, dans la mesure où toute activité d'analyse implique un travail d'interprétation et donc de subjectivité. Celle-ci a été abordée tout au long du processus de correction des productions automatisées, les journalistes ayant pris en charge cette dimension évaluative, sans pour autant avoir envisagé la longueur de ce processus, qui s'est échelonné sur une année, de même que la charge de travail que cela engendrerait. Les journalistes ont objectivé ce travail par leur expertise du domaine d'application ainsi que par leur expérience dans la couverture des marchés boursiers. Par ailleurs, les résultats des évaluations automatiques réalisées dans le cadre de cette recherche, malgré leur caractère limité puisqu'elles ne portent pas sur la qualité des

¹¹² Voir "Evaluations basées sur des jugements humains", p.258

contenus, ont été régulièrement transmis au responsable du projet à *L'Echo*¹¹³, qui a également organisé et supervisé le processus de correction, conscient qu'il s'agissait là d'un travail supplémentaire pour les journalistes qui était pourtant essentiel pour garantir l'adéquation des productions automatisées avec les exigences posées.

VII. Gestion des relations : rester à l'écoute de toutes les parties

Le dernier principe de l'ISO 9000 s'intéresse à toutes les parties prenantes d'un projet, considérant que "*pour un succès durable, une organisation gère ses relations avec les parties intéressées, telles que les fournisseurs*". Dans sa mise en œuvre, il recouvre l'optimisation des relations dans le court et long-terme, le partage des expertises et des ressources, une collaboration pour l'amélioration et le développement des activités. Ce principe suppose également de reconnaître les succès de chacun, dans la mesure où chacun influe sur le projet et sa réussite. En filigrane, il renvoie au rôle de l'agent de médiation, induit dans le principe de la norme relatif au *leadership*. Ce principe de gestion des relations a participé à la complexité du projet, estime l'agent de médiation. "*Par exemple, notre prestataire IT n'est là qu'une journée par semaine dans les locaux, et cela complique encore les choses pour synchroniser les agendas et réaliser des allers-retours simples entre chacune des parties. (...) La complexité vient aussi de l'organisation du projet et de la disponibilité des journalistes. J'ai appris énormément en termes de gestion de projet, et je ne connaissais pas très bien les marchés boursiers. J'ai donc acquis des savoirs et des savoir-faire. (...) Il y avait beaucoup de challenges mais j'ai appris à gérer cela, en suivant le projet de manière hebdomadaire pour être certain que les informations circulent. Si je ne faisais pas le lien entre toutes les parties, la communication ne passait pas forcément très bien entre les pôles parce que chacun a ses spécificités et sa manière de travailler*"¹¹⁴. Pour cet agent de médiation, il a donc également s'agit de comprendre les modes de fonctionnement de chacune des parties prenantes, pour assurer une gestion optimale des relations.

4.2.4.2 Évaluations automatiques des textes générés

Les évaluations automatiques se sont déroulées au fur et à mesure de l'envoi de nouvelles versions de textes rédigés via le moteur de rédaction de Syllabs. Ceux-ci ont été comparés aux textes fournis par les journalistes, lesquels avaient pour objet de servir de modèle pour la génération automatique de textes. Le corpus total des textes cibles (générés de manière automatique) est composé de 112 échantillons (Figure 4.8 et Figure 4.9), tandis que le corpus total des textes sources (rédigés par les journalistes) est composé compte 87 échantillons dont certains consistent en deux variantes d'un même texte. Chaque texte cible a été mis en correspondance avec les textes sources écrits par les journalistes¹¹⁵. Ces derniers ont été post-édités de manière légère : les valeurs "XXX" contenues dans le corpus ont été remplacées par des valeurs (aléatoires) exprimées en pourcentage. Les titres n'ont pas été pris en compte, n'étant pas prévus dans les premiers échantillons fournis par les journalistes¹¹⁶.

¹¹³ Voir "Évaluations automatiques des textes générés", p.253

¹¹⁴ Communication personnelle, Bruxelles, le 21/03/2019.

¹¹⁵ Voir annexe 22, "Échantillon de textes sources et de textes cibles", p.359

¹¹⁶ Ces évaluations ont été réalisées en ligne, URL : <https://www.ohmybox.info/linguistics/fr>

”

L'Europe connaissait une fin de séance en baisse aujourd'hui.

En baisse à la clôture de la séance, le Footsie britannique affichait une perte de 0,43% à 7.458,28 points. À 5.476,17 points, le CAC 40 parisien a reculé de 0,33% ce mardi, l'indice Footsie a rétrogradé de 0,43% ce jour, à 7.458,28 points tandis que l'indice DAX s'est délesté de 0,6% à 12.356,11 points. L'indice Bel20, quant à lui, a perdu 0,73% à 3.747,22 points.

FIGURE 4.8 – Échantillon de texte généré de manière automatique, prélevé le 08/01/2019

”

Les principaux marchés européens majoritairement en baisse en début de séance

Les principaux marchés d'Europe évoluent en repli à l'ouverture de la séance. Le CAC 40 perd 0,03% à 5.844,02 points. À 599,25 points, l'indice AEX cède du terrain, à hauteur de 0,12%. L'indice Bel 20 rétrograde de 0,11% à 3.923,39 points.

Rédigé par Quotebot le 11/12/2019 à 09h01

FIGURE 4.9 – Échantillon de texte généré de manière automatique, prélevé le 11/12/2019

Quatre vagues d'évaluation se sont déroulées tout au long du développement du projet¹¹⁷ :

- **Vague 1, août 2018.** À partir d'un échantillon de 21 textes prélevés de manière aléatoire dans les documents relatifs aux deux premières versions des textes générés par le moteur de rédaction de Syllabs (32 combinaisons de textes sources et de textes cibles).
- **Vague 2, septembre 2018.** À partir d'un échantillon de 22 textes, soit l'ensemble du corpus fourni par Syllabs dans une troisième version de textes (35 combinaisons).
- **Vague 3, janvier 2019.** À partir d'un échantillon de 55 textes prélevés de manière aléatoire via l'interface de démonstration de Syllabs, accessible en ligne en accès protégé¹¹⁸ (55 combinaisons).
- **Vague 4, décembre 2019.** À partir d'un échantillon de 14 textes prélevés par un journaliste du service "Investir" dans le système de gestion de contenus de *L'Echo*, dans lequel parviennent les productions de Syllabs, au cours d'une seule journée (13 combinaisons).

Le score Flesch-Kincaid (Figure 4.10) est, en moyenne pour l'ensemble du corpus, de 63,33 pour les textes rédigés par le moteur de rédaction de Syllabs (textes cibles). Ce score est légèrement supérieur à celui engrangé, pour l'ensemble du corpus, par les textes rédigés par les journalistes, qui est de 62,34. Pour autant, il ne s'agit pas d'une très grande différence. La lisi-

¹¹⁷ Voir "Évaluations automatiques", p.231

¹¹⁸ Voir annexe 21, "Interface de démonstration de Syllabs", p.358

bilité des textes sources et des textes cibles peut être considérée comme satisfaisante et plutôt concordante, bien qu'une plus grande stabilité des valeurs soit observée en ce qui concerne les textes rédigés de manière automatique. Dans les deux cas, les valeurs médianes sont assez proches des valeurs moyennes.

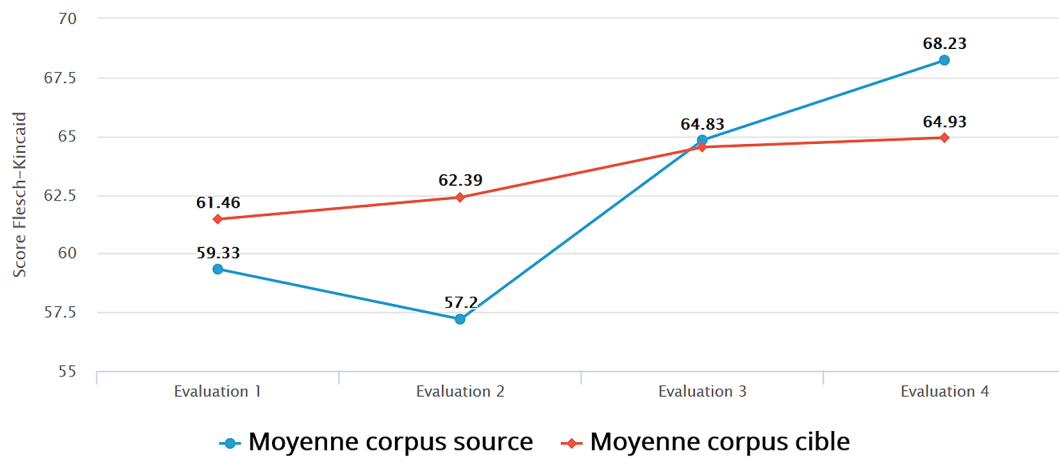


FIGURE 4.10 – Comparaison du score Flesch-Kincaid, relatif à la lisibilité des textes sources et des textes cibles

Les résultats obtenus pour l'Index Coleman Yau (Figure 4.11) témoignent d'un degré d'intelligibilité correct dans les deux cas, et cela compte tenu du fait qu'il s'agisse de textes courts qui comportent des données chiffrées. Cet indicateur est, en moyenne, plus élevé en ce qui concerne les textes cibles (20,3 de moyenne, contre 16,67 pour les textes rédigés par les journalistes du service "Investir"). On observe peu d'évolution dans le temps en ce qui concerne les textes cibles, tandis que les variations sont plus marquées, d'une évaluation à l'autre, en ce qui concerne les textes sources. Dans les deux cas, les valeurs médianes sont assez proches des valeurs moyennes.

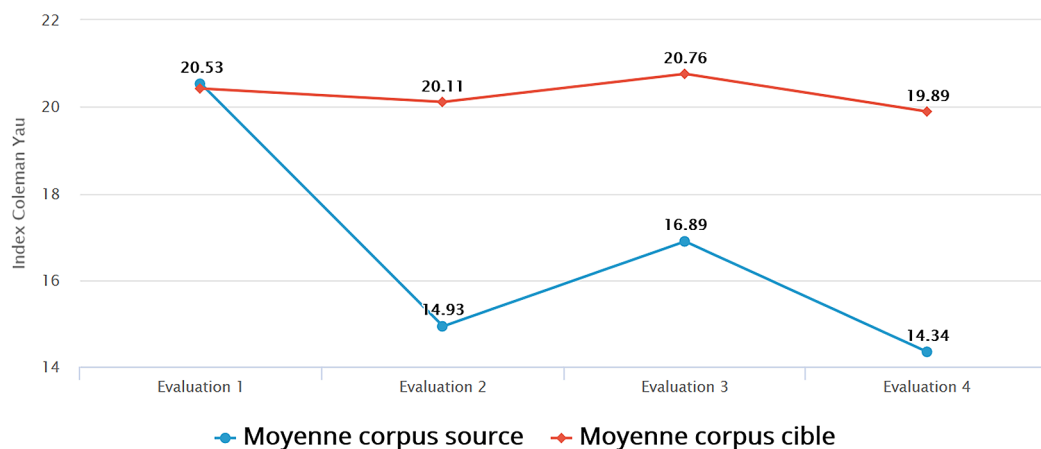


FIGURE 4.11 – Comparaison de l'index Coleman Yau, relatif au degré d'intelligibilité des textes sources et des textes cibles

En ce qui concerne le grade Flesch-Kincaid (Figure 4.12), les textes rédigés par des journalistes correspondent généralement à un niveau d'éducation plus élevé (le grade de 8,14 correspond à un niveau d'éducation "grand public") que les textes rédigés de manière automatique (6,82). Dans la troisième vague d'évaluation, 4 textes du corpus cible et 5 textes du corpus source dépassent le seuil de 10, qui correspond à un niveau d'éducation supérieur (soit 7,2% et 21,74%). Dans les deux cas, les valeurs médianes sont assez proches des valeurs moyennes.

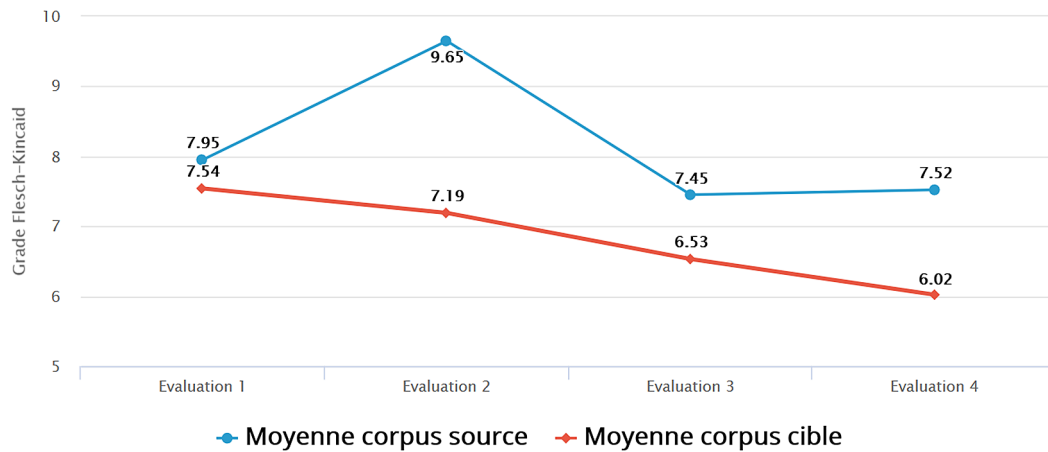


FIGURE 4.12 – Comparaison du grade Flesch-Kincaid, relatif au niveau d'éducation nécessaire pour comprendre les textes sources et les textes cibles

Les résultats obtenus pour l'index ARI (Automated Readability Index, Figure 4.13) correspondent entre les textes sources et les textes cibles, avec un score moyen respectif de 11,76 et de 12,08 (et des valeurs médianes assez proches). Qu'ils soient générés de manière automatique ou rédigés par des journalistes, les textes sont plutôt destinés à un public d'adultes, ce qui correspond aux audiences de *L'Echo*.

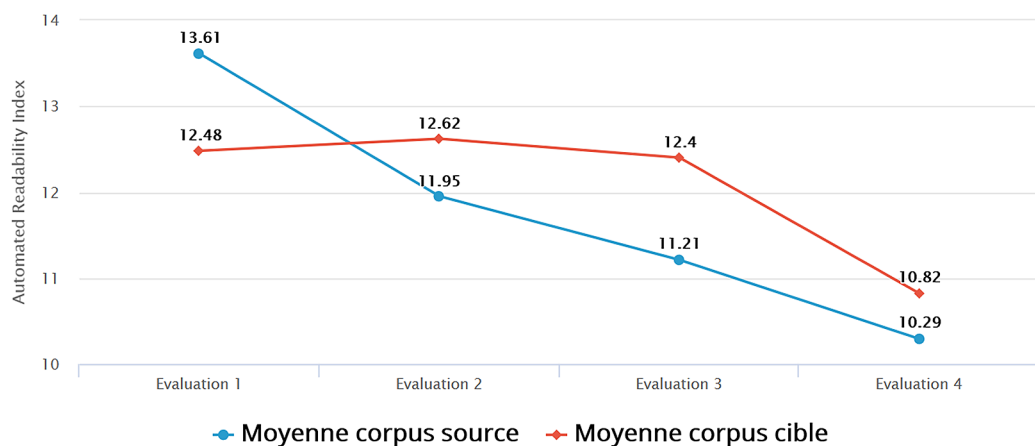


FIGURE 4.13 – Comparaison de l'Automated Readability Index, relatif au degré d'éducation nécessaire pour comprendre les textes sources et les textes cibles

Si l'on compare la moyenne générale des résultats enregistrés par les textes rédigés par des journalistes à ceux générés de manière automatique, on constate généralement de meilleures performances pour ces derniers (Figure 4.14), bien que les différences soient peu marquées en ce qui concerne la lisibilité (Score Flesch-Kincaid) et le niveau d'éducation (ARI). Les textes cibles seraient un peu plus intelligibles (Index Coleman Yau), tandis que les textes sources s'adresseraient à un public un peu plus éduqué (Grade Flesch-Kincaid).

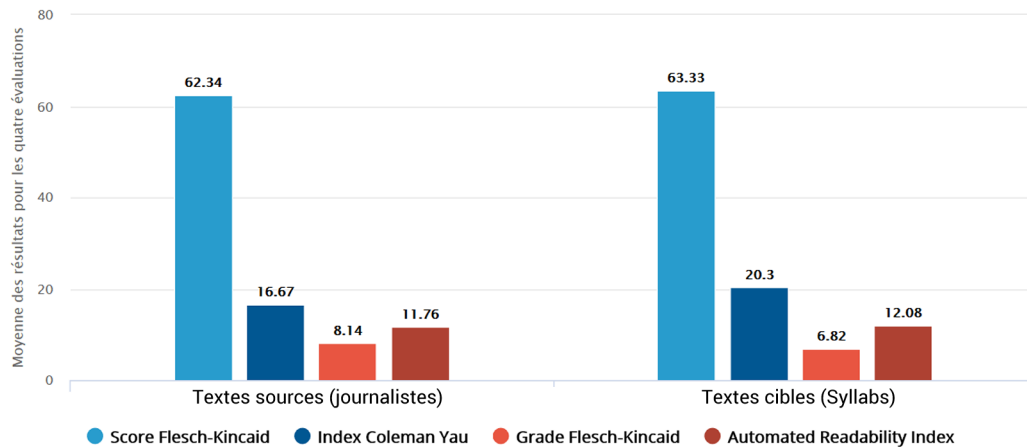


FIGURE 4.14 – Moyenne des indicateurs automatiques pour les textes sources et les textes cibles

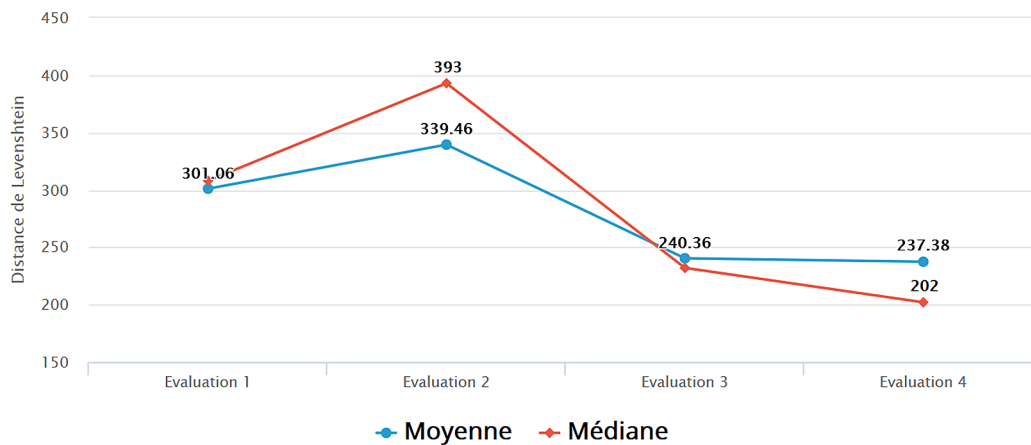


FIGURE 4.15 – Distance de Levenshtein entre les textes sources et les textes cibles

Les observations relatives à la distance de Levenshtein (Figure 4.15) montrent une évolution positive qui s'amorce lors de la troisième vague d'évaluation, la distance moyenne ayant tendance à s'amenuiser (de 301 à 240). Considérant qu'il s'agit de textes courts (entre 200 et 600 caractères), cet indicateur ne peut amener à conclure que les générations sont conformes, dans tous les cas, aux textes fournis par les journalistes. Une raison possible consiste dans la variation des textes générés pour un seul et même phénomène observé, ainsi que dans des valeurs ou indices boursiers ayant pu varier d'un texte à un autre. Une autre explication peut être envisagée dans la manière dont a été conçu le moteur de rédaction : s'il est adapté à chaque nou-

veau projet, il va potentiellement suivre de mêmes procédures. Cela étant, la progression observée peut laisser penser que les textes générés ont été améliorés au fur et à mesure de leurs versions, de manière à mieux respecter les gabarits de textes rédigés par les journalistes ou à correspondre aux demandes de correction formulées par ces derniers.

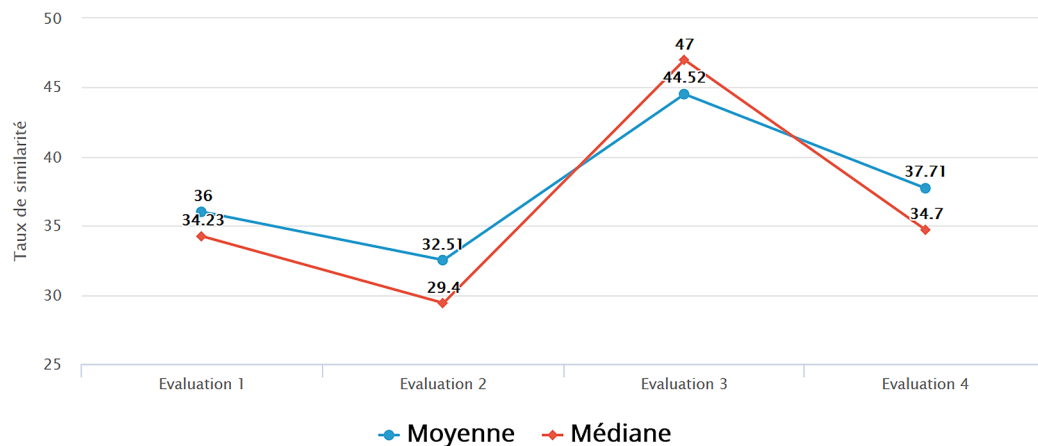


FIGURE 4.16 – Évolution du taux de similarité entre les textes sources et les textes cibles

L'observation de l'évolution de taux de similarité (Figure 4.16) témoigne du peu de concordances entre les textes générés et les textes définis par les journalistes lors de la conception de "Quotebot". Ce taux est passé de 36% lors de la première vague d'évaluation à 44,52% lors de la troisième, pour redescendre à 37,71% à la quatrième. Si l'on observe les échantillons ayant réalisé plus de 50% de taux de similarité, ils sont au nombre de 15,6% dans la première vague, de 11,4% dans la deuxième, de 43,63% dans la troisième, et de 15,38% dans la dernière. Les correspondances les plus élevées ont été constatées dans les textes les plus courts (moins de 100 signes). Toutefois, les textes rédigés par les journalistes au départ de l'expérience ont pu évoluer, au fur et à mesure de leurs corrections. De plus, les textes générés de manière automatique ont été conçus pour présenter un minimum de variabilité lexicale. Si, sur un plan théorique, il donne des indications pouvant paraître *a priori* intéressantes, il ne s'agissait sans doute pas de l'indicateur le plus pertinent dans le contexte de cette expérience.

4.2.4.3 Évaluations basées sur des jugement humain

Un premier questionnaire d'évaluation a été soumis aux six journalistes du service "Investir" ainsi qu'aux trois journalistes en charge de l'édition web en soirée. Les réponses ont été enregistrées entre le 27/02 et le 13/03/2019, soit après que les journalistes aient pris connaissance des premiers échantillons fournis par Syllabs. Sept personnes ont participé à cette évaluation. Parmi celles-ci, cinq disent avoir participé au processus de conception de "Quotebot".

Q1. Pourquoi ont-ils participé à la conception de "Quotebot" ?

Deux évaluateurs n'ont pas participé à la conception de "Quotebot", l'un parce qu'il n'était pas présent à la rédaction, l'autre parce qu'il n'y avait pas été convié. Parmi ceux qui y ont participé,

aucun d'entre eux ne s'est senti contraint. Leurs motivations résident, pour une large majorité, dans leur intérêt pour le projet (quatre évaluateurs sur cinq). Deux évaluateurs sur cinq étaient également motivés par la curiosité, tandis que deux autres ont trouvé ce projet amusant.

Q2. "Quotebot" répond-t-il aux attentes des journalistes?

À ce stade de la conception de "Quotebot", la réponse est "Oui" pour la majorité des évaluateurs (quatre sur sept). Parmi les deux évaluateurs n'ayant pas participé au processus de conception, un seul a estimé que "Quotebot" répondait à ses attentes. Ce qui signifie que trois répondants sur les cinq qui ont participé à la conception de "Quotebot" estiment que ce projet rencontre bien leurs attentes. Ceux-ci estiment que c'est parce que le dispositif leur fera gagner du temps (3 répondants) et que les textes pourront être publiés tels quels (un répondant). Un évaluateur a fait état d'une raison non listée dans les propositions : "*Quotebot devrait nous aider dans notre travail (à travers les hit-parades et les alertes)*". La seule raison évoquée pour expliquer en quoi "Quotebot" ne répond pas aux attentes des journalistes (trois évaluateurs sur trois) est liée au fait que les textes nécessitent des retraitements et que cela ira plus vite de les rédiger soi-même. Les avis sont donc partagés quant à l'efficacité du système de génération automatique.

3. "Quotebot" écrit-il comme eux?

Les avis sont unanimes : l'ensemble des évaluateurs estime que "Quotebot" n'écrit pas comme eux. Alors que cinq d'entre eux ont participé à la conception et/ou à la rédaction des templates devant servir de base aux textes générés, ils n'estiment donc pas que les exigences posées ont été rencontrées. Lors de la première évaluation relative au calcul des indices de similarité entre les textes rédigés par les journalistes et les productions du moteur de rédaction de Syllabs, il avait été observé que le taux de similarité réalisé pour les productions soumises à l'évaluation des journalistes était de 44,52%, avec une distance de Levenshtein moyenne de 240, pour des textes qui oscillent entre 200 et 600 caractères. La différence relevée dans les évaluations automatiques correspondrait donc avec les jugements humains. Toutefois, à l'examen des raisons mises en avant par les répondants pour expliquer en quoi ils estiment que "Quotebot" n'écrit pas comme eux, il apparaît que les résultats des évaluations automatiques ne sont pas corrélées par les journalistes.

Les raisons évoquées sont que "Quotebot" :

- commet des erreurs (quatre répondants), mais un évaluateur précise que "*Ce sont des erreurs liées à la première phase de test du projet*";
- est trop standardisé (trois répondants);
- est trop répétitif (un répondant).

4. Comment les journalistes perçoivent-ils la qualité des productions de "Quotebot"?

Sur les sept variables (ou indicateurs de qualité) qui étaient proposées aux journalistes, celles du sens et de la lisibilité obtiennent les scores les plus élevés : leur moyenne est de 4 sur 5. Les

variables relatives à l'exactitude, la cohérence, la fiabilité, l'utilisabilité, l'utilité et la clarté se positionnent de manière mitigée, avec un score moyen de 3 sur 5 (Figure 4.17).

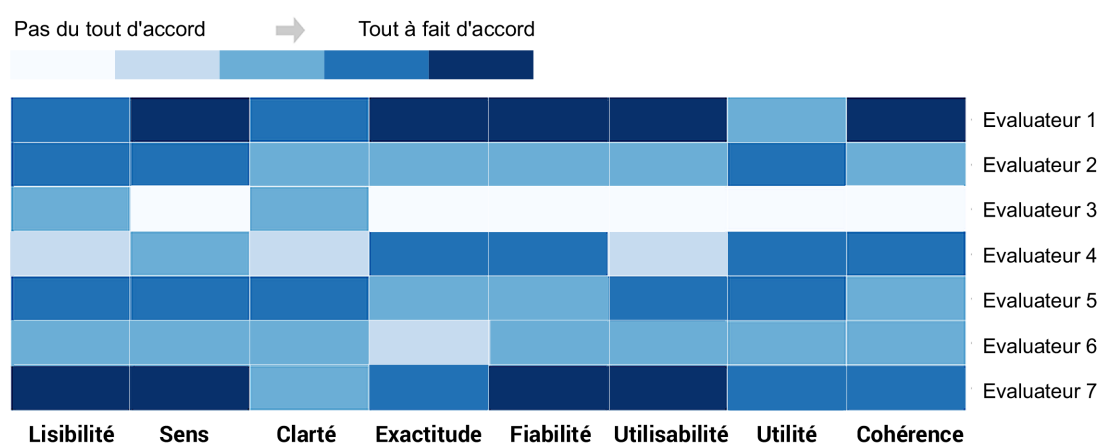


FIGURE 4.17 – Évaluation des productions de "Quotebot"

À un niveau plus granulaire en ce qui concerne les scores les plus élevés, il est observé que ceux-ci portent davantage sur les variables de sens, de lisibilité et d'utilité (4 répondants se disant "d'accord" ou "tout à fait d'accord"); tandis que les scores les plus bas sont plutôt relatifs aux indicateurs d'exactitude et d'utilisabilité. Toutefois, les avis des journalistes apparaissent comme partagés, avec un positionnement généralement moyen au regard des variables examinées.

5. "Quotebot" doit-il être amélioré?

Tous les évaluateurs estiment que le système d'automatisation doit être amélioré, et ils ont commenté leurs réponses :

- "Les productions de "Quotebot" sont encore très simples et il existe une grande quantité de données qui pourraient encore donner lieu à des productions plus poussées de "Quotebot"."
- "Le projet n'est pas encore totalement terminé. Les remarques de la première phase de test n'ont pas encore été totalement prises en compte. On est plus très loin toutefois de l'aboutissement du projet."
- "Les données utilisées sont incorrectes. D'où les erreurs."
- "Génération de textes plus agréables à lire et fiabilité des chiffres donnés."
- "Il commet parfois des erreurs et le délai pour réagir à un événement est trop long."
- "Ex. Le Bel 20 : -0,02%. Peut-on dire pour autant qu'il est en baisse?"
- "Enrichir son vocabulaire, notamment en termes de langage 'plus familier'."

Par ailleurs, deux commentaires libres ont été déposés :

- "Un investissement important en temps et en argent a été nécessaire pour développer des productions "Quotebot" relativement simples. Il s'agit toutefois d'un investissement non récurrent, tandis que les productions de "Quotebot" devraient être utilisables quasiment

indéfiniment. Si des améliorations doivent être apportées à "Quotebot", notamment pour approfondir ses productions en y intégrant davantage de données, cela nécessitera vraisemblablement un investissement en temps et en argent encore plus élevé, mais toujours avec la perspective de pouvoir ensuite bénéficier de ces productions sans véritable limite dans le temps : ces investissements sont donc amortissables à très long terme. La question qui peut se poser est la suivante : est-ce le travail des journalistes de participer à un tel projet qui nécessite un investissement aussi élevé ou ne faudrait-il pas consacrer les moyens financiers dégagés pour ce type de projet pour envisager l'acquisition d'une "fintech"¹¹⁹ qui fera ce travail plus vite et mieux? Mais on entre là dans des considérations stratégiques qui relèvent de la responsabilité de l'éditeur auquel je vous renvoie pour une éventuelle réponse à cette question."

— *"Je suis bien curieux de voir le projet mis en production permanente, ainsi que les alertes et le portefeuille."*

À ce stade du développement du projet, les avis sont partagés quant à la qualité de "Quotebot". On observe qu'il y a unanimité pour conclure que "Quotebot" ne rédige pas comme un journaliste de la rédaction et qu'il doit encore être amélioré. Il est également souligné l'intérêt de disposer d'un système d'alerte en cas d'observations d'évolution anormale de valeurs, alors que ce volet du projet sera finalement abandonné pour des raisons de planning. Si "Quotebot" apparaît comme potentiellement utile, en l'état, il n'est pas encore utilisable. Les journalistes réclament davantage de qualité, que ce soit sur le plan de l'écriture ou sur celui de la précision dans le traitement des données. Si l'on compare les évaluations automatiques aux évaluations basées sur des jugements humains, les différences entre les textes sources et les textes cibles sont confirmées mais elles sont ici nuancées en termes d'exactitude et de fiabilité. À ce stade du projet, "Quotebot" ne rencontre donc pas pleinement le critère d'adéquation de ses productions avec les savoir-faire et exigences journalistiques.

Un deuxième questionnaire d'évaluation a été soumis aux mêmes journalistes précédemment sollicités. Les réponses ont été enregistrées entre le 21/12/2019 et le 03/01/2020, soit avant la dernière phase de test de "Quotebot". Quatre journalistes y ont répondu. Les mêmes questions leur ont été posées, à la différence qu'il leur était demandé s'ils avaient pris connaissance des dernières productions de "Quotebot". Trois journalistes ont répondu par l'affirmative, le dernier répondant indiquant qu'il n'en a pas eu le temps. Leurs avis étaient partagés quant à la qualité du dispositif d'automatisation : un seul journaliste a estimé que "Quotebot" a été amélioré et que ses productions utilisables. "Quotebot" correspond aux attentes de ce même répondant, qui a estimé que les productions automatisées lui feront gagner du temps. Pour les autres répondants, les textes ne sont pas exploitables en l'état et nécessitent un retraitement. Aucun d'entre eux n'estime que "Quotebot" écrit comme eux. Les raisons sont à trouver dans

¹¹⁹ Contraction de finance et technologie, le terme "Fintech" désigne des entreprises innovantes, "*utilisant les technologies du numérique, du mobile, de l'intelligence artificielle, etc., pour fournir des services financiers de façon plus efficace et moins chère*", définition proposée par le journal *La Tribune*, "C'est quoi une Fintech?", Delphine Curry, 07/04/2017, consulté le 13/02/2020, URL : <https://www.latribune.fr/entreprises-finance/banques-finance/c-est-quoi-une-fintech-680118.html>

sa standardisation (1 répondant), et dans le fait qu'il commette encore des erreurs (2 répondants). Les deux autres raisons évoquées sont : "*Le texte produit n'est pas toujours top*" et "*La base de données de textes n'a pas été suffisamment bien réalisée*". L'examen des sept indicateurs de qualité montre que les valeurs moyennes sont moins bonnes pour cette seconde évaluation (Figure 4.18) : selon les évaluateurs, les productions de "Quotebot" font moins sens, et sont moins lisibles, fiables et utilisables. Il convient d'indiquer qu'un journaliste a tiré les résultats à la hausse, attribuant généralement de meilleurs scores à chacun de ces indicateurs. Les avis sont donc partagés quant à une éventuelle utilisation finale des productions de "Quotebot", telles qu'elles sont délivrées à ce stade du développement de l'outil.

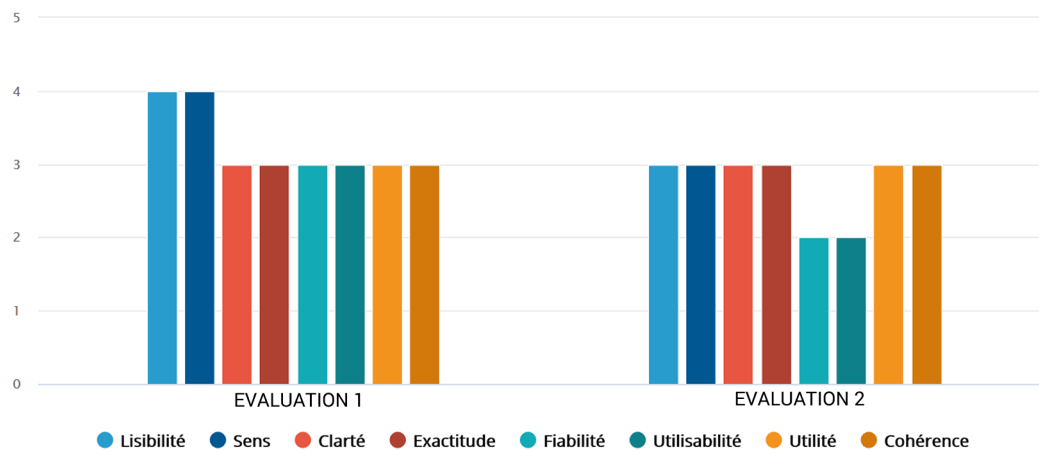


FIGURE 4.18 – Comparaison des deux évaluations des journalistes à propos de la qualité des productions de "Quotebot"

Les journalistes avaient également la possibilité de déposer un commentaire libre. Un premier a précisé que : "*Questionnaire complété avant la phase finale de tests*"; tandis qu'un autre a indiqué que : "*À la fin de la réalisation du projet, il est apparu que certaines données fondamentales de 'Quotebot' (parties de phrases utilisées dans le programme) n'étaient pas correctement formulées ou articulées avec les données chiffrées, si bien qu'un retravail de l'ensemble des données était nécessaire, alors que le budget alloué au projet était déjà épuisé. La complexité de la réalisation de ce projet semble avoir été sous-estimée. En l'état, les articles rédigés par Quotebot ne sont donc ni fiables, ni utilisables, selon moi, car ils devraient être irréprochables alors que, dans l'état actuel d'achèvement du projet, il subsiste des risques d'erreurs, de mauvaises formulations, d'aberrations, etc. 'Quotebot' ne peut donc pas (encore?) être utilisé à ce stade*".

Les résultats de cette deuxième évaluation basée sur des jugements humains ne permettent pas de conclure que les journalistes aient généralement perçu une amélioration dans les contenus générés de manière automatique, et cela après neuf mois de tests et de corrections. Ceci ne corrèle pas avec les évaluations automatiques des derniers échantillons de textes, où il avait été constaté relativement moins de différences entre les textes sources et les textes cibles. Du point de vue des journalistes, cela ne témoigne pas de la capacité du moteur de rédaction de Syllabs à pouvoir être adapté à leurs demandes précises, en termes de traduction de leur exper-

tise et de leurs exigences dans la configuration du logiciel. Ce constat doit être pondéré par les problèmes de qualité des données rencontrés à différents moments du développement du moteur de rédaction, qui peuvent également être un facteur amenant les journalistes à considérer que les productions de "Quotebot" ne sont pas fiables. Cela étant, de nouvelles (et dernières) générations de textes devaient encore faire l'objet de nouveaux (et derniers) tests, qui n'étaient pas encore réalisés au moment de clôturer cette thèse.

4.3 Analyse des observations

"Quotebot" s'est déployé dans un service où l'on traite habituellement d'actualités boursières, et donc de faits chiffrés, et où la manipulation des logiciels informatiques est bien ancrée dans les pratiques professionnelles. Les journalistes envisagent ce projet comme un outil supplémentaire ou comme une source d'informations s'ajoutant à un arsenal existant. Les exigences de sa conception sont celles que vont poser les journalistes tout au long de son processus de développement. S'agissant d'usages finaux dans lesquels les utilisateurs de "Quotebot" sont invités à se projeter, l'expression de leurs besoins va prendre différentes formes quant aux fonctionnalités attendues de l'outil : rédaction de textes relatifs aux temps forts d'une journée boursière, génération de graphiques ou de tableaux permettant de visualiser les valeurs ayant enregistré les plus fortes hausses ou les plus fortes baisses, prise en charge automatisée de l'historique des valeurs boursières, et alertes destinées à attirer l'attention des journalistes sur les valeurs se comportant de manière inhabituelle. Ces besoins seront traités selon un ordre de priorités que les journalistes ont eux-mêmes défini. Le système d'alertes, pourtant considéré comme important, sera abandonné en cours de route pour des raisons de calendrier. Une accumulation de retards dans la mise en production de l'artefact n'a donc pas permis de satisfaire l'ensemble de ces besoins et de tenir le délai de développement initialement imparti au projet. Deux ans après le lancement "*kick off*" du projet, "Quotebot" était toujours en cours de développement, une poignée de mois le séparant alors de sa mise en service officielle.

À l'instar de la première étude de cas, la mise en perspective du cycle de vie du logiciel avec le modèle SCOT a permis de suivre le développement de l'expérience tant sur son plan social que technique, dans un processus itératif où les étapes relatives au design et à l'implémentation n'ont pas été forcément entreprises de manière distincte entre les agents sociaux impliqués dans le monde du journalisme et dans celui de la technique. La phase de flexibilité adaptative, qui correspond au paramétrage du moteur de rédaction, aura duré un peu moins d'un an, la difficulté se situant au niveau de la récolte de la "bonne donnée" au "bon moment". Le processus de tests et de corrections – qui peut être considéré comme un processus de négociation s'inscrivant dans le cadre de la phase stabilisation de l'artefact technologique – fut particulièrement long, dès lors qu'il s'est échelonné sur une année.

Dans l'esprit d'une adéquation aux usages, l'intégration des journalistes dans le processus de conception peut être considérée comme un moyen de parvenir à cette fin, étant supposé qu'un ajustement entre concepteurs et utilisateurs constitue un préalable à un usage effectif. Pour autant, les usages finaux ne peuvent pas forcément être déduits mécaniquement d'un proces-

sus de conception (Akrich 1990), et cela d'autant plus que les résultats ne correspondent pas aux attentes des journalistes. Au moment de clôturer cette thèse, la stabilisation de l'artefact n'est pas constatée et ses usages opérationnels demeurent potentiellement incertains, en raison d'exigences qualitatives – tant sur le plan des données que sur celui des textes générés – qui ne sont pas encore rencontrées. La principale explication de cet état de fait serait à trouver dans le manque d'expertise, en matière d'informations boursières, de la société en charge du développement de "Quotebot", qui ne s'est pas trouvée en capacité, à ce stade, de fournir des productions automatisées qui satisfassent l'ensemble des journalistes. Cela étant, les erreurs rencontrées dans les générations de contenus ne sont pas toutes imputables à Syllabs. Aussi, le projet "Quotebot" a-t-il été l'occasion de se rendre compte de problèmes de qualité dans la fourniture des données, alors que celles-ci font l'objet d'un contrat payant et que, de ce fait, leur qualité est normalement assurée. Le manque de fraîcheur de certaines données est également apparu comme problématique dans le contexte de la couverture d'une information en temps réel.

Si "Quotebot" a bien pour objet de soutenir des routines journalistiques, sa conception a induit une reconfiguration des pratiques éditoriales, ici pensées en termes de processus à automatiser. Dans ce projet, il ne s'agissait pas seulement de paramétrer un moteur de rédaction en fonction de choix éditoriaux et d'exigences journalistiques – ce qui correspond à la définition du "quoi dire" et du "comment le dire" –, mais aussi d'être en mesure de les médier et de les traduire conformément aux intentions des journalistes. Bien que les choix posés ne se distinguent pas, sur un plan conceptuel, d'un processus éditorial classique, cela a impliqué de nouvelles formes de travail pour les journalistes. En premier lieu, la rédaction des gabarits de texte (ou "templates") a supposé que les journalistes "déconstruisent" leur propre manière d'écrire pour en proposer des versions standardisées qui conviennent aux caractéristiques d'une production automatisée. Ils ont également dû réfléchir aux possibilités de variation de chacun des gabarits qu'ils ont définis, de même qu'ils ont dû envisager la manière dont l'évolution des valeurs allait influencer sur la mise en récit. Aussi, les journalistes ont-ils développé une nouvelle forme de savoir-faire. Ceci témoigne de la modification de la nature des rôles, tâches et flux de travail induite par le développement de la production automatisée d'informations dans les rédactions (Diakopoulos 2019). De plus, à l'inverse de ce qui avait été constaté dans la première étude de cas, les journalistes n'ont éprouvé aucun mal à se projeter dans leurs usages finaux, ces derniers ayant d'emblée perçu les avantages relatifs du dispositif (un gain de temps dans des routines professionnelles s'opérant en flux tendu et une source d'informations complétant utilement un arsenal existant). Celui-ci rencontre leur adhésion dès les débuts du projet ("*On va la faire travailler cette machine*"), mais celle-ci s'étiolera au fil du temps ("*En l'état, les articles rédigés par Quotebot ne sont donc ni fiables, ni utilisables, selon moi, car ils devraient être irréprochables*").

En second lieu, l'implication de certains journalistes dans le processus de correction s'est avérée bien plus fastidieuse que prévu, en témoignent les nombreuses erreurs qu'ils ont relevé dans un mouvement d'aller-retours avec Syllabs. Pour un journaliste impliqué dans ce processus, "*la difficulté, c'est que c'est un travail qui vient en plus d'un travail quotidien, et (lorsque l'on*

suit) les marchés en temps réel, (il faut) être constamment aux aguets. Travailler sur un projet comme ça, ça demande de se concentrer et de se focaliser sur la matière, donc c'est deux rythmes de travail différents. Comme on a un effectif limité, j'ai toujours travaillé sur 'Quotebot' un peu à la marge, et donc ce n'était pas évident parfois (...) Ce qui est assez chouette c'est que tu développes et crée quelque chose, ça change un peu de la routine du Market Live"¹²⁰. Ceci tranche avec de précédentes études où les journalistes n'ont pas témoigné d'une expérience particulièrement agréable lorsqu'ils étaient amenés à "former" des algorithmes pour choisir le phrasé le plus approprié (Linden 2017).

Toutefois, en ce qui concerne un autre journaliste, la charge de travail induite par ce processus a agi en tant qu'élément déclencheur d'une crise ayant eu pour conséquence la paralysie du projet pendant plusieurs mois. Cette contestation, qualifiée de violente, se trouvait plus largement dirigée contre l'organisation générale de l'entreprise de presse et son management, puisque la remise en cause portait également contre la rédaction en chef, qui a changé en cours de projet. Ce type d'attitude face à l'introduction d'une innovation technologique dans une rédaction n'est pas neuf. Cela fut notamment observé au début des années 2000, dans le contexte du développement de politiques de convergence, où les résistances exprimées par les journalistes ont été expliquées de manière multiple : qu'elles soient dirigées contre une redistribution des rôles au sein des rédactions, contre des modifications organisationnelles et/ou du processus de production de l'information, ou contre la direction et le management des entreprises de presse (Garrisson 2001, cité par Deuze 2008). Précisons que le journaliste à l'origine de la crise qui entraîna le gel du projet "Quotebot" est celui qui avait lancé, sur le ton de la boutade lors d'une réunion de travail, que la société en charge du développement de l'artefact se posait en potentielle concurrente des journalistes ("*C'est eux qui vont nous remplacer? On les aide à nous remplacer?*").

Du point de vue des études des logiciels, un système de production automatisée de contenus journalistiques rencontre les cinq caractéristiques d'un "nouveau média" (Manovich 2010) : (1) il fournit une représentation numérique des informations; (2) sa modularité est symbolisée par la séparation entre les bases de données informatisées fournies en "entrée" et le code informatique qui permettra de générer les "sorties"; (3) le niveau de variabilité du système dépend des valeurs des données; (4) ses processus sont automatisés et résultent d'un transcodage culturel alimenté par des expériences éditoriales antérieures; (5) l'automatisation est rendue possible par la conjugaison du codage numérique et de la structure modulaire de l'artefact. Ce faisant, il peut donc être considéré comme une forme remixée d'une forme médiatique préexistante (Manovich 2013). Les productions de "Quotebot" s'inscrivent dans un projet éditorial qui est également remixé, celui-ci étant caractérisé par un partage des tâches entre des agents évoluant dans deux mondes sociaux distincts, lesquels répondent à des exigences différentes où la "subjectivité" du journalisme est opposée à la "rationalité" de la technologie. Cette promesse d'objectivité reste néanmoins illusoire, dans la mesure où un programme informatique n'est pas développé en dehors de son contexte social et culturel (Gillespie 2014). De même, une ac-

¹²⁰ Communication personnelle, par téléphone, le 10/09/2019.

tivité éditoriale suppose une série de choix qui ne peuvent être considérés comme neutres, dès lors qu'ils impliquent des décisions humaines.

Dans un projet d'automatisation se concentrant sur l'aspect lié à la prise en charge de tâches répétitives, le journalisme semble ici moins envisagé comme une idéologie professionnelle que comme une activité occupationnelle. Toutefois, en tant que participantes actives aux processus journalistiques, les technologies d'automatisation – et leurs concepteurs – impliquent une nouvelle remise à plat de ce qu'est le journalisme : qui le fait et que fait-il? (Primo & Zago 2015). Ce faisant, elles contribuent au floutage des contours de la profession ou à leur émiettement (Ruellan 1992, 2005). Dès lors que des linguistes et des informaticiens sont impliqués dans un processus éditorial, ils font acte de journalisme mais sans en partager l'idéologie, les savoirs et savoir-faire. Bien que Syllabs exerce une activité éditoriale en agissant comme fournisseur de contenus journalistiques, son discours est d'affirmer ne pas se substituer aux journalistes et encore moins de se revendiquer comme une agence de presse ou un média. De ce fait, elle trace une frontière claire et non équivoque entre les deux mondes, et cette position est partagée par l'ensemble des agents de la société. Pour autant, cette posture reste ambiguë : en témoignent des discours de communication pouvant apparaître comme contradictoires, ou encore la déception exprimée à propos des formats de textes à automatiser, plutôt courts et qui lui laissent peu de marge de manœuvre en termes de "créativité". Par son implication dans une activité éditoriale (et donc également dans la construction du sens journalistique), Syllabs agit en tant qu'actrice à part entière du processus journalistique : reconnaître ceci implique un changement de paradigme. Cela pose, plus largement, deux questions : une première, qui concerne la manière dont il conviendrait d'inclure ces agents du monde de la technique dans celui du journalisme et selon quelles modalités ; une seconde, relative à l'élaboration d'un cadre commun où l'activité technique doit être envisagée, d'abord et avant tout, comme une activité journalistique.

Lorsque les agents du monde du journalisme sont convoqués pour poser les jalons du système d'automatisation, ils bouleversent le flux de travail de Syllabs, peu habituée à recevoir des directives aussi précises dans le contexte d'une collaboration avec un média d'information. Bien que son moteur de rédaction n'ait pas été spécifiquement conçu à des fins journalistiques, celui-ci fait l'objet d'une adaptation "sur mesure" lors de chaque nouveau projet. Dans le contexte du développement de "Quotebot", en s'engageant dans un jeu de l'imitation, il va donc potentiellement reproduire et incarner des normes professionnelles journalistiques (Domingo 2008, Anderson 2013). Toutefois, l'expertise éditoriale, de même que celle du domaine d'application, qui relèvent des savoirs et savoir-faire des journalistes, n'ont pas pu être "imitées" parfaitement par le moteur de rédaction. La correspondance des productions automatisées avec les exigences journalistiques apparaît, ici, comme un préalable aux usages. Celles-ci portent tant sur la forme (utilisation d'un vocabulaire et d'expressions spécifiques) que sur le fond (précision des valeurs et capacité à rendre compte d'une information boursière). Les dernières évaluations des productions de "Quotebot" ne permettent pas, à ce stade, d'en envisager des usages finaux, dès lors que celles-ci ne répondent pas aux critères de fiabilité et d'utilisabi-

lité posés par les journalistes. Bien qu'ils ne partagent pas tous ce point de vue, celui-ci semble pourtant dominer.

Lorsque les journalistes sont invités à évaluer le dispositif, ils sous-entendent que les erreurs commises par le système ne permettront pas de l'utiliser et qu'il sera plus rapide de traiter eux-mêmes les données. Ils sont également unanimes à considérer que "Quotebot" n'écrit pas comme eux... précisément en raison des erreurs commises et d'un manque de fiabilité qu'ils attribuent à ses productions. Le manque d'expertise de Syllabs en matière d'information économique et financière a été considéré comme l'un des facteurs ayant ralenti le projet car ce qui paraissait évident aux yeux des journalistes ne l'était pas forcément du côté du prestataire de services : *"C'est ça qui prend du temps (...) Les personnes chez Syllabs ne connaissent pas vraiment bien les marchés boursiers, donc il y avait certaines choses chez nous qui paraissaient évidentes et qu'ils ne comprenaient pas ou alors qu'ils les traitaient de manière erronée"*, indique un journaliste¹²¹ ; *"Le manque de connaissance de la matière, malgré les efforts, est problématique. Il a fallu faire œuvre de pédagogie (...) Si c'était à refaire, on aurait organisé une session de deux jours sur le fonctionnement des marchés financiers, de manière à disposer d'une base de connaissance commune qui permette aussi à Syllabs (...) de moins appliquer les choses de manière moins littérale"*, explique le responsable du projet à *L'Echo*¹²².

Sur le plan des représentations, les imaginaires technologiques des journalistes peuvent être considérés comme plutôt ouverts au principe d'une automatisation. Lors des premières réunions de travail, ils ont témoigné d'attitudes enthousiastes, percevant d'emblée les bénéfices potentiels du dispositif. Leurs imaginaires ont été davantage mobilisés dans la manière dont ils ont pu se projeter dans leurs usages finaux, sans pour autant abandonner au système d'information la mise en sens journalistique : celle-ci relève de leur domaine, en décidant de publier un contenu tel quel, de l'enrichir de leur expertise avant publication, ou de ne pas le publier du tout. En ce sens, l'autonomie du système ne lui est que partiellement reconnue. Il s'agit également de favoriser l'expérience utilisateur en s'adaptant à son expérience habituelle, dès lors que les productions de "Quotebot" apparaissent directement dans le système de gestion de contenus que les journalistes utilisent au quotidien. Par ailleurs, un éventuel rapport affectif aux productions générées de manière automatiques – voire à l'outil en lui-même – a émergé dans le contexte du départ à la retraite d'un journaliste. Cela a été nuancé par un autre journaliste, qui estime que celles-ci renferment également un part de lui *"mais pas l'essentiel, c'est surtout le travail de la société en charge du développement de 'Quotebot'"*¹²³.

Dans un contexte professionnel où la métaphore du robot est susceptible d'induire un rapport anxigène à l'artefact¹²⁴, d'éventuelles résistances avaient été envisagées à l'aube du projet. C'est pourquoi la question de la communication autour de "Quotebot" a été appréhendée avec prudence : *"On sait que cela pourrait freiner le projet, voire de l'arrêter comme cela a été*

¹²¹ Communication personnelle, par téléphone, le 10/09/2019.

¹²² Communication personnelle, Bruxelles, le 19/12/2019.

¹²³ Communication personnelle, par téléphone, le 10/09/2019.

¹²⁴ Voir " Imaginaires du 'robot journalisme'", p.77

le cas dans une rédaction économique française. On s'est dit qu'une communication générale allait peut-être donner une importance trop grande à un phénomène isolé au sein de la rédaction. Cela concerne des contenus qui n'en remplacent pas d'autres mais qui viennent les compléter, les enrichir ou aider les journalistes. (...) La piste envisagée a été celle d'en parler à la société des journalistes, en lui expliquant que cela ne vise pas à remplacer les journalistes (...) (dans leurs) analyses des marchés", explique le responsable du développement et de la transformation numérique de L'Echo¹²⁵. Sur le site web de L'Echo, il n'est pas fait appel à la métaphore du robot pour présenter le projet, même si celui-ci est décrit comme un " rédacteur d'un nouveau genre" : "Comment libérer les journalistes des tâches répétitives et chronophages pour qu'ils puissent se concentrer sur ce qu'ils font le mieux : l'analyse et la mise en contexte? C'est la question à laquelle L'Echo a essayé de répondre en créant 'Quotebot', un générateur automatique de textes qui produit de petites synthèses factuelles à partir de données financières. Un travail conjoint entre les journalistes de la rédaction et Syllabs, une entreprise spécialisée dans la génération automatique de texte. Ensemble, nous avons mis en place des scénarios qui s'adaptent en fonction des variations de la Bourse et sont publiés tout au long de la journée (...) dans le Market Live Blog"¹²⁶.

En tant qu'actantes du processus éditorial, les technologies d'automatisation ne sont pas purement mécaniques : elles se trouvent au centre d'un jeu de médiations et de traductions qui les impliquent activement dans la construction du sens (Latour 2005). Si leurs processus peuvent apparaître comme enchâssés dans une "boîte noire" technologique dont on ne connaîtrait que les entrants et les extrants, la compréhension de la manière dont se sont opérées les interactions au sein de la chaîne sociotechnique ont permis une mise en visibilité partielle de ses processus, mais elle reste insuffisante : le moteur de rédaction développé par Syllabs est le fruit de longues années de recherche et ses mécanismes relèvent d'un évident secret industriel.

Les interactions entre les agents sociaux participant à la chaîne sociotechnique ont été facilitées par la figure d'un agent de médiation, dont le rôle a été fondamental, que ce soit pour clarifier les demandes des journalistes ou pour gérer un planning impliquant une variété d'intervenants. L'une des difficultés fut, pour lui, d'établir un agenda commun alors que le projet ne faisait pas nécessairement l'objet d'un caractère prioritaire, notamment en ce qui concerne les prestataires IT externes à Mediafin qui étaient en charge de l'implémentation des flux d'information transmis par Syllabs dans le système de gestion de contenus de L'Echo. La distance physique entre chacune des parties prenantes au projet a également été perçue comme une difficulté supplémentaire, les communications devant obligatoirement passer par des canaux numériques ou téléphoniques.

Pour ce gestionnaire de projet, qui dit avoir gagné en compétences grâce à cette expérience – en ce compris en matière d'informations boursières –, cela a impliqué de rester à l'écoute des besoins exprimés par les différentes parties prenantes, tout en tenant compte des contraintes administratives, techniques et journalistiques inhérentes au projet. Cette mission – qu'il définit

¹²⁵ Communication personnelle, Bruxelles, le 02/02/2018.

¹²⁶ "Qui est Quotebot?", Nicolas Becquet, L'Echo, 22/02/2019, URL : <https://www.lecho.be/les-marches/homepage/qui-est-quotebot/10100844.html>

lui-même comme celle d'un facilitateur – s'inscrivait dans le champ de ses activités habituelles. Sa double compétence, journalistique et technologique, lui a permis de faire le lien entre deux mondes sociaux aux pratiques et normes distinctes. Il a également participé de manière active à la construction du sens journalistique de l'artefact technologique, et cela d'autant plus qu'il se trouve à l'origine du projet et qu'il a chapeauté l'ensemble du processus de conception et de correction de "Quotebot".

Une autre difficulté fut, pour lui, de faire face à une accumulation de retards dans la mise en œuvre de "Quotebot", alors que "*dans l'ensemble du projet, tout s'est très bien passé. Il a suscité l'adhésion de la rédaction et l'enthousiasme de la direction*"¹²⁷. À l'été 2019, rien ne laissait présager qu'une crise interne allait éclater, celle-ci ayant sans doute mis le doigt "*sur des manquements dans le projet ou sur ses fragilités*"¹²⁸. Mais si "Quotebot" a survécu à cette tempête, un autre obstacle allait encore le ralentir. Cette fois, il s'agissait d'une variable indépendante : celle d'une réorganisation complète du système informatique de l'entreprise, mobilisant les ressources internes au média. C'est la raison de laquelle le projet a été postposé au second trimestre 2020. Aussi, la complexité inhérente au projet "Quotebot" tient-elle dans ses aspects multidimensionnels : à la fois sociaux, techniques et organisationnels.

¹²⁷ Communication personnelle, Bruxelles, le 19/12/2019.

¹²⁸ Communication personnelle, Bruxelles, le 19/12/2019.

Conclusions générales

Le journalisme est une profession dont les conditions d'exercice n'ont cessé de se transformer au cours de ces vingt dernières années, l'introduction d'innovations technologiques dans les rédactions n'ayant pas toujours contribué à ce qu'elles s'en trouvent améliorées. Dans le même temps, alors que les entreprises de presse traversent une crise sans précédents, le climat ambiant d'injonction à l'innovation sous-entend que leur salut ne se trouverait pas ailleurs que dans la réponse technologique. Il s'agissait dans, cette thèse, de regarder le verre à moitié plein, en se plaçant dans une position méta qui permette un recul nécessaire face aux discours les plus pessimistes et les plus optimistes. Si le phénomène de l'automatisation de la production d'informations remet une nouvelle fois en question la place et le rôle des journalistes dans une société de l'information, l'objectif de cette thèse était de comprendre comment celui-ci pouvait être envisagé comme un allié du journalisme. S'intéresser aux mécanismes de la structuration de nouveaux usages professionnels, compte tenu d'une complexité multifactorielle, supposait de s'intéresser aux représentations sociales et culturelles d'un phénomène technique sujet à controverses.

En premier lieu, il est apparu que les imaginaires technologiques liés à ce phénomène sont pluriels. Ils peuvent autant relever de la relation que les journalistes entretiennent en général avec les technologies informatiques, de leur rapport souvent difficile au fait chiffré, et la représentation duale – entre opportunité et menace – véhiculée par l'usage répandu de la métaphore du robot, comme cela a été souligné dans l'analyse du discours médiatique des journalistes. Ces imaginaires induisent un premier rapport à l'objet qui, à la lumière des deux études de cas présentées en deuxième partie de cette thèse, peuvent influencer sur la manière dont celui-ci sera abordé en tant qu'outil du journalisme. Les représentations des journalistes peuvent également être conditionnées par les avantages perçus de l'artefact d'automatisation, par une culture organisationnelle peu ou prou ouverte au phénomène technique en général, ainsi que par une approche par données dans le journalisme déjà ancrée dans les pratiques professionnelles. Trois enjeux principaux apparaissent en filigrane : celui de la gestion des angoisses professionnelles induites par le phénomène de la production automatisée d'informations, dès lors qu'il est aussi envisagé comme une menace sur l'emploi ou l'identité du journaliste ; celui d'un nouvel équilibre à trouver dans une relation homme machine, en termes de complémentarité et de valeur ajoutée du journaliste ; et celui d'une mise à plat des logiques informatisées à l'œuvre dans un artefact d'automatisation de la production d'informations, amenant le journaliste à considérer ses pratiques sous l'angle des processus qui les sous-tendent.

En second lieu, l'examen des dimensions culturelles de l'automatisation de la production d'informations, ici abordée sous l'angle des dimensions de la culture professionnelle du journalisme, met en relief les challenges que doit rencontrer un artefact d'automatisation de la production d'informations lorsqu'il est envisagé en tant qu'outil du journalisme. Ceux-ci sont à

la fois techniques, s'agissant de garantir aux journalistes que les productions automatisées sont précises et fiables – et donc utilisables –, et journalistiques, s'agissant de s'inscrire dans la logique d'un processus éditorial qui rencontre les savoirs et savoir-faire des professionnels de l'information. Dans cette perspective, la problématique de la qualité des données apparaît comme un axe fondamental à traiter en amont du système d'automatisation, en termes de prévention des anomalies et de gestion des erreurs, mais aussi en termes de gestion de la qualité dans le temps. Des problèmes de qualité des données ont été rencontrés dans les deux études de cas et, malgré des caractéristiques qui leur sont propres, leur traitement a supposé une contribution humaine qu'il ne faut pas négliger : sans données de qualité, pas de génération de qualité. Si l'on pouvait s'attendre à de potentiels problèmes lorsqu'il s'agit de données fournies en *open data*, le fait de souscrire un abonnement payant auprès d'un prestataire spécialisé ne garantit pas nécessairement que leur qualité soit assurée.

En supposant qu'aucun problème ne soit rencontré sur le plan de la qualité des données, il apparaît que l'expertise du domaine d'application reste la pierre angulaire présidant à tout processus d'automatisation de contenus. Cette expertise doit nécessairement être partagée entre les agents sociaux en charge du développement de l'artefact d'automatisation et les journalistes. Les observations réalisées dans le cadre de la deuxième étude de cas démontrent qu'une connaissance non partagée a pour conséquence un taux d'erreur important dans les productions automatisées, ce qui ne permet pas aux journalistes de les utiliser. Dans les deux études de cas, il est apparu que les développeurs d'un système d'automatisation doivent également être légitimes aux yeux des journalistes, pour qui les fondamentaux du métier doivent être garantis. Dans la première étude cas, cette légitimité a été attribuée à notre passé professionnel de journaliste tandis, que dans la deuxième étude de cas, celle-ci n'a pas été reconnue au prestataire de services auquel il a toutefois été reconnu l'expertise technique. C'est pourquoi nous soutenons que si les journalistes devraient désormais développer une forme de "pensée computationnelle", en vue de favoriser les dialogues avec les agents sociaux du monde de la technique, l'inverse devrait également être vrai. Ces derniers devraient développer une forme de "pensée journalistique", pour faciliter une compréhension mutuelle, voire une meilleure adaptation à un processus éditorial journalistique qui ne peut être envisagé de la même manière que dans d'autres domaines, tels que le marketing ou la communication. Cela implique un changement de paradigme : avant d'être techniques, les activités liées à la production automatisée d'informations doivent d'abord être considérées comme journalistiques. Ceci soulève, en filigrane, l'enjeu d'un cadre normatif qui devrait être communément partagé, considérant que le journalisme engage la responsabilité sociale de chacun de ses acteurs envers les audiences des médias d'information envers lesquelles ils sont redevables. Cela étant, cette problématique ne va pas apparaître lorsque des journalistes assurent le paramétrage d'un logiciel d'automatisation de la production d'informations, dans la mesure où ils ne se trouvent pas en état de dépendance envers des tiers, et qu'ils conservent ainsi le contrôle de la chaîne éditoriale. Mais cela implique de nouvelles pratiques professionnelles, qui sont à aborder en termes de processus puisque les journalistes sont amenés à déconstruire leurs manières de faire habituelles en vue de les standardiser.

Si nous considérons les deux hypothèses posées au début de cette thèse, à la lumière des observations réalisées dans le cadre des expériences "Bxl'air bot" et "Quotebot", leurs réponses ne peuvent être envisagées sur un mode binaire et doivent donc être nuancées.

Hypothèse 1 – Les remises en question professionnelles et identitaires induites par la métaphore du robot journalisme vont être source de résistance et ne vont pas permettre de réaliser des usages journalistiques.

Selon que l'on se place du point de vue des imaginaires technologiques induits par la métaphore du "robot journaliste" ou de celui de la résistance à l'innovation technologique, des logiques d'usage et de non-usage vont émerger d'une manière qui ne va pas forcément converger. Ces logiques dépendent des contextes organisationnels, culturels et sociaux au sein desquels va se déployer un système d'automatisation de la production d'informations, ainsi que de représentations où ces technologies sont abordés dans leurs dimensions symboliques et idéologiques (Mumford 1965). Cela peut soit entrer en contradiction avec l'idéologie du journalisme, dès lors qu'elles sont perçues comme une menace pour un emploi humain qui n'est alors plus considéré que sous l'angle d'une somme de tâches standardisables (ou mécanisables) ; soit renforcer la symbolique du métier d'informer, lorsque la technologie est entendue comme un moyen d'incarner des savoirs et savoir-faire professionnels, en vue de soutenir des activités journalistiques humaines.

Dans la première étude de cas, où le cadre technologique de la rédaction s'inscrit dans la perspective de pratiques professionnelles présentant peu d'appétence pour les technologies numériques et pour une approche par données dans le journalisme, la perception de la menace sur l'emploi a perduré tout au long de l'expérience et, cela, malgré des avantages relatifs qui furent davantage perçus lors de la mise en récit des productions automatisées (faire gagner du temps dans la récolte de l'information, constituer une source d'informations inédite). Dès lors que le traitement automatisé des données se limite à proposer une photographie chiffrée d'un phénomène au cours d'une période déterminée, le travail de contextualisation et d'analyse du journaliste est celui de la mise en sens qui donne corps au travail journalistique. Les imaginaires liés au mythe de Frankenstein ("*la méga menace du robot qui piquera ma place est toujours bien présente*") peuvent donc bien constituer un frein à l'usage, bien que ce mythe ne puisse complètement se réaliser : puisqu'il ne lui est pas reconnu faire acte de journalisme, le système d'automatisation ne peut prendre le pas sur l'humain. *A contrario*, l'image d'un robot sympathique, à mettre en relation avec le mythe du robot hollywoodien, participe à donner une image positive du média. Elle favorise également une adoption symbolique de l'artefact, qui consiste dans l'acceptation de l'idée de l'innovation, bien qu'elle reste confinée à un usage discursif. La métaphore du robot peut aussi jouer un rôle de dédramatisation du phénomène, s'agissant de jouer sur la corde de l'humour pour légitimer l'adoption symbolique d'une "nouvelle recrue" d'un genre particulier : "*Le Bxl'air bot ne prend pas beaucoup de place et ne sert pas le café*", "*Il démontre une mémoire infailible des chiffres et travaille 24 h sur 24 sans jamais*

prendre de pause". Pour autant, cette adoption symbolique n'a pas toujours été partagée par tous ("*À la limite, on n'a pas besoin du robot*"), ce qui laisse également supposer une résistance face à un dispositif qui remettrait en question le rôle du journaliste dans sa mission d'informer ou une manière habituelle de travailler, qui privilégie le terrain.

Les non-usages de l'artefact "Bxl'air bot" ne sont pas toujours liés aux seuls imaginaires véhiculés par la métaphore du robot : certains journalistes n'éprouvaient pas d'intérêt pour la thématique proposée, tandis que d'autres avaient du mal à appréhender une démarche journalistique s'appuyant sur des données en tant que matériau de base à l'information ("*Les chiffres me font peur*"). Les non-usages peuvent encore s'expliquer par un imaginaire technologique plus général, où les technologies sont considérées comme chronophages et nécessitant un temps d'apprentissage et d'adaptation. Pour ces raisons, l'expérience "Bxl'air bot" a pu placer les journalistes face à leurs propres compétences, alors que le propos de l'artefact était précisément de leur faire gagner du temps en prenant en charge des tâches ingrates. La seule journaliste qui en a fait usage dans ses productions journalistiques reconnaît d'ailleurs que les journalistes n'auraient eu ni le temps ni la patience de s'atteler à un tel travail. Les non-usages peuvent donc être ici liés à une forme de technophobie, couplée à une résistance technologique, voire à un refus idéologique. Nous retrouvons donc ici face à une variété de situations qui fondent les logiques de non-usage : une absence de besoin face à un dispositif que la rédaction n'a pas sollicité, un désintérêt relatif au domaine d'application, un refus de remettre en question une approche du journalisme que la rédaction qualifie elle-même de "traditionnelle", et une résistance contre la symbolique ambivalente de l'objet.

Dans la seconde étude de cas, où le cadre technologique du service d'informations boursières peut être défini à travers des pratiques professionnelles intégrant les technologies numériques, et où le fait chiffré constitue le matériau de base à l'information, la métaphore du robot a eu peu d'impact. L'adoption symbolique, rencontrée à l'unanimité dès les débuts du projet, est celle d'une machine destinée à faire gagner du temps aux journalistes, de manière à leur permettre de se concentrer sur d'autres tâches moins répétitives et à plus forte valeur ajoutée. La menace perçue n'est pas celle du dispositif d'automatisation en tant que tel, mais éventuellement celle de la concurrence de nouveaux acteurs technologiques qui ne sont pas traditionnellement liés au monde du journalisme ("*C'est eux qui vont nous remplacer?*"). Des imaginaires relevant d'un isomorphisme du système avec les journalistes ayant participé à sa conception vont faire apparaître une dimension affective dans le rapport à l'objet ou à ses productions. Bien que certains d'entre eux aient admis que l'artefact contienne une petite part d'eux-mêmes, les journalistes n'ont pas souhaité rebaptiser "Quotebot" par un petit nom qui lui conférerait un caractère plus "humain". Toutefois, la charge représentationnelle liée à la métaphore du robot a incité à la prudence en ce qui concerne la communication à l'ensemble de la rédaction, dans le souci de ne pas induire un rapport anxigène vis-à-vis d'un artefact conçu pour soutenir un service en particulier. De la même manière, celle-ci n'est pas utilisée dans la présentation officielle du projet, en le désignant pour ce qu'il est : un générateur automatique de textes.

Hypothèse 2 – Les usages journalistiques vont dépendre de l'adéquation des productions automatisées avec les savoir-faire et exigences du journalisme.

Les expériences "Bxl'air bot" et "Quotebot" témoignent de la nécessité d'inclure les journalistes-utilisateurs finaux dans le processus de conception d'un artefact de production automatisée d'informations, dès lors que celui-ci a pour objet de soutenir leurs pratiques professionnelles. Il s'agit non seulement de rencontrer la manière dont les journalistes envisagent leurs routines, mais aussi les logiques éditoriales dans lesquelles le système d'automatisation va s'inscrire. L'implication active des journalistes dans le processus de conception d'un artefact d'automatisation de la production d'informations peut être considérée comme une première forme d'usage où il s'agit, pour eux, de se projeter dans leurs usages finaux, que ce soit en termes de fonctionnalités ou de résultats attendus. Cette approche inclusive est susceptible d'amener les journalistes à réfléchir sur leurs propres processus rédactionnels, voire à ce qu'ils déconstruisent leurs pratiques dans le cadre d'une nouvelle forme de travail, laquelle consiste à élaborer des gabarits de texte en vue de leur automatisation. Cela permet de démystifier une technologie pouvant apparaître comme complexe, voire duale sur le plan de ses représentations. Faire des journalistes des acteurs de l'innovation ne va pas garantir des usages finaux, les professionnels de l'information devant nécessairement trouver un intérêt opérationnel à partir du moment où ils sont confrontés à la matérialité des productions de l'artefact. La perception des bénéfices d'un dispositif d'automatisation de la production d'informations – un gain de temps et une source d'informations supplémentaire, dans les deux cas – favorise les usages mais ne les mobilise pas pour autant.

La complémentarité entre le journaliste et l'artefact d'automatisation apparaît de manière non équivoque : c'est à ce premier qu'appartient la mise en sens de l'information, qu'il s'agisse d'intégrer tout ou partie des productions automatisées dans ses propres productions, ou de décider si la production automatisée sera publiée ou non. Cette mise en sens, qui comprend des activités de contextualisation ou d'analyse, ne peut être automatisée. Mais pour que cette complémentarité soit réalisée, l'adéquation des productions automatisées avec les savoir-faire et exigences journalistiques est un préalable, celle-ci ne pouvant être garantie qu'à partir du moment où les agents sociaux du monde de la technique apparaissent comme légitimes aux yeux des journalistes. Le jeu d'interactions entre l'ensemble des agents sociaux impliqués dans le processus de conception de l'artefact peut être compris comme un processus de négociation et de traduction sociotechnique, où les possibilités et limites de l'artefact vont apparaître clairement. Pour parvenir à rencontrer les exigences journalistiques d'actualité, d'exactitude et de fiabilité, qui font référence aux dimensions cognitive et normative de la culture professionnelle du journalisme, la qualité des données doit être garantie, et l'expertise du domaine d'application être partagée avec les agents sociaux en charge du développement de l'artefact. Dans les deux études de cas, le système de production automatisée d'informations n'est pas considéré comme autonome : seule la médiation journalistique donne lieu à une information journalistique.

Dans la première étude de cas, un seul usage d'appropriation des productions de l'artefact a été constaté mais l'on ne peut affirmer que celui-ci soit complètement réalisé. Aussi, la journaliste a-t-elle gardé ses distances avec un dispositif "non sollicité", auquel elle reconnaît néanmoins des qualités en ce qui concerne la collecte et l'analyse primaire des données. Dans la deuxième étude de cas, les erreurs de formulation et d'analyse constatées, de manière récurrente, par un journaliste, sont l'élément déclencheur d'une résistance dirigée plus largement contre l'organisation, dès lors que l'implication de ce journaliste dans le processus de test et de corrections a été perçu comme induisant une charge de travail supplémentaire. À l'examen des évaluations réalisées auprès des journalistes en cours d'expérience, il apparaît que ceux-ci estiment que "Quotebot" n'écrit pas comme eux car il commet des erreurs. Seule la résolution de ces problèmes permettra d'en envisager l'usage.

Quatre logiques d'usage ont été identifiées dans ces deux études de cas (Caradec 2001) : une logique utilitaire liée à l'utilité de l'objet (la coordinatrice du magazine *Alter Échos*, lorsqu'elle consulte l'application en ligne en vue de préparer un dossier thématique) ; une logique identitaire, en adéquation (ou non) avec l'individu (les journalistes de *L'Echo*, lorsqu'ils estiment que "Quotebot" n'écrit pas comme eux) ; une logique de médiation, où l'usage va être encouragé par l'intervention d'un tiers (lorsque dans ces deux études de cas, les agents de médiation sociotechnique vont jouer un rôle pédagogique en expliquant le dispositif et ses processus) ; et une logique d'évaluation, où les usages seront fonction de leurs représentations positives (lorsque les journalistes du service "Investir" estiment que le dispositif les aidera à gagner du temps) ou négatives (lorsque la dualité de la métaphore du robot contribue à de non usages au sein de la rédaction d'*Alter Échos*).

Le processus d'adoption d'une innovation technologique n'est jamais immédiat, et il peut être lié à différents schémas d'utilisateur/usagers. De plus, la formation des usages nécessite du temps et ne peut être automatiquement déduite du processus de conception (Akrich 1990), même si les utilisateurs finaux y sont activement impliqués. À la lumière des résultats de cette recherche, il apparaît donc que les mécanismes de structuration des usages relèvent d'un processus complexe qui va dépendre de facteurs endogènes, liés au contexte organisationnel et aux routines journalistiques, et de facteurs exogènes, liés aux cadres plus larges de l'imaginaire technologique et de la relation homme-machine. Ceux-ci peuvent donc varier d'un projet à un autre, d'une organisation médiatique à une autre, et d'un contexte professionnel à un autre. Le point de convergence est à trouver dans le respect des fondamentaux du journalisme, dans le cadre d'un partenariat où le journaliste conserve le monopole de la mise en sens à travers son travail d'analyse et de contextualisation. Cette posture permet d'affirmer la valeur ajoutée du journaliste, sans remettre pas en question son identité professionnelle. Malgré leurs potentiels dans la prise en charge de tâches répétitives et chronophages, les systèmes d'automatisation de la production d'informations ne sont pas sans limites. Lorsqu'ils deviennent des outils du journalisme, celles-ci sont clairement balisées, et cela d'autant plus qu'aucune profession ne peut être réduite à la somme de ses tâches. Il s'agit là d'un savoir-faire et d'une expérience du métier qui ne peuvent être automatisés.

Un processus de traduction sociotechnique

Les observations empiriques réalisées dans le cadre de cette thèse ont permis d'identifier trois variables sociales et culturelles à l'œuvre lors de la construction sociale d'un artefact d'automatisation de la production d'informations : les intentions, les médiations et les représentations. Sur un plan conceptuel et théorique, celles-ci s'inscrivent dans la dynamique de ce que l'on peut appeler les cercles de la traduction sociotechnique, qui permettent de comprendre comment s'opère un processus de traduction d'exigences journalistiques en solutions technologiques (Figure 4.19).

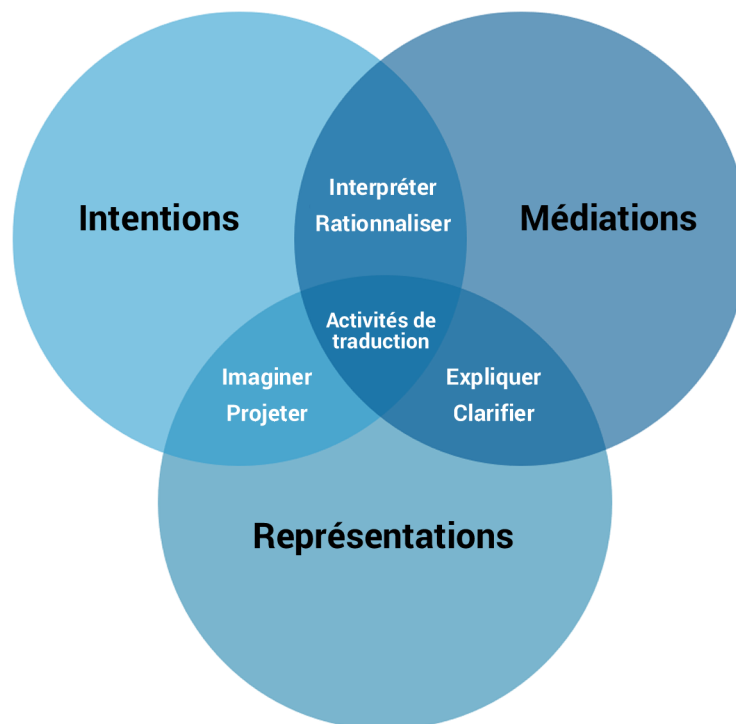


FIGURE 4.19 – Cercles de la traduction sociotechnique (schéma de l'auteur)

Ce processus de traduction sociotechnique préside au développement d'un artefact qui tient compte de contraintes qui sont à la fois techniques et journalistiques. Aucune des activités à l'œuvre dans ce processus ne peut être considérée comme neutre ou de manière isolée, qu'il s'agisse des intentions que les journalistes vont exprimer de manière implicite ou explicite, des représentations qui vont participer à forger leurs intentions, et des activités de médiation qui, malgré leurs objectifs de rationalisation technique, relèvent toujours d'une forme d'interprétation. Ces interactions s'opèrent sur un plan collectif et individuel : collectif, car le journalisme peut être considéré comme un bien partagé par un ensemble d'individus dans un contexte professionnel et organisationnel déterminé; individuel, car chaque journaliste impliqué dans ce processus dispose de sa propre expérience et de ses propres représentations sociales et culturelles. S'engage alors un jeu de négociations où l'activité de médiation va s'efforcer de trouver un terrain d'entente entre toutes les parties prenantes.

La formulation des intentions est conditionnée par les représentations des journalistes, qu'elles soient liées à l'imaginaire véhiculé par l'objet "robot", ou à la manière dont les journalistes se projettent dans leurs usages finaux (à quoi l'outil va-t-il me servir et de quelle manière sera-t-il le plus efficace pour répondre à mes besoins?). Les activités de médiation sont susceptibles d'influer sur les intentions, en tant qu'appui à la définition des exigences journalistiques ou à la prise de décision; ainsi que sur les représentations, en participant à la démystification d'une solution technologique pouvant être perçue comme une boîte noire dont on ne connaîtrait que les entrants (données) et les extrants (productions générées de manière automatique). Aussi, l'activité de médiation sociotechnique peut-elle essentiellement s'entendre dans ses aspects pédagogiques. Elle a pour objet d'expliquer le fonctionnement de l'artefact (en termes de processus), de clarifier les exigences journalistiques (les intentions), et de manager un projet dont les parties prenantes peuvent être issues de cultures professionnelles différentes. Un agent de médiation disposant d'un profil professionnel hybride permet de fluidifier les dialogues entre les agents sociaux du monde du journalisme et du monde de la technique.

Les activités de médiation sociotechnique sont constitutives de la problématique de l'imaginaire technologique (explication ou démystification de la technologie) et de celle des dimensions culturelles de l'automatisation de la production d'informations (facilitation des traductions des savoirs et savoir-faire journalistiques). Aussi, peuvent-elles être considérées comme un facteur de facilitation des usages, en s'assurant que les traductions techniques rencontrent les exigences professionnelles que les journalistes vont eux-mêmes formuler dans le cadre d'un processus inclusif. Mais elles ne vont pas garantir que leurs intentions – lesquelles découlent également de leur expertise et de la manière dont ils vont se projeter dans leurs usages – soient toujours bien comprises ou interprétées, pas plus qu'elles ne vont influencer de manière directe sur les usages finaux. Ceux-ci vont dépendre, pour une large part, de la capacité de l'artefact à reproduire des savoirs et savoir-faire professionnels. En ce sens, les conditions pour un partenariat fructueux entre journalistes et technologies d'automatisation de la production d'informations ne peuvent être définies qu'en fonction de leurs contextes d'usage, où vont se déployer des cultures professionnelles et des imaginaires technologiques potentiellement hétérogènes. Le dénominateur commun est à trouver dans une traduction technologique qui soit fidèle aux exigences journalistiques. Il s'agit là du préalable à une collaboration où le journaliste n'envisage pas ce partenariat d'égal à égal, dès lors qu'il conserve la main sur la manière dont l'information est délivrée aux audiences.

Mise en perspective et pistes d'approfondissement

Cette recherche s'inscrit dans le prolongement des études du journalisme qui se sont concentrées sur la diffusion d'innovations technologiques dans les rédactions, dès le début des années 2000, dans le contexte du développement des politiques de convergence. Plus particulièrement, elle fait écho au tournant matériel des études du journalisme, où technique et social ont été abordés sous l'angle de leurs influences réciproques, en proposant un cadre théorique

inspiré par les études des sciences et des technologies (STS). Cette recherche témoigne de cette interdépendance entre ces deux variables, du caractère continu du processus de la diffusion d'innovations, et de l'importance du contexte de développement de l'innovation sociotechnique (Boczkowski 2004).

Les conclusions de cette thèse corrént avec plusieurs observations empiriques menées à cette époque, en ce qui concerne les mécanismes de résistance, qui caractérisent une culture professionnelle refusant de bousculer les valeurs véhiculées par l'idéologie professionnelle du journalisme (Deuze 2005). Celles-ci prennent appui sur une modification des pratiques professionnelles, un affaiblissement de la ligne de démarcation entre les équipes journalistiques et techniques (Cottle & Ashton 1999), ainsi que sur une remise en cause de la direction et du management des entreprises de presse (Ursell 2001). Il s'agissait également de résistances contre le phénomène technique, vecteur d'imaginaires utopiques et dystopiques (Mosco 2005, Domingo 2008, Jarrige 2014). Dans cette thèse, des formes similaires de résistance ont été constatées, à ceci près que les imaginaires portés par la métaphore du "robot journaliste" exacerbent une représentation duale de la technologie, dans un contexte économique fragile pour les entreprises de presse, et donc pour les journalistes. Au pire, les technologies d'automatisation de la production d'informations constitueraient un danger pour l'emploi et l'identité journalistique. Au mieux, en les libérant de tâches répétitives et chronophages, elle permettrait aux professionnels de l'information de disposer d'outils performants pour soutenir leurs pratiques. L'apport de cette recherche est donc celui de fournir une meilleure compréhension de cette perception duale, non seulement à travers une analyse du discours médiatique des journalistes, mais aussi dans une mise en situation où des journalistes sont confrontés à des artefacts d'automatisation. Cette recherche montre également qu'il ne s'agit pas d'une transformation des pratiques, mais bien d'une intégration dans les pratiques (Plesner 2009). Un autre apport est celui d'une approche transdisciplinaire qui permet de déconstruire les mécanismes d'un processus de production automatisée d'informations, dans une mise en relief, à la fois théorique et empirique, des challenges journalistiques et techniques qui les sous-tendent. Les observations réalisées dans la première étude de cas corrént également avec les recherches qui, au cours des vingt dernières années, ont mis en lumière le rapport difficile que les journalistes entretiennent avec les chiffres (Maier 2000, 2003, Schmitz-Weiss & Retis-Rivas 2018).

La particularité de cette thèse est d'avoir placé le curseur sur les usages des journalistes à travers deux études de cas. Il s'agit des premières études empiriques menée au sein de rédactions sur une période aussi longue, et au cours desquelles les journalistes ont été invités à s'impliquer activement dans la conception d'artefacts d'automatisation. Là aussi, il s'agissait d'une perspective originale, dès lors que les journalistes sont souvent tenus à l'écart des processus d'innovation. De plus, dans cette thèse, nous assistons à la naissance du phénomène en Belgique francophone : "Bxl'air bot" et "Quotebot" ont pour particularité d'être les deux premiers artefacts d'automatisation qui y ont été développés dans un contexte journalistique. Malgré la richesse des enseignements de ces études, qui se complètent en raison de leurs différences, il s'agit aussi de la principale limite de cette thèse : réalisée dans d'autres rédactions et dans

d'autres contextes, les résultats de ces observations ne seraient pas forcément convergents. Toutefois, compte tenu du caractère récent du phénomène, il n'y avait pas de points de comparaison pour les y confronter. De plus, le projet "Quotebot" n'étant pas clôturé au moment de mettre le point final à cette thèse, l'étude de ses usages est donc limitée à celle de sa conception. Pour autant, cette thèse offre des premières clés de lecture permettant de comprendre les mécanismes de structuration des usages professionnels lorsqu'un artefact d'automatisation est abordé dans son double statut d'objet – en tant qu'élément qui agit comme un moyen de médiation pour produire du journalisme (De Maeyer & Le Cam 2015) – et d'outil du journalisme. Partant, elle esquisse les contours des conditions qui permettent d'envisager une complémentarité entre le journaliste et la machine. Mais elle ne fournit pas un mode d'emploi qui pourrait fonctionner dans tous les cas, dès lors que la problématique des usages ne peut être abordée que dans les particularités (et la complexité) de ses contextes de structuration.

Les méthodes mise en œuvre dans le cadre de ces deux études empiriques, en ce compris les modèles d'évaluation conçus dans le cadre de cette recherche, consistent en un autre apport original de cette thèse. C'est ce en quoi ce travail de recherche ouvre la voie à des recherches empiriques ultérieures qui souhaiteraient suivre la construction sociotechnique d'un artefact d'automatisation tout au long de son processus de conception et de développement, dans une perspective inclusive et transdisciplinaire. Ces recherches pourraient être élargies à d'autres formes de processus automatisés, considérant que l'automatisation concerne désormais l'ensemble de la chaîne de production journalistique. Cette thèse a également ouvert une porte sur la problématique de la personnalisation automatisée d'informations, qui n'est pas sans véhiculer des enjeux éthiques et sociaux importants, et qui serait également à aborder en termes de processus. Une autre piste d'approfondissement consiste à envisager le management d'innovations technologiques sous le prisme conceptuel des cercles de la traduction sociotechnique, qui pourraient ainsi être complétés et affinés. Une dernière piste concerne l'étude de la position ambiguë de ces nouveaux acteurs du monde du journalisme, dans leur refus d'admettre que leurs activités ne peuvent être réduites à la seule approche technique, dès lors qu'ils participent à un processus éditorial et donc journalistique.

Retour sur un processus de recherche et des situations d'immersion

Les deux études de cas présentées dans cette thèse ont pour particularité de m'avoir activement impliquée dans le processus de développement d'artefacts d'automatisation de la production d'informations. Cette approche embarquée s'est traduite de manière différente dans chacune des rédactions étudiées, dès lors que mon implication n'était pas de même nature et que les contextes d'expérience présentaient d'importantes différences, en termes organisationnels, professionnels et opérationnels. Celles-ci m'ont conduit à mettre en place des dispositifs de recherche adaptés à chacun des contextes d'étude, où mon rôle de chercheuse en immersion a été mis à l'épreuve d'une quête constante de ce délicat équilibre à trouver, entre un engagement actif sur les terrains d'enquête et une nécessaire mise à distance critique des observations réalisées.

Dans ces deux études de cas, les objectifs visaient la production de connaissances à propos d'un phénomène récent, encore peu étudié sur un plan empirique. Je cherchais à comprendre les mécanismes sociaux d'adhésion et de résistance à un projet d'automatisation de la production d'informations, dès lors que celui-ci est conçu en tant qu'outil au service du journalisme. Ma première hypothèse de départ était que les imaginaires véhiculés par ce que l'on appelle vulgairement "le robot journalisme" étaient susceptibles de freiner le développement d'un tel projet. Elle était nourrie par plusieurs travaux dans les domaines de la sociologie de l'innovation et des études du journalisme (Flichy 2001, Massit-Folléa 2002, Musso 2009, Domingo 2008), lesquels convergent sur le fait que l'usage d'une technologie dépend, en partie, de la représentation que l'on s'en fait. Elle l'était aussi par des contacts réguliers avec le monde professionnel. Ayant assisté à plusieurs conférences sur le phénomène de la production automatisée d'informations (en France, en Allemagne, au Portugal, en Autriche, en Pologne et en Belgique), j'étais interpellée par ces questions qui, tel un mantra, revenaient invariablement : "*Devons-nous avoir peur?*", "*Allons-nous être remplacés?*".

Dans le même temps, je m'interrogeais sur la manière de traduire des exigences journalistiques dans un système d'automatisation, pour que cela ait du sens sur un plan opérationnel. Leur intégration rigoureuse permettrait-elle une approche professionnelle qui serait envisagée en termes de complémentarité? L'ouvrage "Race Against The Machine" des économistes américains Erik Brynjolfsson et Andrew McAfee (2011) avait agi comme l'élément déclencheur de

cette seconde réflexion. Il concluait, en substance, que si les humains s'engageaient dans une course contre les machines, celle-ci serait perdue d'avance. La seule manière d'appréhender le phénomène de l'automatisation et de la robotisation consisterait à s'engager dans une course non pas contre, mais avec les machines. Les scientifiques abordaient plus largement la question d'une quatrième révolution industrielle, à propos de laquelle plusieurs études prospectives prédisent de lourdes conséquences sur l'emploi humain (Ford 2015) et contribuent, de ce fait, à alimenter un climat anxieux.

Mes compétences professionnelles et techniques – en développement numérique et en linguistique computationnelle – ont été mises au service de cette recherche, me permettant d'élaborer des outils et des méthodes d'évaluation originaux, s'agissant de les adapter aux caractéristiques des terrains étudiés. Mon bagage acquis lors du master en sciences et technologies de l'information et de la communication m'a orientée vers une littérature académique qui fut appréhendée en termes de complémentarité. En effet, il s'agissait d'étudier deux groupes socio-professionnels, dans une mise en situation où l'objet technique ne peut être abordé autrement que par la manière dont il est construit culturellement et socialement : un système de production automatisée d'informations n'émerge pas de lui-même, il est le fruit d'une réflexion humaine engageant une traduction sociotechnique forgée par des intentions, des représentations et des médiations. Il incorpore des savoir-faire qui, s'ils sont techniques, sont aussi – et avant tout – journalistiques, et cela d'autant plus lorsque leur propos est de soutenir des routines professionnelles. Cette approche m'a permis de penser en termes de processus, dans le souci de comprendre l'agencement, l'imbrication entre un processus technique et un processus journalistique. Mon passé professionnel – qui inclut près de quinze années d'activités au sein d'une union professionnelle de journalistes – a également été un appui dans ma manière d'appréhender les groupes sociaux étudiés, en tant qu'observatrice privilégiée du terrain de la presse belge francophone, consciente de ses particularités, mais aussi de ses difficultés.

Sans ce bagage et cette somme d'intérêts personnels et professionnels, cette recherche ne se serait pas engagée sur la voie de la transdisciplinarité. À l'instar d'un photographe variant ses angles de prises de vue pour représenter une réalité donnée, ma démarche était celle de faire varier les angles d'observation en vue de mieux comprendre un phénomène récent (Denis 2009). Cette posture n'a pas seulement été théorique : elle a également trouvé un ancrage pratique. Consciente de la prise de risques que cela impliquait, j'ai considéré que seul l'équilibre de mon positionnement me permettrait une étude rigoureuse. La quête de cet équilibre a induit une tension, largement éprouvée dans la recherche en sciences sociales, mais elle était aussi inévitable en raison de la proximité induite entre le chercheur et l'objet de son étude (Elias 1993, Horvat 2013) ¹²⁹.

¹²⁹ Voir aussi, les sections 0.3.2 et 0.4 dans l'introduction de cette thèse, pp.26-30

"Bxl'air bot" : une perspective expérimentale

À l'époque où ont débuté mes travaux de recherche, il n'existait pas, en Belgique francophone, de rédaction engagée dans un processus d'automatisation de la production d'informations. Je ne pouvais me satisfaire du seul examen approfondi de productions académiques pour une compréhension avancée du phénomène, convaincue de la nécessité de disposer d'un terrain empirique (Flyvbjerg 2006). Dès lors, s'il n'existait aucun objet d'étude, pourquoi ne pas en créer un? Cette idée a germé progressivement au cours de la deuxième année de recherche, après avoir répertorié différents dispositifs d'automatisation de la production d'informations mis en place dans des médias en Europe et aux États-Unis, mais aussi après avoir exploré le domaine de l'*open data* belge, en vue de trouver un domaine d'application qui convienne à un processus d'automatisation. Les données devaient impérativement rencontrer des critères de qualité qui soient à la fois techniques et journalistiques, en vue de leur automatisation (Dierickx 2017). Surtout, cette automatisation devait faire sens sur un plan journalistique.

Le domaine de la qualité de l'air en région bruxelloise m'est apparu comme la piste la plus pertinente à explorer : les données étaient accessibles et mises à jour régulièrement. Dans une perspective journalistique, il s'agissait d'un domaine donnant lieu à des controverses que les médias relayaient régulièrement car si elles donnent "à voir" un état de fait à un instant T, les données ne disent rien des causes et des conséquences de la pollution atmosphérique. J'avais trouvé ce qui pourrait être un objet d'étude susceptible de faire sens, celui que je proposerais à la rédaction d'un média d'informations pour étudier ce que des journalistes en feraient... ou pas. L'idée de mettre mes compétences techniques au profit de cette recherche n'a donc pas été instantanée, elle fut nourrie par une longue réflexion. Elle fut également motivée par une autre nécessité : celle de faire l'expérience de la technologie, en vue de mieux en comprendre ses possibilités et limites : jusque-là, je les avais étudiées sur le seul plan théorique, notamment dans mon mémoire de master en sciences et technologies de l'information et de la communication. Il s'agissait donc de mieux comprendre un phénomène technique, avant d'expliquer les situations sociales (Barbier 1996 :14).

Je me retrouvais ainsi engagée dans un processus de recherche-action, une démarche qui trouve son origine dans les travaux de Kurt Lewin (Van Trier 1980, Rhéaume 1982), considérant que l'action permet de mieux comprendre une réalité sociale, tout en s'inscrivant dans un processus expérimental (Lewin 1952, cité par Jouison-Laffitte, 2009). Les caractéristiques de la recherche-action sont de viser à résoudre un problème dans une mise en situation (et dans le cadre d'un processus itératif), en vue de produire des connaissances scientifiques sur la situation étudiée. La démarche qui allait ici s'engager s'inscrivait pleinement dans une dynamique de *Participatory Action Research*, en plaçant l'accent sur une collaboration entre des acteurs de l'organisation et le chercheur, dans une dimension à la fois pratique (résolution du problème) et théorique (production de connaissances sur la base de l'action) (Goyette & Lessard-Hébert 1987, Jouison-Laffitte 2009).

En raison de ma participation active à l'expérience, cela a induit une recherche constante de l'équilibre entre l'engagement du chercheur et l'analyse critique. Bien que le principe de neutralité soit illusoire, dès lors que ce type de démarche implique un engagement actif du chercheur, une manière d'atteindre cet équilibre était de documenter et d'argumenter chacun de mes choix, dans le souci de la rigueur scientifique. Dans l'expérience "Bxl'air bot", cette démarche a été nourrie par la littérature relative aux bonnes pratiques à l'œuvre dans les domaines du développement de bases de données informatisées, de la conception d'interfaces, de l'ergonomie des systèmes informatiques et des modes de narration pilotés par des données. Cette forme de recherche-action ne peut être considérée comme une méthodologie de recherche, mais bien comme un outil contribuant au développement de méthodes de recherche à visée qualitative (Rhéaume, 1982). Il s'agissait donc bien d'un moyen et non d'une fin en soi.

Ma proximité professionnelle avec la rédaction du magazine *Alter Échos*, en particulier avec sa coordinatrice, fut considérée comme une opportunité pour cette recherche. Nous avons eu l'occasion de travailler ensemble sur la refonte graphique et éditoriale du site web du magazine, et j'avais pu constater son vif intérêt pour les possibilités que peuvent permettre un environnement numérique. Par ailleurs, le magazine relayait régulièrement des actualités relatives à la qualité de l'air bruxellois, bien que cette matière ne constitue par le centre de ses préoccupations éditoriales. Toutefois, je ne souhaitais pas lui soumettre un projet qui soit purement théorique, voire abstrait dans sa représentation, mais lui donner "à voir" un premier prototype. Il s'agissait là d'une autre prise de risques : la coordinatrice de la rédaction aurait pu refuser, mais c'est avec enthousiasme qu'elle a accepté d'accueillir cette expérience. Je n'avais aucun rapport de propriété à l'objet que j'avais développé : il avait été conçu sans avoir eu le sentiment qu'il m'appartienne d'une quelconque manière. Ce détachement de l'objet technique, qui m'apparaît aujourd'hui comme un préalable, me semblait naturel à l'époque. Dès lors que celui-ci intégrait un contexte d'expérience, il s'agissait de laisser les participants en disposer (ou non) à leur guise. Mon rôle devenait alors celui d'une médiatrice du code informatique, en charge de la maintenance du système ainsi que du contrôle, dans le temps, de la validité des données. Cela a également impliqué de me documenter, en amont, sur le domaine d'application de la qualité de l'air et de solliciter des experts, les processus d'automatisation ne pouvant être appréhendés sans le préalable de la connaissance d'un domaine que je découvrais dans le contexte de cette recherche (Reiter & Dale 1997).

Dans cette expérience, ce que j'allais observer, c'est ce que les journalistes allaient faire ou non du "Bxl'air bot". Aussi, il n'était pas important qu'ils fassent usage de l'outil : d'emblée, il a été considéré que ne pas en faire usage était tout aussi important. La recherche en sociologie des usages a d'ailleurs souvent souligné cet aspect, considérant les mécanismes complexes qui forgent tant les usages que les non-usages, ces derniers ne pouvant être uniquement attribués à l'incompétence d'un utilisateur ou à un objet mal conçu (Akrich 1990, Proulx 1994, Boudoukhane 2006, Kellner *et al.* 2010). Bien sûr, nous pouvons considérer qu'à partir du moment où ce projet avait été présenté à la rédaction, deux types de biais étaient susceptibles d'influer sur le déroulement de l'expérience : celui du choix de la thématique, dont il est ensuite apparu

qu'elle ne rencontrait pas les intérêts personnels et professionnels de la majeure partie de la rédaction; et celui d'une version prototypée de l'artefact qui résultait d'un processus qui m'avait exclusivement engagée jusque-là, bien que celui-ci ait été rigoureusement balisé.

Le propos était de soumettre à la rédaction un objet qui rencontre de "bonnes pratiques", à la fois techniques et journalistiques. En quelque sorte, il s'agissait d'un modèle qui serait "idéal", sans pour autant prétendre être parfait, mais qui présentait aussi le risque que les journalistes développent rapidement une attitude positive envers lui. De manière à rencontrer les besoins des deux journalistes impliquées dans le processus de confrontation à l'objet sociotechnique, il s'agissait également de faire preuve d'un maximum d'ouverture et de disponibilité. Rester à l'écoute des demandes de toutes natures était un autre préalable. Dans un souci de transparence, les processus, de même que ses possibilités et limites, ont été expliqués et documentés, posant le postulat selon lequel une meilleure compréhension du système d'information faciliterait l'identification de leurs besoins. Consciente de la nature interprétative de ma fonction, chaque nouveau développement apporté au dispositif sociotechnique devait être validé avant sa mise en ligne. En ce sens, mes différentes interactions avec les agents sociaux impliqués dans ce processus n'ont pas été sans influencer leur manière de l'envisager, dès lors que leur réflexivité a été encouragée. Ceci peut être considéré non pas comme un biais, mais comme un levier pour la recherche (Brasseur 2012).

Si la finalité de l'expérience consistait à observer et analyser la manière dont les journalistes allaient s'emparer ou non de l'artefact technologique, la mise à distance critique s'imposait comme une nécessité. Dans cette perspective, je me suis attelée à définir une série de garde-fous : celui de distinguer mes relations de recherche avec la rédaction des relations professionnelles que j'entretenais, par ailleurs, avec elle; celui, aussi, de me placer à distance "physique" lors des réunions de rédaction, de manière à souligner la dimension observante de la démarche de recherche. Un autre garde-fou fut d'avoir recours à des questionnaires dont les réponses ont été enregistrées de manière anonyme, en vue de permettre une plus grande liberté de parole. En effet, l'une de mes craintes était qu'un journaliste formule des réponses pour me "faire plaisir" ou ne pas me "froisser". De plus, ma démarche n'était pas celle de promouvoir un outil au sein d'une rédaction mais bien de comprendre ce que les journalistes allaient en faire (ou non) et pourquoi. Je souhaitais que cette posture soit sans équivoque. La mise à distance critique a également eu lieu, de manière pratique, dans la rédaction de cette thèse : j'ai attendu trois mois avant de me lancer dans le travail d'analyse et d'écriture.

"Quotebot" : une perspective évaluative

L'expérience "Bxl'air bot" devait être la seule étude de cas développée dans le cadre cette recherche. Mais alors que se clôturait la phase d'analyse des observations, une seconde opportunité s'est présentée, sous la forme d'une invitation à participer à un projet d'automatisation de la production d'informations pour les quotidiens du groupe Mediafin. Le responsable du développement et de la transformation numérique de *L'Echo*, que je connaissais par ailleurs sur un plan professionnel, souhaitait introduire un dossier de financement auprès du Google Digital

News Initiative. Il m'a invitée à faire partie de l'expérience, en tant que chercheuse et consultante, en raison de mon expertise dans le domaine de la génération automatique de textes. Le propos était d'accompagner les journalistes dans le développement d'un système qu'ils devaient définir en amont de son développement technique, mais aussi de réaliser des analyses tout au long du projet, dans une visée évaluative. Cette opportunité a été considérée car il s'agissait d'un nouveau terrain de recherche présentant des caractéristiques très différentes de la première rédaction étudiée. Il pouvait donc être abordé de manière complémentaire, en étudiant ce qui non seulement pouvait rejoindre les observations du premier cas, mais aussi ce qui pouvait l'en séparer. En raison de ses particularités, ce projet représentait donc aussi un potentiel d'enrichissement de la démarche de recherche. Le dossier a été accepté par Google environ deux mois après son introduction, mon rôle ayant alors été décrit comme s'inscrivant dans le cadre d'un partenariat : "*Developed in partnership with robotic writer specialist Syllabs and academic researcher Laurence Dierickx (ULB) – and supported by the DNI Fund – Quotebot extracts market data from a structured database to automatically produce text and graphics*"¹³⁰.

Dans le cadre de la préparation de ce dossier, j'avais élaboré un protocole de recherche que j'avais soumis, pour validation, au responsable du développement et de la transformation numérique de *L'Echo*. Ce protocole était orienté sur le concept d'expérience utilisateur (UX), tel que défini par l'ISO, considérant que les usages du dispositif dépendraient de sa capacité à satisfaire les exigences posées par les journalistes, tout en leur fournissant une expérience "agréable". Ce protocole disposait que "*Compte tenu des apports d'une approche croisée en matière d'évaluation et du contexte de la recherche, visant six journalistes attachés à service en particulier et portant sur leurs usages d'un 'brouillon automatisé' généré de manière automatique, plusieurs dispositifs d'évaluation sont envisagés dans leur complémentarité. Ceux-ci doivent être mis en œuvre dès lors que les premiers tests de Syllabs seront effectués. Dans un souci d'adéquation aux usages, il est préconisé que des ajustements soient possibles en fonction des retours des utilisateurs*". Il présentait ensuite, de manière détaillée, les quatre modes d'évaluation qui seraient réalisés tout au long de l'expérience, en vue de proposer une analyse "objectivée" de la situation : des évaluations par métriques automatiques, pour établir en quoi les productions de Syllabs correspondent aux gabarits de textes fournis par les journalistes ; des évaluations *in situ* avec les utilisateurs, de manière à voir en quoi et comment les productions automatisées sont intégrées dans les routines journalistiques ; des évaluations par questionnaire, envisagées de manière anonyme pour assurer une liberté de parole aux utilisateurs, susceptibles d'être freinés lors de réunions pilotées par un manager ; et des évaluations des utilisateurs via des entretiens, destinées à comprendre la manière dont les journalistes envisagent leur relation avec le dispositif.

Ces différents modes d'évaluation n'ont pas pu être tous réalisés, en raison du retard engrangé dans le planning du développement de "Quotebot", qui n'était pas terminé au moment de clôturer cette thèse. Toutefois, ceux qui l'ont été ont apporté des données pertinentes qui me

¹³⁰ Source "Elevating quality in journalism. Digital News Innovation Fund Report 2018", Google DNI, 2018 (PDF), document accessible via <https://alt-dot-gweb-dni-v2.appspot.com/>, consulté le 08/03/2010.

permettaient d'enrichir l'analyse de travail de recherche. C'est également dans cette même optique de complémentarité et d'enrichissement de l'analyse que la gestion du management du projet "Quotebot" a été soumise à la grille de lecture des principes de la norme ISO 9000 (Dierickx 2020). Ces apports méthodologiques ont été nourri par mon cursus en sciences et technologies de l'information et de la communication, au cours duquel le concept de qualité, compte tenu de sa complexité, a été défini comme une adéquation entre des besoins exprimés par les utilisateurs (de manière implicite ou explicite) et les usages de ces mêmes utilisateurs (Boydens et van Hooland 2011). Dans mon rôle de consultante, des rapports ont été transmis régulièrement à propos des résultats de ces analyses, et ils ont pu mettre en lumière des problèmes qui pouvaient sembler moins apparents ou sensibiliser le responsable du projet aux critères de qualité des productions automatisées. Cela étant, cette mission de consultance peut être davantage comprise dans sa dimension pédagogique, caractérisée par la transmission et le partage d'une expertise à propos des technologies et processus de génération automatique de textes.

Cette dimension pédagogique a été prégnante tout au long de ma mission d'accompagnement de la rédaction dans le processus de conception de "Quotebot". Aussi, les journalistes ont-ils été invités à s'engager dans une forme de déconstruction de leurs productions journalistiques, en vue de définir des gabarits de textes à automatiser. Cela a impliqué une réflexion sur les possibilités de variations de textes en fonction de celles des valeurs boursières, sur les expressions de référence caractérisant le domaine d'application, et sur la gestion d'un catalogue de synonymes. Ces séances de consultance ont donc pu permettre aux journalistes impliqués de mieux comprendre en quoi consistait un processus de génération automatique de textes. J'ai été attentive à ne pas outrepasser ce périmètre, en vue de faciliter l'émergence d'une réflexion qui soit propre aux journalistes, quant à leur manière d'envisager l'artefact.

Mon expertise dans le domaine de la production automatisée d'informations journalistiques a donc été partagée, tout en ayant pu permettre de répondre à d'éventuelles interrogations quant à ses processus techniques théoriques, alors qu'il n'y avait aucun accès aux processus effectifs mis en œuvre par le prestataire de services. Il s'agissait donc d'interventions dans le cadre limité de réunions avec les journalistes, qui m'ont ouvert la porte sur un important matériel de recherche – tels que les appels passés entre les représentants de Syllabs et de Mediafin et les documents de travail stockés sur une plateforme collaborative en ligne – auquel je n'aurais pas pu accéder autrement. L'un des objectifs du responsable du développement et de la transformation numérique de *LEcho* était que je puisse rendre compte, le plus fidèlement possible, du déroulé de l'expérience. Bien que mon rôle ait été actif lors de la phase d'accompagnement des journalistes, qui s'est limitée à six réunions de travail, cela m'a donné l'opportunité d'accéder à un poste d'observatrice privilégiée. Et si cela a impliqué moins de tensions dans ma posture de chercheuse, la recherche de l'équilibre entre l'engagement et la mise à distance critique est resté constante tout au long de cette seconde expérience.

Dans ces deux études de cas, il s'agissait d'une situation de recherche en immersion dont les paramètres étaient différents, étant donné la nature distincte de mon rôle actif, et cela malgré des projets présentant des finalités communes visant à soutenir des pratiques journalistiques. Leur autre point commun est d'avoir pu me placer au cœur de deux projets d'automatisation de la production d'informations et de deux rédactions, dans l'objectif d'en comprendre et analyser les dynamiques sociales. Ma situation embarquée a impliqué une gestion rigoureuse de mes différentes interactions avec les groupes sociaux étudiés, laquelle est passée par la définition de garde-fous et la mise en œuvre de méthodes de collectes de données empiriques qui soient clairement définies et balisées. Ces deux expériences furent riches en enseignements, et cela d'autant plus que le phénomène de l'automatisation de la production d'informations, en raison de son caractère récent, a été peu étudié sur un plan empirique. Même si cela a induit, à plusieurs reprises, un sentiment de se trouver dans une position inconfortable, cette difficulté ne pouvait être éludée car ce type de démarche n'est pas sans prise de risques. Pour qu'un tel projet de recherche puisse fonctionner, ces risques doivent être constamment évalués, et la démarche de recherche remise régulièrement en question : la nécessaire distance critique n'est donc pas seulement celle du chercheur par rapport à son terrain d'enquête, elle est aussi celle du chercheur face à lui-même.

Bibliographie

ABRAN, A., KHELIFI, A., SURYN, W. et SEFFAH, A. (2003). Usability meanings and interpretations in ISO standards. *Software Quality Journal*, 11(4):325–338.

AKRICH, M. (1990). De la sociologie des techniques à une sociologie des usages. *Techniques & Culture*, 16:83–110.

AKRICH, M. (1993b). Les formes de la médiation technique. *Réseaux*, 60:87–98.

AKRICH, M. (2006). Les utilisateurs, acteurs de l'innovation. In AKRICH, M., CALLON, M. et LATOUR, B., éditeurs : *Sociologie de la traduction : textes fondateurs*, Collection Sciences Sociales, pages 253–265. Presses des Mines, Paris.

AKRICH, M. (2010). Comment décrire les objets techniques? *Techniques & Culture. Revue semestrielle d'anthropologie des techniques*, 54-55:205–219.

ALLEN, C., WALLACH, W. et SMIT, I. (2006). Why machine ethics? *Intelligent Systems, IEEE*, 21(4):12–17.

ANANNY, M. et CRAWFORD, K. (2015). A liminal press : Situating news app designers within a field of networked news production. *Digital Journalism*, 3(2):192–208.

ANCIAUX, A., DURAZO HERRMANN, J. et GUAZINA, L. (2017). Etudes comparatives sur le journalisme, les médias et la politique. *Sur le journalisme-About journalism-Sobre jornalismo*, 6(2):4–11.

ANDERSON, C. et KREISS, D. (2013). Black boxes as capacities for and constraints on action : Electoral politics, journalism, and devices of representation. *Qualitative Sociology*, 36(4):365–382.

ANDERSON, C. W. (2013). Towards a sociology of computational and algorithmic journalism. *New Media & Society*, 15(7):1015–1021.

ANDERSON, C. W. (2018a). *Apostles of certainty : Data journalism and the politics of doubt*. Oxford Studies in Digital Politics. Oxford University Press, Oxford.

ANDERSON, C. W. (2018b). Genealogies of data journalism. In BOUNEGRU, L. et STRAY, J., éditeurs : *The Data Journalism Handbook 2. Towards a Critical Data Practice*, volume <https://datajournalism.com/read/handbook/two>. European Journalism Centre.

- ANDERSON, C. W. et DE MAEYER, J. (2015). Objects of journalism and the news. *Journalism*, 16:3–9.
- ANDERSON, C. W., DOWNIE, L. et SCHUDSON, M. (2016). *The news media : What everyone needs to know*. Oxford University Press, Oxford.
- ANDERSON, C. W. et SCHUDSON, M. (2009). Objectivity, professionalism, and truth seeking in journalism. *In The handbook of journalism studies*, pages 108–121. Routledge, Londres.
- ANTHEAUME, A. (2013). *Le Journalisme numérique*. Nouveaux débats. Presses de Sciences Po (PFNSP), Paris.
- ANTHEAUME, A. (2016). Les robots de l'information. *In ANTHEAUME, A., éditeur : Le journalisme numérique*, Nouveaux Débats, pages 175–194. Presses de Sciences Po (PFNSP), Paris.
- ANTTILA, J. et JUSSILA, K. (2017). ISO 9001 : 2015-a questionable reform. What should the implementing organisations understand and do? *Total Quality Management & Business Excellence*, 28(9-10):1090–1105.
- BAKKER, P. (2014). Mr. Gates returns : Curation, community management and other new roles for journalists. *Journalism Studies*, 15(5):596–606.
- BARBIER, R. (1996). *La recherche action*. Anthropos, Paris.
- BARDOEL, J. et D'HAENENS, L. (2004). Media responsibility and accountability : new conceptualizations and practices. *Communications*, 29(1):5–25.
- BASS, F. M. (1980). The relationship between diffusion rates, experience curves, and demand elasticities for consumer durable technological innovations. *Journal of business*, 53:S51–S67.
- BATINI, C., CAPPIELLO, C., FRANCALANCI, C. et MAURINO, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys (CSUR)*, 41(3):1–52.
- BECKER, H. S. et GEER, B. (1957). Participant observation : The analysis of qualitative field data. *In Field research*, pages 376–394. Routledge, Londres.
- BELZ, A. et REITER, E. (2006). Comparing automatic and human evaluation of NLG systems. *In EACL 2006, 11st Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy.
- BELZ, A. et REITER, E. (2009). An investigation into the validity of some metrics for automatically evaluating natural language generation systems. *Computational Linguistics*, 35(4):529–558.
- BENJAMIN, S. M. (2012). Algorithms and speech. *University of Pennsylvania Law Review*, 161:1445–1494.
- BENSAFTA, K. M. et GERVASIO, S. (2011). Chocs, chocs de volatilité et contagion entre les marchés boursiers. *Revue économique*, 62(2):277–311.

- BENZELL, S. G., KOTLIKOFF, L. J., LAGARDA, G. et SACHS, J. D. (2015). Robots are us : Some economics of human replacement. Rapport technique, National Bureau of Economic Research, USA.
- BERNIER, M.-F. (2006). L'idéal journalistique : comment des prescripteurs définissent le "bon" message journalistique. *Les Cahiers du Journalisme*, 16:8–45.
- BERNS, T. (2011). L'efficacité comme norme. *Dissensus*, 4:<https://popups.uliege.be/2031-4981/index.php?id=727>.
- BERRET, C. et PHILLIPS, C. (2016). Teaching data and computational journalism. *Columbia Journalism School*.
- BERRY, D. M. (2011). The computational turn : Thinking about the digital humanities. *Culture Machine*, 12.
- BERTIN, J. (2013). *Sémiologie graphique : les diagrammes, les réseaux, les cartes*. Editions de l'École des Hautes Etudes en Sciences Sociales, Paris.
- BERTRAND, C.-J. (1993). Les M* A* R* S en europe ou les moyens d'assurer la responsabilité sociale. *Communication et langages*, 97(1):22–30.
- BEVAN, N. (1995). Human-computer interaction standards. *Advances in Human Factors Ergonomics*, 20:885–885.
- BIJKER, W. E., HUGHES, T. P. et PINCH, T. (1987). *The social construction of technological systems : New directions in the sociology and history of technology*. MIT Press, Cambridge, Massachusetts.
- BOCZKOWSKI, P. J. (2004). The mutual shaping of technology and society in videotex newspapers : Beyond the diffusion and social shaping perspectives. *The Information Society*, 20(4):255–267.
- BOCZKOWSKI, P. J. (2010). Ethnographie d'une rédaction en ligne argentine. *Réseaux*, 2:43–78.
- BOCZKOWSKI, P. J. (2015). The material turn in the study of journalism : Some hopeful and cautionary remarks from an early explorer. *Journalism*, 16:65–68.
- BONNARDEL, N., PIOLAT, A. et LE BIGOT, L. (2011). The impact of colour on website appeal and users' cognitive processes. *Displays*, 32(2):69–80.
- BORGES-REY, E. (2016). Unravelling data journalism : A study of data journalism practice in British newsrooms. *Journalism Practice*, 10(7):833–843.
- BOUDOKHANE, F. (2006). Comprendre le non-usage technique : réflexions théoriques. *Les Enjeux de l'Information et de la Communication*, 2006(1):13–22.
- BOUMAZA, M. et CAMPANA, A. (2007). Enquêter en milieu « difficile ». *Revue Française de Science Politique*, 57(1):5–25.

- BOUNEGRU, L. (2013). Le datajournalisme en perspective. In STRAY, J., BOUNEGRU, L., CHAMBERS, L. et KAYSER-BRIL, N., éditeurs : *Guide du datajournalisme : Collecter, analyser et visualiser les données*, pages 23–28. Eyrolles, Paris.
- BOUTET, A. et TRÉMENBERT, J. (2009). Mieux comprendre les situations de non-usages des TIC. le cas d'internet et de l'informatique. *Les Cahiers du numérique*, 5(1):69–100.
- BOYDENS, I. (1999). *Informatique, normes et temps : évaluer et améliorer la qualité de l'information : les enseignements d'une approche herméneutique appliquée à la base de données "LATG" de l'O.N.S.S.* Academia Bruylant, Louvain-la-Neuve.
- BOYDENS, I. (2010). Hiérarchie et anarchie. *Les Cahiers du Numérique*, 6(3):77–101.
- BOYDENS, I. (2012). L'océan des données et le canal des normes. *Les Annales des Mines*, 67:22–29.
- BOYDENS, I. (2014a). Dix bonnes pratiques pour améliorer et maintenir la qualité des données. (En ligne, site consulté le 12/03/2016). <https://www.smalsresearch.be/dix-bonnes-pratiques-pour-ameliorer-et-maintenir-la-qualite-des-donnees/>.
- BOYDENS, I. (2014b). Open data et e-government. *Research Note 33, Smals Research*.
- BOYDENS, I. et van HOOLAND, S. (2011). Hermeneutics applied to the quality of empirical databases. *Journal of Documentation*, 67(2):279–289.
- BOYER, D. et HANNERZ, U. (2006). Introduction : Worlds of journalism. *Ethnography*, 7:5–17.
- BOZDAG, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, 15(3):209–227.
- BRADSHAW, P. (2015). Data journalism. In ZION, L. et CRAIG, D., éditeurs : *Ethics for Digital Journalists : Emerging Best Practices*, pages 202–219. Routledge, Londres.
- BRADSHAW, P. (2018). Data journalism teaching, fast and slow. *Asia Pacific Media Educator*, 28(1):55–66.
- BRASSEUR, M. (2012). L'interaction du chercheur avec son terrain en recherche-action : deux cas d'accompagnement individuel des managers. *Recherches en Sciences de Gestion*, (2):103–118.
- BRODIE, M. L. (1980). Data quality in information systems. *Information & Management*, 3(6): 245–258.
- BRYNJOLFSSON, E. et MCAFEE, A. (2012). *Race against the machine : How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming Employment and the economy*. Digital Frontier Press, Lexington, Massachusetts.
- BURRELL, J. (2016). How the machine 'thinks' : Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1):1–12.

- BUTON, F. et MARIOT, N. (2006). *Socio-histoire du dictionnaire des idées*. 2e volume de la collection des "Notionnaires". Encyclopaedia Universalis, Paris.
- CARADEC, V. (2001). "Personnes âgées" et "objets technologiques" : Une perspective en termes de logiques d'usage. *Revue française de sociologie*, 1(41):117–148.
- CARLSON, M. (2015). The robotic reporter : Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, 2(4):1–16.
- CARLSON, M. (2019). News algorithms, photojournalism and the assumption of mechanical objectivity in journalism. *Digital Journalism*, 7(8):1117–1133.
- CASWELL, D. et DÖRR, K. (2018). Automated journalism 2.0 : Event-driven narratives : From simple descriptions to real stories. *Journalism Practice*, 12(4):477–496.
- CATENACCIO, P., COTTER, C., DE SMEDT, M., GARZONE, G., JACOBS, G., MACGILCHRIST, F., LAMS, L., PERRIN, D., RICHARDSON, J. E., VAN HOUT, T. *et al.* (2011). Towards a linguistics of news production. *Journal of Pragmatics*, 43(7):1843–1852.
- CEFAÏ, D. (1996). La construction des problèmes publics. définitions de situations dans des arènes publiques. *Réseaux*, 14(75):43–66.
- CHALABY, J. K. (1996). Journalism as an anglo-american invention : A comparison of the development of french and anglo-american journalism, 1830s-1920s. *European Journal of Communication*, 11(3):303–326.
- CHAMBAT, P. (1994). Usages des technologies de l'information et de la communication (TIC) : évolution des problématiques. *Technologies de l'information et société*, 6(3):249–270.
- CHARAUDEAU, P. (1995). Une analyse sémiolinguistique du discours. *Langages*, 17:96–111.
- CHARAUDEAU, P. (2006). Discours journalistique et positionnements énonciatifs. frontières et dérives. *Semen. Revue de sémio-linguistique des textes et discours*, 22:1166–1177.
- CHARAUDEAU, P., MAINGUENEAU, D. et ADAM, J. (2002). *Dictionnaire d'analyse du discours*. Le Seuil, Paris.
- CHARETTE, R. N. (2005). Why software fails. *IEEE spectrum*, 42(9):36.
- CHAVAND, F. (2017). *Des données à l'information*. ISTE Editions, Londres.
- CHENG, W. (2013). Corpus-based linguistic approaches to critical discourse analysis. In CHAPPELLE, C., éditeur : *The encyclopedia of applied linguistics*, page 1353–1360. Blackwell, Oxford.
- CHIANG, F. et MILLER, R. J. (2008). Discovering data quality rules. *Proceedings of the VLDB Endowment*, 1(1):1166–1177.
- CHRISTIN, A. (2016). Is journalism a transnational field? Asymmetrical relations and symbolic domination in online news. *The Sociological Review Monographs*, 64(2):212–234.

- CHUNG, D. S. et NAH, S. (2014). Negotiating journalistic professionalism : A case study of Ohmynews in a participatory media climate. *Journalism Practice*, 8(4):390–406.
- CLERWALL, C. (2014). Enter the robot journalist. *Journalism Practice*, 8(5):519–531.
- CODDINGTON, M. (2015). Clarifying journalism's quantitative turn : A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital journalism*, 3(3):331–348.
- COHEN, S. et ALLI (2011). Computational journalism. *Communications of the ACM*, 54(10):66–71.
- COLPAERT, P., JOYE, S., MECHANT, P., MANNENS, E. et Van de WALLE, R. (2013). The 5 stars of open data portals. In METTEG13, éditeur : *Proceedings of the 7th International Conference on Methodologies, Technologies and Tools Enabling e-Government*, pages 61–67, University of Vigo, Espagne.
- COMTET, I. (2009). Entre usage professionnel des TIC et structure organisationnelle : la capacité au bricolage comme compétence adaptative. *Études de communication*, 2:119–134.
- CORNU, D. (2009). *Journalisme et vérité : l'éthique de l'information au défi du changement médiatique*. Le champ éthique. Labor et Fides, Genève.
- COTTLE, S. et ASHTON, M. (1999). From BBC newsroom to BBC newscentre : On changing technology and journalist practices. *Convergence*, 5(3):22–43.
- COURMONT, A. (2015). Open data et recomposition du gouvernement urbain : de la donnée comme instrument à la donnée comme enjeu politique. *Informations sociales*, 5:40–50.
- COUTANT, A. (2015). Les approches sociotechniques dans la sociologie des usages en SIC. *Revue française des sciences de l'information et de la communication*, 6:<http://journals.openedition.org/rfsic/1271>.
- COX, G. et MCLEAN, C. A. (2013). *Speaking code : coding as aesthetic and political expression*. MIT Press, Cambridge, Massachusetts.
- COX, M. (2000). The Development of Computer-assisted Reporting. *Paper presented at AEJMC 2000, Phoenix, AZ, August 9–12*.
- CRISP (1959). La presse financière en BELGIQUE. *Courrier Hebdomadaire*, 35:1–11.
- CURTIN, P. A. et MAIER, S. R. (2001). Numbers in the newsroom : A qualitative examination of a quantitative challenge. *Journalism & Mass Communication Quarterly*, 78(4):720–738.
- DAEWON, K. et SEONGCHEOL, K. (2017). Newspaper journalists' attitudes towards robot journalism. *Telematics and Informatics*, 2(35):340–357.
- DAGIRAL, E. et PARASIE, S. (2010). Presse en ligne : où en est la recherche? *Réseaux*, 2:13–42.
- DAGIRAL, E. et PARASIE, S. (2011). Portrait du journaliste en programmeur. l'émergence d'une figure du journaliste hacker. *Les Cahiers du Journalisme*, 22(23):144–154.

- DAGIRAL, E. et PARASIE, S. (2012). Data-driven journalism and the public good : "computer-assisted-reporters" and "programmer-journalists" in chicago. *New Media Society*, 15(6):853–871.
- DAGIRAL, E. et PARASIE, S. (2013). Des journalistes enfin libérés de leurs sources? promesse et réalité du "journalisme de données". *Sur le journalisme-About journalism-Sobre jornalismo*, 1(2):52–63.
- DALE, R. (1995). An introduction to natural language generation. *European Summer School in Logic, Language and Information, ESSLLI'95*.
- DALE, R. et WHITE, M. (2007). Shared tasks and comparative evaluation in natural language generation. *In Proceedings of the Workshop on Shared Tasks and Comparative Evaluation in Natural Language Generation*, pages 1–6.
- DANIEL, A., FLEW, T. et SPURGEON, C. (2010). The promise of computational journalism. *In Media, Democracy and Change : Refereed Proceedings of the Australian and New Zealand Communications Association Annual Conference*, pages 1–19, Canberra. Australia and New Zealand Communication Association.
- DANLOS, L. (1991). Génération automatique de textes en langue naturelle. *Linx*, 4(1):197–214.
- DANLOS, L. et ROUSSARIE, L. (2000). Ingénierie des langues. *In DANLOS, L., éditeur : La génération automatique de textes*, page 354. Hermès.
- DAVIS, M. (1987). The moral authority of a professional code. *Nomos*, 29:302–337.
- DAVIS, M. (2012). "Ain't no one here but us social forces" : Constructing the professional responsibility of engineers. *Science and Engineering Ethics*, 18(1):13–34.
- de BENS, E. (1973). *De Belgische dagbladpers onder Duitse censuur (1940-1944)*. De Nederlandsche Boekhandel.
- DE BENS, E. (2001). *De pers in België : het verhaal van de Belgische dagbladpers : gisteren, vandaag en morgen*. Lannoo Campus.
- de BURGH, H. (2008). *Investigative Journalism*. Routledge, Londres.
- DE MAEYER, J. et LE CAM, F. (2015). The material traces of journalism : A socio-historical approach to online journalism. *Digital journalism*, 3(1):85–100.
- DE MAEYER, J., LIBERT, M., DOMINGO, D., HEINDERYCKX, F. et LE CAM, F. (2015). Waiting for data journalism : A qualitative assessment of the anecdotal take-up of data journalism in french-speaking belgium. *Digital Journalism*, 3(3):432–446.
- DE MAEYER, J. et MALCORPS, S. (2015). Des controverses sociotechniques aux discours méta-journalistiques. *Hermès, La Revue*, 3:171–180.

- de VILLERS, M.-E. (2001). Analyse linguistique d'un titre de presse : illustration d'une norme. In RAYMOND, D. et LAFRANCE, A., éditeurs : *Norme et médias*, pages 97–98. Les publications du Québec, Terminogramme, Québec.
- DEFLEUR, M. H. (2013). *Computer assisted investigative reporting : Development and methodology*. Routledge, Londres.
- DEGENNE, A. et VERGÈS, P. (1973). Introduction à l'analyse de similitude. *Revue française de sociologie*, 4(14):471–512.
- DELFORCE, B. (1996). La responsabilité sociale du journaliste : donner du sens. *Les Cahiers du Journalisme*, 2:16–32.
- DEMEULENAERE, P. (2012). L'interprétation des terrains et le recours à des propositions théoriques. *L'Année sociologique*, 62(1):67–91.
- DENIS, J. (2009). Une autre sociologie des usages? Pistes et postures pour l'étude des chaînes sociotechniques. In *Article de synthèse pour le programme TIC & Migrations*. MSH Paris.
- DESROSIÈRES, A. (2008a). *Pour une sociologie historique de la quantification : L'argument statistique*. Collection Sciences sociales. Presses des Mines, Paris.
- DESROSIÈRES, A. (2008b). La statistique, outil de gouvernement et outil de preuve. In DESROSIÈRE, A., éditeur : *Pour une sociologie historique de la quantification*, pages 7–19. Presses des Mines, Paris.
- DESROSIÈRES, A. (2008). *L'argument statistique : gouverner par les nombres*. Collection Sciences sociales. Presses des Mines, Paris.
- DEUZE, M. (2002). National news cultures : A comparison of Dutch, German, British, Australian, and US journalists. *Journalism & Mass Communication Quarterly*, 79(1):134–149.
- DEUZE, M. (2003). The web and its journalisms : Considering the consequences of different types of newsmedia online. *New Media & Society*, 5(2):203–230.
- DEUZE, M. (2005). What is journalism? Professional identity and ideology of journalists reconsidered. *Journalism*, 6(4):442–464.
- DEUZE, M. et BARDOEL, J. (2001). Network journalism : Converging competences of media professionals and professionalism. *Australian Journalism Review*, 23(1):91–103.
- DEUZE, M. et WITSCHGE, T. (2018). Beyond journalism : Theorizing the transformation of journalism. *Journalism*, 19(2):165–181.
- DEVELOTTE, C. et RECHNIEWSKI, E. (2001). Discourse analysis of newspaper headlines : a methodological framework for research into national representations. *The Web Journal of French Media Studies*, 4(1):1–12.
- DIAKOPOULOS, N. (2015). Algorithmic accountability : Journalistic investigation of computational power structures. *Digital Journalism*, 3(3):398–415.

- DIAKOPOULOS, N. (2019). *Automating the news : How algorithms are rewriting the media*. Harvard University Press, Cambridge, Massachusetts.
- DIAKOPOULOS, N. (2019b). Towards a design orientation on algorithms and automation in news production. *Digital Journalism*, 7(8):1180–1184.
- DIERICKX, L. (2015). Possibilités et limites de la génération automatique de textes dans un contexte journalistique. *Master en Sciences et Technologies de l'Information et de la Communication, Université Libre de Bruxelles*.
- DIERICKX, L. (2018). News bot for the newsroom : How building data quality indicators can support journalistic projects relying on real-time open data. *In Global Investigative Journalism Conference 2017 Academic Tracks*. Investigative Journalism Education Consortium, Wits University, 16-19 November, Johannesburg.
- DIERICKX, L. (2020). Journalists as end-users : Quality management principles applied to the design process of news automation. *First Monday*, 25(4).
- DOMINGO, D. (2008). Interactivity in the daily routines of online newsrooms : Dealing with an uncomfortable myth. *Journal of Computer-Mediated Communication*, 13(3):680–704.
- DOMINGO, D. et LE CAM, F. (2014). Journalism in dispersion : Exploring the blurring boundaries of newsmaking through a controversy. *Digital Journalism*, 2(3):310–321.
- DOMINGO, D. et PATERSON, C. (2008). *Making online news : The ethnography of new media production*. Peter Lang Publishing, Oxford.
- DONSBACH, W. et KLETT, B. (1993). Subjective objectivity. how journalists in four countries define a key term of their profession. *Gazette (Leiden, Netherlands)*, 51(1):53–83.
- DOR, D. (2003). On newspaper headlines as relevance optimizers. *Journal of Pragmatics*, 35(5): 695–721.
- DORAY, P. (2015). Construction sociale des technologies. *In PRUD'HOMME, J., DORAY, P. et BOUCHARD, E., éditeurs : Sciences, technologies et sociétés de A à Z*, pages 57–61. Presses de l'Université de Montréal, Montréal.
- DÖRR, K. N. (2016). Mapping the field of algorithmic journalism. *Digital Journalism*, 4(6):700–722.
- DÖRR, K. N. et HOLLNBUCHNER, K. (2016). Ethical challenges of algorithmic journalism. *Digital Journalism*, 5(4):404–419.
- DU BOIS, J. W. (2007). The stance triangle. *Stancetaking in discourse : Subjectivity, evaluation, interaction*, 164(3):139–182.
- DUMONT, J., GREVISSE, B. et RINGLET, G. (1998). *La presse écrite en Belgique*. Espace Nord. Éditions Labor, Charleroi.
- DURKHEIM, E. (2013). *Professional ethics and civic morals*. Routledge, Londres.

- DUTTON, W. H. et KRAEMER, K. L. (1980). Automating bias. *Society*, 17(2):36–41.
- ECKERSON, W. W. (2002). Data quality and the bottom line. *TDWI Report, The Data Warehouse Institute*.
- EKBIA, H. et NARDI, B. (2014). Heteromation and its (dis) contents : The invisible division of labor between humans and machines. *First Monday*, 19(6).
- EKDALE, B., SINGER, J. B., TULLY, M. et HARMSSEN, S. (2015). Making change : Diffusion of technological, relational, and cultural innovation in the newsroom. *Journalism & Mass Communication Quarterly*, 92(4):938–958.
- ELIAS, N. (1993). *Engagement et distanciation : contributions à la sociologie de la connaissance*. Fayard, Paris.
- ELLUL, J. (1954). *La technique ou l'enjeu du siècle*. Sciences politiques. Armand Colin, Paris.
- ENGLBRETSON, R. (2007). *Stancetaking in discourse : Subjectivity, evaluation, interaction*, volume 164. John Benjamins Publishing, Amsterdam.
- ESPELAND, W. N. et STEVENS, M. L. (2008). A sociology of quantification. *European Journal of Sociology/Archives Européennes de Sociologie*, 49(3):401–436.
- FAIRCLOUGH, N. (2003). *Analysing discourse : Textual analysis for social research*. Routledge, Londres.
- FANTA, A. (2017). Putting europe's robots on the map : Automated journalism in news agencies. *Reuters Institute Fellowship Paper, Reuters Institute for the Study of Journalism, University of Oxford*, pages 2017–09.
- FEENBERG, A. (1995). *Alternative modernity : The technical turn in philosophy and social theory*. University of California Press, Oakland, California.
- FINK, K. (2019). The biggest challenge facing journalism : A lack of trust. *Journalism*, 20(1):40–43.
- FISH, S. E. (1976). How to do things with Austin and Searle : Speech act theory and literary criticism. *MLN*, 91(5):983–1025.
- FLANAGIN, A. J., FLANAGIN, C. et FLANAGIN, J. (2010). Technical code and the social construction of the internet. *New Media & Society*, 12(2):179–196.
- FLESCH, R. (1948). A new readability yardstick. *Journal of applied psychology*, 32(3):221.
- FLICHY, P. (1995). *L'innovation technique : récents développements en sciences sociales, vers une nouvelle théorie de l'innovation*. Sciences et société. La Découverte, Paris.
- FLICHY, P. (2001). La place de l'imaginaire dans l'action technique. *Réseaux*, 109:52–73.
- FLICHY, P. (2008). Technique, usage et représentations. *Réseaux*, 2:147–174.

- FLICHY, P. et PARASIE, S. (2013). Sociologie des bases de données. *Réseaux*, 2(3):178–179.
- FLYVBJERG, B. (2006). Five misunderstandings about case-study research. *Qualitative inquiry*, 12(2):219–245.
- FOOT, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5:5–15.
- FORD, M. (2015). *The rise of the robots : Technology and the threat of mass unemployment*. Oneworld Publications, Londres.
- FOUCART, T. (2001). L'interprétation des résultats statistiques. *Mathématiques et sciences humaines. Mathematics and social sciences*, 153:21–28.
- FOX, C., LEVITIN, A. et REDMAN, T. (1994). The notion of data and its quality dimensions. *Information processing & management*, 30(1):9–19.
- FOX, P. (2014). Data life cycle : Introduction, definitions and considerations. *Rensselaer Polytechnic Institute, 25 September 2014, New York*.
- FRIEDMAN, B. et NISSENBAUM, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, 14(3):330–347.
- FRIENDLY, M. (2008). The golden age of statistical graphics. *Statistical Science*, 23(4):502–535.
- FULK, J. (1993). Social construction of communication technology. *Academy of Management journal*, 36(5):921–950.
- FULLER, M. (2003). *Behind the blip : Essays on the culture of software*. Autonomedia, New York.
- FULLER, M., CUBITT, S. et MALINA, R. (2008). *Software Studies : A lexicon*. EBSCO Ebook Academic Collection. MIT Press, Cambridge, Massachusetts.
- FULLER, M. et GOFFEY, A. (2012). *Evil Media*. MIT Press, Cambridge, Massachusetts.
- FÜRBER, C. et HEPP, M. (2010). Using semantic web resources for data quality management. *Knowledge Engineering and Management by the Masses*, 6317:211–225.
- GABIELKOV, M., RAMACHANDRAN, A., CHAINTREAU, A. et LEGOUT, A. (2016). Social clicks : What and who gets read on twitter? *ACM SIGMETRICS Performance Evaluation Review*, 44(1): 179–192.
- GANGLBAUER, E., SCHRAMMEL, J., DEUTSCH, S. et TSCHELIGI, M. (2009). Applying psychophysiological methods for measuring user experience : possibilities, challenges and feasibility. *In Workshop on user experience evaluation methods in product development*. Citeseer.
- GARRIC, N. (2013). *Introduction à la linguistique*. HU Linguistique. Hachette, Paris.
- GARRISON, B. (2001). Diffusion of online information technologies in newspaper newsrooms. *Journalism*, 2(2):221–239.

- GEIGER, R. S. (2014). Bots, bespoke, code and the materiality of software platforms. *Information, Communication & Society*, 17(3):342–356.
- GESLIN, P. *et al.* (2002). Les formes sociales d'appropriations des objets techniques, ou le paradigme anthropotechnologique. *Ethnographiques*, 1.
- GHEZZI, C., JAZAYERI, M. et MANDRIOLI, D. (2002). *Fundamentals of software engineering*. Prentice Hall, Upper Saddle River, New Jersey.
- GILLESPIE, T. (2014). Relevance of the algorithms. In TARLETON (IN), G., P.J., B. et K.A., F., éditeurs : *Media Technologies : Essays on Communication, Materiality, and Society*, Inside Technology, pages 167–194. MIT Press, Cambridge, Massachusetts.
- GINGRAS, A.-M. (2009). *Médias et démocratie : le grand malentendu*. Presses de l'Université du Québec, Québec.
- GOMAA, W. H. et FAHMY, A. A. (2013). A survey of text similarity approaches. *International Journal of Computer Applications*, 68(13):13–18.
- GOËTA, S. et MABI, C. (2014). L'open data peut-il (encore) servir les citoyens? *Mouvements*, 3:81–91.
- GOYETTE, G. et LESSARD-HÉBERT, M. (1987). *La recherche-action : ses fonctions, ses fondements et son instrumentation*. Presses de l'Université du Québec, Québec.
- GRAEFE, A. (2016). Guide to automated journalism. *Tow Center for Digital Journalism*, DOI : 10.7916/D80G3XDJ.
- GRAEFE, A., HAIM, M., HAARMANN, B. et BROSIUS, H.-B. (2015). Perception of automated computer-generated news : credibility, expertise, and readability. *11th Dubrovnik Media Days, 15 September 2015, Dubrovnik*.
- GRANJON, F. et DENOUEL, J. (2011). Penser les usages sociaux des technologies numériques d'information et de communication. In *Communiquer à l'ère numérique. Regards croisés sur la sociologie des usages*, pages 7–43. Presses des Mines, Paris.
- GRAY, J. (2014). Towards a genealogy of open data. In *General Conference of the European Consortium for Political Research, 8 September 2014, Glasgow*.
- GRAY, J. (2018). Computational imaginaries : Some further remarks on Leibniz, Llull, and rethinking the history of calculating machines. In VEGA, A., WEIBEL, P. et ZIELINSKI, S., éditeurs : *DIA-LOGOS : Ramon Llull's Method of Thought and Artistic Practice*, pages 321–328. University of Minnesota Press, Minneapolis.
- GREVISSE, B. (1998). Introduction. autorégulation ou déontologie? Les conditions d'un débat sur les pratiques journalistiques. *Recherches en communication*, 9(9):5–24.
- GREVISSE, B. (2003). Légimité, éthique et déontologie. *Hermès, La Revue*, 1:223–230.

- GULER, I., GUILLÉN, M. F. et MACPHERSON, J. M. (2002). Global competition, institutions, and the diffusion of organizational practices : The international spread of ISO 9000 quality certificates. *Administrative science quarterly*, 47(2):207–232.
- GYNNILD, A. (2014). Journalism innovation leads to innovation journalism : The impact of computational exploration on changing mindsets. *Journalism*, 15(6):713–730.
- HABASH, N. (2004). The use of a structural n-gram language model in generation-heavy hybrid machine translation. In *Natural Language Generation*, pages 61–69. Springer, Cham.
- HAFEZ, K. (2002). Journalism ethics revisited : A comparison of ethics codes in europe, north africa, the middle east, and muslim asia. *Political communication*, 19(2):225–250.
- HAIM, M. et GRAEFE, A. (2017). Automated news : Better than expected? *Digital Journalism*, 5(8):1044–1059.
- HAINAUT, J.-L. (2012). *Bases de données : concepts, utilisation et développement*. Dunod, Malakoff.
- HALLIN, D. C. et MANCINI, P. (2004). *Comparing media systems : Three models of media and politics*. Cambridge University Press, Cambridge.
- HAMILTON, J. T. et TURNER, F. (2009). Accountability through algorithm : Developing the field of computational journalism. In *Report from the Center for Advanced Study in the Behavioral Sciences, Stanford, Summer Workshop*, pages 27–41.
- HAMMOND, P. (2017). From computer-assisted to data-driven : Journalism and big data. *Journalism*, 18(4):408–424.
- HANITZSCH, T. (2006). Mapping journalism culture : A theoretical taxonomy and case studies from indonesia. *Asian Journal of Communication*, 16(02):169–186.
- HANITZSCH, T. (2007). Deconstructing journalism culture : Toward a universal theory. *Communication theory*, 17(4):367–385.
- HANSEN, M., ROCA-SALES, M., KEEGAN, J. M. et KING, G. (2017). Artificial intelligence : Practice and implications for journalism. *Tow Center for Digital Journalism, Columbia University*, DOI : 10.7916/d8x92prd.
- HARDT, H. (1990). Newswriters, technology, and journalism history. *Critical Studies in Media Communication*, 7(4):346–365.
- HARRIS, C. E., DAVIS, M., PRITCHARD, M. S. et RABINS, M. J. (1996). Engineering ethics : What? why? how? and when? *Journal of Engineering Education*, 85:93–96.
- HARRIS, Z. S. et DUBOIS-CHARLIER, F. (1969). Analyse du discours. *Langages*, 13:8–45.
- HAUG, A., STENTOFT ARLBJORN, J. et PEDERSEN, A. (2009). A classification model of ERP system data quality. *Industrial Management & Data Systems*, 109(8):1053–1068.

- HAUG, A., ZACHARIASSEN, F. et VAN LIEMPD, D. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management*, 4(2):168–193.
- HEIKKILÄ, H., DOMINGO, D., PIES, J., GLOWACKI, M., KUS, M. et BAISSNÉE, O. (2012). Media accountability goes online. *A Transnational Study no Emerging Practices and Innovations (Media ACT Working paper)*. Journalism Research and Development Centre and Universidad de Tampere, Tampere.
- HENNINGER, M. (2013). Data-driven journalism. *Reassessing Journalism*, 157:157–184.
- HERMANS, L., VERGEER, M. et D'HAENENS, L. (2009). Internet in the daily life of journalists : Explaining the use of the internet by work-related characteristics and professional opinions. *Journal of Computer-Mediated Communication*, 15(1):138–157.
- HERRERO CECILIA, J. (2007). Syntaxe, sémantique et pragmatique des titres des nouvelles de la presse française construits en forme de phrase nominale ou averbale : aspects cognitifs et communicatifs. In BONNET, D., CHAVES GARCÍA, M. J. et DUCHÊNE, N., éditeurs : *Littérature, langages et arts : rencontres et création*, page 97. Universidad de Huelva.
- HIMMA, K. E. et TAVANI, H. T. (2008). *The handbook of information and computer ethics*. John Wiley & Sons, Hoboken, New Jersey.
- HO-DAC, L.-M., JACQUES, M.-P. et REBEYROLLE, J. (2004). Sur la fonction discursive des titres. In PORHIEL, S. et KLINGLER, D., éditeurs : *L'unité texte*, pages 125–152. Perspectives, Pleyben.
- HORVAT, N. (2013). Le regard critique en situation d'immersion. Observation in situ d'un dispositif de vidéosurveillance. *Socio-Anthropologie*, 27:59–69.
- HOWARD, A. (2014). The art and science of data-driven journalism (report). *Tow Center for Digital Journalism*, DOI; 10.7916/D8Q531V1.
- HUH, Y., KELLER, F., REDMAN, T. C. et WATKINS, A. (1990). Data quality. *Information and Software Technology*, 32(8):559–565.
- JANSSEN, M., CHARALABIDIS, Y. et ZUIDERWIJK, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4):258–268.
- JANSSEN, N. A., HOEK, G., SIMIC-LAWSON, M., FISCHER, P., VAN BREE, L., TEN BRINK, H., KEUKEN, M., ATKINSON, R. W., ANDERSON, H. R., BRUNEKREEF, B. *et al.* (2011). Black carbon as an additional indicator of the adverse health effects of airborne particles compared with pm10 and pm2.5. *Environmental health perspectives*, 119(12):1691–1699.
- JARRIGE, F. (2014). *Technocritiques : Du refus des machines à la contestation des technosciences*. Hors collection Sciences Humaines. La Découverte, Paris.
- JOURÉGUIBERRY, F. (2008). De l'usage des technologies de l'information et de la communication comme apprentissage créatif. *Education et sociétés*, 2:29–42.

- JOURÉGUIBERRY, F. (2012). Retour sur les théories du non-usage des technologies de communication. In PROULX, S. et KLEIN, A., éditeurs : *Connexions : communication numérique et lien social*, pages 335–350. Presses Universitaires de Namur, Namur.
- JEANNERET, Y. (2007). Usages de l'usage, figures de la médiatisation. *Communication & langages*, 151(1):3–19.
- JOANNES, A. (2010). *Bases de données et visualisation de l'information*. CFPJ Editions, Paris.
- JOUËT, J. (1993). Pratiques de communication et figures de la médiation. *Réseaux*, 11(60):99–120.
- JOUËT, J. (2000). Retour critique sur la sociologie des usages. *Réseaux*, 18(100):487–521.
- JOUËT, J. et SFEZ, L. (1993). Usages et pratiques des nouveaux outils de communication. *Dictionnaire critique de la communication*, 1:371–376.
- JOUISSON-LAFFITTE, E. (2009). La recherche action : oubliée de la recherche dans le domaine de l'entrepreneuriat. *Revue de l'Entrepreneuriat*, 8(1):1–35.
- JUNG, T., KASS, C., SCHRAMM, T. et ZAPF, D. (2017). So what really is user experience? An experimental study of user needs and emotional responses as underlying constructs. *Ergonomics*, 60(12):1601–1620.
- JURAN, J. M. (1999). How to think about quality. In JURAN, J. M., GODFREY, A. B., HOOGSTOEL, R. E. et G., S. E., éditeurs : *Quality-Control Handbook*. McGraw-Hill, New York.
- KAMMER, A. (2013). The mediatization of journalism. *MedieKultur : Journal of media and communication research*, 54:141–158.
- KARAPANOS, E. (2013). User experience over time. In KARAPANOS, E., éditeur : *Modeling users' experiences with interactive systems*, pages 57–83. Springer, Cham.
- KARLSEN, J. et STAVELIN, E. (2014). Computational journalism in Norwegian newsrooms. *Journalism practice*, 8(1):34–48.
- KELLNER, C., MASSOU, L. et MORELLI, P. (2010). (re) penser le non-usage des TIC. *Questions de communication*, 18.
- KINCAID, J. P., FISHBURNE JR, R. P., ROGERS, R. L. et CHISSOM, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Rapport technique, Naval Technical Training Command Millington TN Research Branch.
- KITCHIN, R. et DODGE, M. (2011). *Code/space : Software and everyday life*. MIT Press, Cambridge, Massachusetts.
- KLEIN, H. K. et KLEINMAN, D. L. (2002). The social construction of technology : Structural considerations. *Science, Technology & Human Values*, 27(1):28–52.

- KLONGLAN, G. E. et COWARD, E. W. (1970). The concept of symbolic adoption : A suggested interpretation. *Rural Sociology*, 35(1):77–83.
- KOSTOFF, R. N., BOYLAN, R. et SIMONS, G. R. (2004). Disruptive technology roadmaps. *Technological Forecasting and Social Change*, 71(1):141–159.
- KOWALSKI, R. (1979). Algorithm= logic+ control. *Communications of the ACM*, 22(7):424–436.
- KRAEMER, F., van OVERVELD, K. et PETERSON, M. (2011). Is there an ethics of algorithms? *Ethics and information technology*, 13(3):251–260.
- KRONROD, A. et ENGEL, O. (2001). Accessibility theory and referring expressions in newspaper headlines. *Journal of Pragmatics*, 33(5):683–699.
- LAKOFF, G. et JOHNSON, M. (1980). Conceptual metaphor in everyday language. *The journal of Philosophy*, 77(8):453–486.
- LAPASSADE, G. (2017). Observation participante. In ENRIQUEZ, E., LEVY, A. et BARUS-MICHEL, J., éditeurs : *Vocabulaire de psychosociologie*, pages 392–407. Erès.
- LATAR, N. L. (2015). The robot journalist in the age of social physics : the end of human journalism? In EINAÏ, G., éditeur : *The New World of Transitioned Media*, pages 65–80. Springer, Cham.
- LATAR, N. L. (2018). *Robot Journalism : Can human journalism survive?* World Scientific, Singapore.
- LATOUR, B. (2000). La fin des moyens. *Réseaux*, 18(100):39–58.
- LATOUR, B. (2005). *Reassembling the social : An introduction to Actor-Network-Theory*. Clarendon Lectures in Management Studies. Oxford University Press.
- LATZKO-TOTH, G. (2009). L'étude de cas en sociologie des sciences et des techniques. *Centre interuniversitaire de la recherche sur la science et la technologie, Montréal*, Notes de recherche.
- LATZKO-TOTH, G. et PROULX, S. (2015). Appropriation des technologies. In PRUD'HOMME, J., DORAY, P. et BOUCHARD, F., éditeurs : *Sciences, technologies et sociétés de A à Z*, pages 24–26. Presses de l'Université de Montréal, Montréal.
- LAUK, E. et KUŠ, M. (2012). Editors' introduction : Media accountability, between tradition and innovation. *Central European Journal of Communication*, 5(2):168–174.
- LAVIE, T., SELA, M., OPPENHEIM, I., INBAR, O. et MEYER, J. (2010). User attitudes towards news content personalization. *International journal of human-computer studies*, 68(8):483–495.
- LE CAM, F. (2006). États-Unis : les weblogs d'actualité ravivent la question de l'identité journalistique. *Réseaux*, 138:139–158.
- LECOMPTE, C. (2015). Automation in the newsroom. *Nieman Reports*, 3(69):32–45.

- LEON, S. (2018). Accounting for methods in data journalism : Spreadsheets, scripts and programming notebooks. In BOUNEGRU, L. et STRAY, J., éditeurs : *The Data Journalism Handbook 2. Towards a Critical Data Practice*, volume <https://datajournalism.com/read/handbook/two>. European Journalism Centre.
- LEONARD, T. (1992). Databases in the newsroom : Computer-assisted reporting. *Online*, 16(3): 62–65.
- LEPPÄNEN, L., MUNEZERO, M., SIRÉN-HEIKEL, S., GRANROTH-WILDING, M. et TOIVONEN, H. (2017). Finding and expressing news from structured data. In *Proceedings of the 21st International Academic Mindtrek Conference, 20-21 September 2017, Tampere*, pages 174–183. ACM.
- LESAGE, F. et HACKETT, R. A. (2014). Between objectivity and openness : The mediality of data for journalism. *Media and Communication*, 2(2):42–54.
- LETEINTURIER, C. (2016). Continuité/discontinuité des carrières des journalistes français encartés. Etude de deux cohortes de nouveaux titulaires de la carte de presse. *Recherches en Communication*, 43(43):27–57.
- LÉVY, P. (2017). La pyramide algorithmique. *Sens Public*, 29 | <http://sens-public.org/article1275.html>.
- LEWIS, S. C. (2012). The tension between professional control and open participation : Journalism and its boundaries. *Information, Communication & Society*, 15(6):836–866.
- LEWIS, S. C., SANDERS, A. K. et CARMODY, C. (2019). Libel by algorithm? Automated journalism and the threat of legal liability. *Journalism & Mass Communication Quarterly*, 96(1):60–81.
- LEWIS, S. C. et USHER, N. (2013). Open source and journalism : Toward new frameworks for imagining news innovation. *Media Culture Society*, 35(5):4–9.
- LEWIS, S. C. et USHER, N. (2014). Code, collaboration, and the future of journalism. a case study of the hacks/hackers global network. *Digital Journalism*, 2(3):383–393.
- LEWIS, S. C. et WESTLUND, O. (2015a). Actors, actants, audiences, and activities in cross-media news work : A matrix and a research agenda. *Digital Journalism*, 3(1):19–37.
- LEWIS, S. C. et WESTLUND, O. (2015b). Big data and journalism : Epistemology, expertise, economics, and ethics. *Digital Journalism*, 3(3):447–466.
- LINDEN, C.-G. (2017). Decades of automation in the newsroom : Why are there still so many jobs in journalism? *Digital Journalism*, 5(2):123–140.
- LINDEN, C.-G. (2017b). Algorithms for journalism. *The Journal of Media Innovations*, 4(1):60–76.
- LINDEN, T. C.-G. et DIERICKX, L. (2019). Robot journalism : The damage done by a metaphor. *Unmediated : Journal of Politics and Communication*, 2:152–155.

- LINDEN, T. C.-G., TUULONEN, H. E., BÄCK, A., DIAKOPOULOS, N., GRANROTH-WILDING, M., HAAPANEN, L., LEPPÄNEN, L. J., MELIN, M., MORING, T. A., MUNEZERO, M. D. *et al.* (2019). News automation : The rewards, risks and realities of 'machine journalism'. *WAN-IFRA Report*.
- LIU, B. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2(2010):627–666.
- LO, C. K., WIENGARTEN, F., HUMPHREYS, P., YEUNG, A. C. et CHENG, T. (2013). The impact of contextual factors on the efficacy of ISO 9000 adoption. *Journal of Operations Management*, 31(5):229–235.
- LOOSEN, W. (2018). Data-driven gold-standards : What the field values as award-worthy data journalism and how journalism co-evolves with the datafication of society. In BOUNEGRU, L. et STRAY, J., éditeurs : *The Data Journalism Handbook 2. Towards a Critical Data Practice*, volume <https://datajournalism.com/read/handbook/two>. European Journalism Centre.
- LOOSEN, W., REIMER, J. et DE SILVA-SCHMIDT, F. (2017). Data-driven reporting : An on-going (r) evolution? an analysis of projects nominated for the data journalism awards 2013-2016. *Journalism*, DOI : 10.1177/1464884917735691.
- LOWREY, W., BROUSSARD, R. et SHERRILL, L. A. (2019). Data journalism and black-boxed data sets. *Newspaper Research Journal*, 40(1):69–82.
- MADNICK, S. et ZHU, H. (2006). Improving data quality through effective use of data semantics. *Data & Knowledge Engineering*, 59(2):460–475.
- MAIER, S. R. (2000). The digital watchdog's first byte : Journalism's first computer analysis of public records. *American Journalism*, 17(4):75–91.
- MAIER, S. R. (2000b). Digital diffusion in newsrooms : The uneven advance of computer-assisted reporting. *Newspaper Research Journal*, 21(2):95–110.
- MAIER, S. R. (2003). Numeracy in the newsroom : A case study of mathematical competence and confidence. *Journalism & Mass Communication Quarterly*, 80(4):921–936.
- MANOVICH, L. (1999). Database as symbolic form. *Convergence : The International Journal of Research into New Media Technologies*, 5(2):80–99.
- MANOVICH, L. (2010). *Le langage des nouveaux médias*. Perceptions. Les Presses du réel, Dijon.
- MANOVICH, L. (2011). What's visualization? *Visual Studies*, 26(1):36–49.
- MANOVICH, L. (2013). *Software Takes Command*. International Texts in Critical Media Aesthetics. Bloomsbury Academic, Londres.
- MARCONI, F., SIEGMAN, A. et JOURNALIST, M. (2017). The future of augmented journalism : A guide for newsrooms in the age of smart machines. *The Associated Press Handbook*, <https://journalismai.com/2017/02/22/future-of-augmented-journalism-ap-2017/>.

- MARTIN, S. (2004). Vérité et objectivité journalistique : même contestation. *Les Cahiers du Journalisme*, 13:144–163.
- MASSIT-FOLLÉA, F. (2002). Usages des technologies de l'information et de la communication : acquis et perspectives de la recherche. *Le Français dans le Monde*, pages 8–14.
- MAURAND-VALET, A. (2007). Norme ISO 9000 et profil du responsable qualité. *Revue française de gestion*, 11:61–72.
- MCCBRIDE, R. E. D. (2016). The ethics of data journalism. *Digital Commons University of Nebraska-Lincoln*, 9:1–44.
- MCCALLUM, D. R. et PETERSON, J. L. (1982). Computer-based readability indexes. *In Proceedings of the ACM'82 Conference, New York*, pages 44–48. ACM.
- MCCALLUM, E. Q. (2012). *Bad data handbook : Cleaning up the data so you can get back to work*. O'Reilly Media, Farnham.
- MCCARTHY, J. et WRIGHT, P. (2007). *Technology as experience*. MIT Press, Cambridge, Massachusetts.
- MCCHESENEY, R. W. (2012). Farewell to journalism? Time for a rethinking. *Journalism Practice*, 6(5-6):614–626.
- MCCOSKER, A. et MILNE, E. (2014). Coding labour. *Cultural Studies Review*, 20(1):4–29.
- MCQUAIL, D. (2003). *Media accountability and freedom of publication*. Oxford University Press, Oxford.
- MELIN, M., BÄCK, A., SÖDERGÅRD, C., MUNEZERO, M. D., LEPPÄNEN, L. J. et TOIVONEN, H. (2018). No landslide for the human journalist. an empirical study of computer-generated election news in finland. *IEEE Access*, 6:43356–43367.
- MELLADO, C. (2015). Professional roles in news content : Six dimensions of journalistic role performance. *Journalism Studies*, 16(4):596–614.
- MERCIER, A. et PIGNARD-CHEYNEL, N. (2014). Mutations du journalisme à l'ère du numérique : un état des travaux. *Revue française des sciences de l'information et de la communication*, 5 | DOI : 10.4000/rfsic.1097.
- MEYER, P. (1973). *Precision journalism : A reporter's introduction to social science methods*. Indiana University Press, Bloomington.
- MEYER, P. (1991). *The new precision journalism*. Indiana University Press, Bloomington.
- MICÓ, J. L., MASIP, P. et DOMINGO, D. (2013). To wish impossible things* Convergence as a process of diffusion of innovations in an actor-network. *International Communication Gazette*, 75(1):118–137.
- MILLERAND, F. (2008). Usages des NTIC : les approches de la diffusion, de l'innovation et de l'appropriation (1ère partie). *Composite*, 2(1):1–19.

- MILOSAVLJEVIĆ, M. et VOBIĆ, I. (2019). 'Our task is to demystify fears' : Analysing newsroom management of automation in journalism. *Journalism*, DOI : 10.1177/1464884919861598.
- MINSKY, M. L. (1967). *Computation : finite and infinite machines*. Prentice Hall, Upper Saddle River, New Jersey.
- MOESCHLER, J. (1985). *Argumentation et conversation : éléments pour une analyse pragmatique du discours*. Hatier, Paris.
- MOIRAND, S. (2007). *Les discours de la presse quotidienne : observer, analyser, comprendre*. Linguistique Nouvelle Series. Presses Universitaires de France, Paris.
- MONTAL, T. et REICH, Z. (2017). I, Robot.You, Journalist. Who is the author? Authorship, by-lines and full disclosure in automated journalism. *Digital Journalism*, 5(7):829–849.
- MONTI, M. (2019). Automated journalism and freedom of information : Ethical and juridical problems related to ai in the press field. *Opinio Juris in Comparatione*, 1:139–155.
- MOODY, D. L. et SHANKS, G. G. (2003). Improving the quality of data models : empirical validation of a quality management framework. *Information systems*, 28(6):619–650.
- MOOR, J. H. (1985). What is computer ethics? *Metaphilosophy*, 16(4):266–275.
- MOOR, J. H. (1998). Reason, relativity, and responsibility in computer ethics. *ACM SIGCAS Computers and Society*, 28(1):14–21.
- MOSCO, V. (2005). *The Digital Sublime : Myth, Power, and Cyberspace*. MIT Press, Cambridge, Massachusetts.
- MUMFORD, L. (1965). Utopia, the city and the machine. *Daedalus*, 103:271–292.
- MUNIER, B. (2014). Homme et machine : la phase du miroir. *Hermès, La Revue*, 1:143–148.
- MUÑOZ-TORRES, J. R. (2012). Truth and objectivity in journalism : Anatomy of an endless misunderstanding. *Journalism Studies*, 13(4):566–582.
- MUSSO, P. (2008). La "révolution numérique" : techniques et mythologies. *La Pensée*, 355:103–120.
- MUSSO, P. (2009). Usages et imaginaires des TIC : la fiction des frictions. *In L'évolution des cultures numériques : de la mutation du lien social à l'organisation du travail*, pages 201–210. FYP Editions, Limoges.
- NAPOLI, P. M. (2014). Automated media : An institutional theory perspective on algorithmic media production and consumption. *Communication Theory*, 24(3):340–360.
- NARIN, B. (2018). Teaching high tech storytelling : Reorganizing journalism education for programmer journalists and data journalists. *Centro de Publicaciones Pontificia Universidad Católica del Ecuador*, pages 169–201.
- NEVEU, E. (1994). La société de communication et ses interprètes. *Réseaux*, 12(64):9–27.

- NEVEU, E. (2010a). News without journalists. *Brazilian Journalism Research, Brasília*, 6(1):29–54.
- NEVEU, E. (2010b). *Sociologie du journalisme*. Repères. La Découverte, Paris.
- NEVEU, E. (2015). *Sociologie politique des problèmes publics*. Sociologie. Armand Colin, Malakoff.
- NIELSEN, J. (1994). *Usability engineering*. Interactive Technologies. Academic Press, San Diego.
- NIELSEN, J. (1999). *Designing web usability: The practice of simplicity*. New Riders Publishing, Indianapolis.
- NYGREN, G. (2014). Multiskilling in the newsroom-de-skilling or re-skilling of journalistic work? *The journal of media innovations*, 1(2):75–96.
- O'REILLY, T. (2013). Open data and algorithmic regulation. In for AMERICA PRESS, C., éditeur : *Beyond transparency: Open data and the future of civic innovation*, pages 289–300. Goldstein, Brett and Dyson, Lauren and Nemani, Abhi, San Francisco.
- ORLIKOWSKI, W. J. (1992). The duality of technology : Rethinking the concept of technology in organizations. *Organization Science*, 3(3):398–427.
- ORNEBRING, H. (2000). Technology and journalism as-labour : Historical perspectives. *Journalism*, 11(1):57–74.
- PALM, E. et HANSSON, S. O. (2006). The case for ethical technology assessment (eTA). *Technological forecasting and social change*, 73(5):543–558.
- PAPINOT, C. (2014). *La relation d'enquête comme relation sociale : épistémologie de la démarche de recherche ethnographique*. Presses de l'Université Laval, Laval.
- PARASIE, S. (2011). Le journalisme hacker : une nouvelle utopie pour la presse? *La vie des idées*, 21 | <https://lavedesidees.fr/Le-journalisme-hacker.html>.
- PARASIE, S. (2013). Des machines à scandale. *Réseaux*, 2:127–161.
- PARISER, E. (2011). *The filter bubble : What the Internet is hiding from you*. Penguin UK, London.
- PASQUALE, F. (2015). *The black box society : The secret algorithms that control money and information*. Harvard University Press, Cambridge, Massachusetts.
- PATERSON, C. et ZOELLNER, A. (2010). The efficacy of professional experience in the ethnographic investigation of production. *Journal of Media Practice*, 11(2):97–109.
- PÉLISSIER, N. et ROMAIN, N. (1998). Journalisme de presse écrite et nouveaux réseaux d'information. *Les Cahiers du Journalisme*, 5:54–71.

- PEREIRA, F. H., TRÉDAN, O. et LANGONNÉ, J. (2018). Penser les mondes du journalisme. *Hermes, La Revue*, 3:99–106.
- PÉRINE, B., DAMHUIS, L., LAURENT, V., VALENDUC, G. et VENDRAMIN, P. (2010). *Diversité et vulnérabilité dans les usages des TIC. La fracture numérique au second degré*. Société et Avenir. Academia Press, Gent.
- PETRIE, H. et BEVAN, N. (2009). The evaluation of accessibility, usability and user experience. In STEPHANIDIS, C., éditeur : *The universal access handbook*, pages 10–20. CRC Press, Boca Raton, Florida.
- PIERRET, J. (2004). Place et usage de l'entretien en profondeur en sociologie. In KAMINSKI, D. et KOKOREFF, M., éditeurs : *Sociologie pénale : système et expérience*, pages 199–213. Erès, Toulouse.
- PINCH, T. P. et BIJKER, W. E. (2000). The social construction of facts and artifacts : Or how the sociology of science and the sociology of technology might benefit each other. *Technology, Organizations and Innovation. Critical Perspectives on Business and Management*, 2(1):698–717.
- PLAISANCE, P. L. (2000). The concept of media accountability reconsidered. *Journal of Mass Media Ethics*, 15(4):257–268.
- PLATTNER, T. et OREL, D. (2019). Addressing micro-audiences at scale. In *Proceedings of the Computation+Journalism Conference, 1-2 February 2019*, Miami. University of Miami.
- PLESNER, U. (2009). An actor-network perspective on changing work practices communication technologies as actants in newswork. *Journalism*, 10(5):604–626.
- PONTON, C. (1997). Génération automatique de textes : 30 ans de réalisations.
- POWERS, M. (2012). In forms that are familiar and yet-to-be invented. American journalism and the discourse of technologically specific work. *Journal of Communication Inquiry*, 36(1): 24–43.
- PRIMO, A. et ZAGO, G. (2015). Who and what do journalism? An actor-network perspective. *Digital Journalism*, 3(1):38–52.
- PROULX, S. (1994). Une lecture de l'oeuvre de Michel de Certeau : l'invention du quotidien, paradigme de l'activité des usagers. *Communication*, 15(2):171–197.
- PROULX, S. (2005). Penser les usages des technologies de l'information et de la communication aujourd'hui : enjeux-modèles-tendances. *Enjeux et usages des TIC : aspects sociaux et culturels*, 1:7–20.
- PROULX, S. (2006). Pour comprendre l'usage des objets communicationnels, (re)penser le constructivisme. *Degrés, Signe, culture et lien social à l'ère des réseaux*, 126-127:B1–B18.
- PROULX, S. (2015). La sociologie des usages, et après? *Revue française des sciences de l'information et de la communication*, 6 | DOI : 10.1109/ICE.2017.8279939.

- RAHMAN, Z. et WEHRMEYER, S. (2018). Searchable databases as a journalistic product. In BOUNEGRU, L. et STRAY, J., éditeurs : *The Data Journalism Handbook 2. Towards a Critical Data Practice*, volume <https://datajournalism.com/read/handbook/two>. European Journalism Centre.
- RAM, S. (1987). A model of innovation resistance. *ACR North American Advances*, 14:208–212.
- RAYMOND, E. (1999). The cathedral and the bazaar. *Knowledge, Technology & Policy*, 12(3):23–49.
- REBEYROLLE, J., JACQUES, M.-P. et PÉRY-WOODLEY, M.-P. (2009). Titres et intertitres dans l'organisation du discours. *Journal of French Language Studies*, 19(02):269–290.
- REDMAN, T. C. (1996). *Data quality for the information age*. Artech House Telecommunications Library. Artech House, Norwood, Massachusetts.
- REID, R., FRASER-KING, G. et SCHWADERER, W. (2007). *Data lifecycles : managing data for strategic advantage*. Wiley, Hoboken, New Jersey.
- REITER, E. et DALE, R. (1997). Building applied natural language generation systems. *Natural Language Engineering*, 3(1):57–87.
- REITER, E. et DALE, R. (2000). *Building Natural Language Generation Systems*. Cambridge University Press, Cambridge.
- RHÉAUME, J. (1982). La recherche-action : un nouveau mode de savoir? *Sociologie et sociétés*, 14(1):43–51.
- RICOEUR, P. (1991). Ethique et morale. In *Autour du politique*, pages 256–269. Le Seuil, Paris.
- RODGERS, S. (2015). Foreign objects? Web content management systems, journalistic cultures and the ontology of software. *Journalism*, 16(1):10–26.
- ROGERS, E. (2003). *Diffusion of Innovations, 5th Edition*. Free Press, New York.
- ROGERS, S. (2013). *Facts are sacred*. Faber & Faber, Londres.
- ROGERS, S., SCHWABISH, J. et BOWERS, D. (2017). Data journalism in 2017 : The current state and challenges facing the field today. *Google News Lab*.
- ROUX, A. (2007). De l'usage à la pratique : les processus d'appropriation. Emprunts à la théorie de la structuration et empreinte du chercheur. *Communication et Organisation*, 31:124–139.
- RUELLAN, D. (1992). Le professionnalisme du flou. *Réseaux*, 10(51):25–37.
- RUELLAN, D. (2005). Expansion ou dilution du journalisme? *Les Enjeux de l'Information et de la Communication*, 2005(1):77–86.
- RYFE, D. M. (2009). Broader and deeper : A study of newsroom culture in a time of change. *Journalism*, 10(2):197–216.

- SABANOVIC, S. (2010). Robots in society, society in robots mutual shaping of society and technology as a framework for social robot design. *International Journal of Social Robotics*, 2(4):439–450.
- SALTZIS, K. et DICKINSON, R. (2008). Inside the changing newsroom : Journalists' responses to media convergence. *Aslib Proceedings*, 60(3):216–228.
- SANDOVAL-MARTÍN, M. T. et LA-ROSA, L. (2018). Big data as a differentiating sociocultural element of data journalism : the perception of data journalists and experts. *Communication & Society*, 31(4):193–209.
- SARTRE, J.-P. (1986). *L'imaginaire*. Gallimard, Paris.
- SCHMITZ-WEISS, A. et DOMINGO, D. (2010). Innovation processes in online newsrooms as actor-networks and communities of practice. *New Media & Society*, 12(7):1156–1171.
- SCHMITZ-WEISS, A. et RETIS-RIVAS, J. (2018). "I don't like maths, that's why i am in journalism". Journalism student perceptions and myths about data journalism. *Asia Pacific Media Educator*, 28(1):1–15.
- SCHUDSON, M. (2001). The objectivity norm in american journalism. *Journalism*, 2(2):149–170.
- SCHULZE, K. et KRÖMKER, H. (2010). A framework to measure user experience of interactive online products. *In Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research, August 2010, Eindhoven*, pages 1–5. ACM.
- SCRANTON, P. (1995). Determinism and indeterminacy in the history of technology. *Technology and Culture*, 36(2):S31–S53.
- SEAVER, N. (2017). Algorithms as culture : Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4 (2) | DOI : 10.1177/2053951717738104.
- SELWYN, N. (2003). Apart from technology : understanding people's non-use of information and communication technologies in everyday life. *Technology in society*, 25(1):99–116.
- SENER, R. et SMITH, E. A. (1967). Automated readability index. Rapport technique, Cincinnati University.
- SERRANO, Y. (2007). L'objectivité journalistique : droit des citoyens, devoir des journalistes? *Cahiers de Psychologie Politique*, 10 | <http://lo-del.irevues.inist.fr/cahierspsychologiepolitique/index.php?id=914>.
- SHANKS, G. (1999). Semiotic approach to understanding representation in information systems. *In DAMPNEY, C. N. G., éditeur : Proceedings of the Information Systems Foundations Workshop, Ontology, Semiotics and Practice*, Sydney.
- SHANKS, G. et CORBITT, B. (1999). Understanding data quality : Social and cultural aspects. *In Proceedings of the 10th Australasian Conference on Information Systems*, volume 785. Victoria University of Wellington, New Zealand.

- SHOEMAKER, P. et REESE, S. (1996). *Mediating the Message : Theories of Influences on Mass Media Content*. Longman, Londres.
- SINGER, J. B. (2003). Who are these guys? The online challenge to the notion of journalistic professionalism. *Journalism*, 4(2):139–163.
- SINGER, J. B. (2004). Strange bedfellows? The diffusion of convergence in four news organizations. *Journalism studies*, 5(1):3–18.
- SINGER, J. B. (2019). Habits of practice, habits of thought. *Journalism*, 20(1):135–139.
- SINGH, P. J., POWER, D. et CHUONG, S. C. (2011). A resource dependence theory perspective of ISO 9000 in managing organizational environment. *Journal of Operations Management*, 29(1-2):49–64.
- SIRÉN-HEIKEL, S., LEPPÄNEN, L., LINDÉN, C.-G. et BÄCK, A. (2019). Unboxing news automation. *Nordic journal of media studies*, 1(1):47–66.
- SOLER, H. H. (2008). A metaphor corpus in business press headlines. *Ibérica : Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)*, 15:51–70.
- SOLLIE, P. (2007). Ethics, technology development and uncertainty : an outline for any future ethics of technology. *Journal of Information, Communication and Ethics in Society*, 5(4):293–306.
- SOULÉ, B. (2007). Observation participante ou participation observante? Usages et justifications de la notion de participation observante en sciences sociales. *Recherches qualitatives*, 27(1):127–140.
- SPYRIDOU, L.-P., MATSIOLA, M., VEGLIS, A., KALLIRIS, G. et DIMOULAS, C. (2013). Journalism in a state of flux : Journalists as agents of technology innovation and emerging news practices. *International Communication Gazette*, 75(1):76–98.
- STEENSEN, S. (2011). Online journalism and the promises of new technology : A critical review and look ahead. *Journalism Studies*, 12(3):311–327.
- STEINER, C. (2012). *Automate this : How algorithms took over our markets, our jobs, and the world*. Penguin Group, East Rutherford, New Jersey.
- STEINER, T., van HOOLAND, S. et SUMMERS, E. (2013). MJ no more : Using concurrent Wikipedia edit spikes with social network plausibility checks for breaking news detection. In *Proceedings of the 22nd international conference on World Wide Web companion, May 2013, Rio de Janeiro*, pages 791–794. International World Wide Web Conferences Steering Committee.
- STERNE, J. (2014). "What do we want?" "Materiality!" "When do we want it?" "Now!". In TARLETON (IN), G., P.J., B. et K.A., E, éditeurs : *Media Technologies : Essays on Communication, Materiality and Society*. MIT Press, Cambridge, Massachusetts.
- STEVENSON, T. H. et BARNES, F. C. (2001). Fourteen years of ISO 9000 : Impact, criticisms, costs, and benefits. *Business Horizons*, 44(3):45–45.

- STONEMAN, J. (2015). Does open data need journalism? *Reuters Institute for the Study of Journalism, Oxford University*, <https://reutersinstitute.politics.ox.ac.uk/our-research/does-open-data-need-journalism>.
- STRAY, J. (2016). The curious journalist's guide to data. *Tow Center for Digital Journalism*, <https://www.gitbook.com/book/towcenter/curious-journalist-s-guide-to-data/details>.
- STRAY, J., BOUNEGRU, L., CHAMBERS, L. et KAYSER-BRIL, N. (2013). *Guide du datajournalisme : Collecter, analyser et visualiser les données*. Eyrolles, Paris.
- STRONG, D. M., LEE, Y. W. et WANG, R. Y. (1997). Data quality in context. *Communications of the ACM*, 40(5):103–110.
- SUNDAR, S. S. (1998). Effect of source attribution on perception of online news stories. *Journalism & Mass Communication Quarterly*, 75(1):55–68.
- TERZIOVSKI, M. et GUERRERO, J.-L. (2014). ISO 9000 quality system certification and its impact on product and process innovation performance. *International Journal of Production Economics*, 158:197–207.
- TERZIOVSKI, M., POWER, D. et SOHAL, A. S. (2003). The longitudinal effects of the ISO 9000 certification process on business performance. *European Journal of operational research*, 146(3):580–595.
- THOMAS, S. L., NAFUS, D. et SHERMAN, J. (2018). Algorithms as fetish : Faith and possibility in algorithmic work. *Big Data & Society*, 5(1):1–11.
- THURMAN, N., DÖRR, K. et KUNERT, J. (2017). When reporters get hands-on with robo-writing : Professionals consider automated journalism's capabilities and consequences. *Digital Journalism*, 5(10):1240–1259.
- THURMAN, N. et SCHIFFERES, S. (2012). The future of personalization at news websites : Lessons from a longitudinal study. *Journalism Studies*, 13(5-6):775–790.
- TRAN, G. B. et HERDER, E. (2015). Detecting filter bubbles in ongoing news stories. *In UMAP Extended Proceedings*.
- TRÉDAN, O. et DIERICKX, L. (2017). L'open data à rennes ou en quoi l'open data ne change pas le gouvernement des villes. *In Les données et le gouvernement de la ville, 17eme Congrès de l'Association Française de Science Politique*, 10-12 Juillet 2017, Montpellier.
- TRÉDAN, O. (2014). Quand le journalisme se saisit du web : l'exemple du datajournalism. *In LE CAM, F. et RUELLAN, D., éditeurs : Changements et permanences du journalisme*, pages 199–214. L'Harmattan, Paris.
- TUCHMAN, G. (1978). *Making News – A Study in the Construction of Reality*. Free Press, New York.
- TUFTE, E. (2001). *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut.

- TURILLI, M. (2007). Ethical protocols design. *Ethics and Information Technology*, 9(1):49–62.
- TURILLI, M. et FLORIDI, L. (2009). The ethics of information transparency. *Ethics and Information Technology*, 11(2):105–112.
- TURING, A. M. (1950). Computing machinery and intelligence. *Mind*, 49:433–460.
- UNGER, R. et CHANDLER, C. (2012). *A project guide to UX design : For user experience designers in the field or in the making*. Voices That Matter. Pearson Education, New York City, New York.
- URSELL, G. D. (2001). Dumbing down or shaping up? New technologies, new media, new journalism. *Journalism*, 2(2):175–196.
- VALENDUC, G. (2005). *La technologie, un jeu de société : au-delà du déterminisme technologique et du constructivisme social*. Collection "Sciences et Enjeux". Academia-Bruylant, Louvain-la-Neuve.
- VAN DALEN, A. (2013). The algorithms behind the headlines : How machine-written news redefines the core skills of human journalists. *Journalism Practice*, 6(5-6):648–658.
- VAN DEN HOVEN, J. et WECKERT, J. (2008). *Information technology and moral philosophy*. Cambridge University Press, Cambridge.
- VAN DER KAA, H. et KRAHMER, E. (2014). Journalist versus news consumer : The perceived credibility of machine written news. *In Proceedings of Computation+Journalism conference*, New York. Columbia University.
- VAN DER VORST, R. (1998). Engineering, ethics and professionalism. *European journal of engineering education*, 23(2):171–179.
- VAN DIJK, T. A. (1985). Structures of news in the press. *Discourse and communication : New approaches to the analysis of mass media discourse and communication*, 10:69–93.
- VAN LEUVEN, S., RAEYMAECKERS, K., LIBERT, M., LE CAM, F., STROOBANT, J., MALCORPS, S., JACQUET, A., D'HEER, J., HEINDERYCKX, E., VANHAELEWYN, B. *et al.* (2019). *Portrait des journalistes belges en 2018*. Academia Press, Gent.
- VAN TRIER, W. (1980). La recherche-action. *Déviance et société*, 4(2):179–193.
- VANDENDAELE, A. (2018). "Trust me, i'm a sub-editor" "production values" at work in newspaper sub-editing. *Journalism Practice*, 12(3):268–289.
- VANDERHAEGHE, B. (2011). Van directe perssteun naar steun via advertenties : België. *UGent (onuitgegeven masterproef)*.
- VERMEEREN, A. P., LAW, E. L.-C., ROTO, V., OBRIST, M., HOONHOUT, J. et VÄÄNÄNEN-VAINIO-MATTILA, K. (2010). User experience evaluation methods : Current state and development needs. *In Proceedings of the 6th Nordic Conference on Human-Computer Interaction : Extending Boundaries, 16-20 October, Reykjavik*, pages 521–530. ACM.

- VERRA, L., KAROUI, M. et DUDEZERT, A. (2012). Adoption symbolique d'un réseau social pour entreprise : le cas de Bouygues Construction. *In 17ème Conférence Internationale de l'Association Information et Management (AIM), 19 Septembre 2012, Lille.*
- VINCK, D. (1995). *Sociologie des sciences*. Série Sociologie. Armand Colin, Malakoff.
- VINCK, D. (2012). Manières de penser l'innovation. *In MIÈGE, B., éditeur : Les masques de la convergence : enquêtes sur sciences, industries et aménagements*, pages 125–148. Editions des Archives Contemporaines, Paris.
- VOLLE, M. (2006). *De l'informatique : savoir vivre avec l'automate*. Economica, Paris.
- von PAPE, T. et MARTIN, C. (2010). Non-usages du téléphone portable : au-delà d'une opposition binaire usagers/non-usagers. *Questions de communication*, 18:113.
- WADDELL, F. T. (2018). A robot wrote this? How perceived machine authorship affects news credibility. *Digital Journalism*, 6(2):236–255.
- WALDROP, M. M. (2016). The chips are down for Moore's law. *Nature News*, 530:144–147.
- WAND, Y. et WANG, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39(11):86–95.
- WANG, P. (2010). Chasing the hottest IT : Effects of information technology fashion on organizations. *MIS quarterly*, 34(1).
- WANG, R. Y., REDDY, M. P. et KON, H. B. (1995). Toward quality data : An attribute-based approach. *Decision Support Systems*, 13(3):349–372.
- WANG, R. Y. et STRONG, D. M. (1996). Beyond accuracy : What data quality means to data consumers. *Journal of management information systems*, 12(4):5–33.
- WANG, R. Y., ZIAD, M. et LEE, Y. W. (2006). *Data quality*, volume 23 de *Science & Business Media*. Springer, Cham.
- WATINE, T. (2004). Entre réalité, vérité et objectivité : la «perspective» journalistique. *Les Cahiers du Journalisme*, 13:24–38.
- WEBER, M. S. et KOSTERICH, A. (2018). Coding the news : The role of computer code in filtering and distributing news. *Digital Journalism*, 6(3):310–329.
- WHITE, A. (2008). *To tell you the truth. The ethical journalism initiative*. International Federation of Journalists, Bruxelles.
- WIEN, C. (2005). Defining objectivity within journalism. *Nordicom Review*, 26(2):3–15.
- WIENER, N. (1961). *Cybernetics or Control and Communication in the Animal and the Machine*, volume 25. MIT Press, Cambridge, Massachusetts.
- WILLIAMS, R. et EDGE, D. (1996). The social shaping of technology. *Research policy*, 25(6):865–899.

- WÖLKER, A. et POWELL, T. E. (2018). Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism*, DOI : 10.1177/1464884918757072.
- WRIGHT, S. et DOYLE, K. (2018). The evolution of data journalism : A case study of Australia. *Journalism Studies*, pages 1–17.
- YOUNG, M. L. et HERMIDA, A. (2015). From Mr. and Mrs. outlier to central tendencies : Computational journalism and crime reporting at the los angeles times. *Digital Journalism*, 3(3):381–397.
- YUJIAN, L. et BO, L. (2007). A normalized Levenshtein distance metric. *IEEE transactions on pattern analysis and machine intelligence*, 29(6):1091–1095.
- ZAMITH, R. (2019). Transparency, interactivity, diversity, and information provenance in everyday data journalism. *Digital Journalism*, 7(4):470–489.
- ZANDVOORT, H., VANDEPOEL, I. et BRUMSEN, M. (2000). Ethics in the engineering curricula : Topics, trends and challenges for the future. *European journal of engineering education*, 25(4): 291–302.
- ZION, L. et CRAIG, D. (2014). *Ethics for digital journalists : Emerging best practices*. Routledge, Londres.

Sources

Sources primaires

Les sources primaires intégrées dans cette thèse comprennent des entretiens en face à face, un entretien téléphonique, des communications reçues via courriel, des observations participantes et la participation à des conférences. Elles sont présentées ci-dessous par ordre alphabétique.

Nom	Prénom	Date	Lieu	Durée	Type	N° de chapitre
Allen	Robbie	26-01-16	En ligne	00:16:03	Webinaire	2
Alter Echos	Réunion de rédaction 1	30-05-17	Bruxelles	Non enregistré	Observation	3
Alter Echos	Réunion de rédaction 2	06-03-18	Bruxelles	01:09:25	Observation	3
Alter Echos	Réunion de rédaction 3	17-04-18	Bruxelles	57:40:00	Observation	3
AXSemantics	Computational Linguistics Unconference (CLUNC)	12-03-16	Stuttgart	Non pertinent	Conférence	2
Becquet	Nicolas	14-11-17	Bruxelles	00:42:32	Entretien	1, 4
Becquet	Nicolas	02-02-18	Bruxelles	00:42:06	Entretien	4
Becquet	Nicolas	21-03-19	Bruxelles	01:12:01	Entretien	4
Becquet	Nicolas	19-12-19	Bruxelles	00:49:12	Entretien	4
Blancafort	Helena	29-02-16	Paris	00:39:15	Entretien	2.4
Blancafort	Helena	24-07-18	Paris	01:06:58	Entretien	4
BRAL	(deux représentants)	27-03-17	Bruxelles	Non enregistré	Entretien	3
CELINE	(deux représentants)	28-05-17	Bruxelles	01:11:03	Entretien	3
Collet	Marc	26-05-19	En ligne	Non pertinent	Courriel	4
Collin	David	27-06-18	Bruxelles	Non enregistré (4 heures)	Observation	4
Collin	David	10-09-19	Bruxelles	00:28:41	Téléphone	4
Collomb	Aurélie	24-07-18	Paris	00:23:48	Entretien	4
Conférence nationale des métiers du journalisme	Les journalistes doivent-ils innover ?	24-01-19	Paris	Non pertinent	Conférence	4
De Camps	Philippe	25-04-16	En ligne	Non pertinent	Courriel	1
De Camps	Philippe	09-12-18	En ligne	Non pertinent	Courriel	1
de Loupy	Claude	29-02-16	Paris	00:39:15	Entretien	2.4
Dussart	Cédric	29-01-18	Namur	00:18:14	Entretien	1

Nom	Prénom	Date	Lieu	Durée	Type	N° de chapitre
Feulner	Frank	12-03-16	Stuttgart	00:14:46	Entretien	2
Garely	Aude	18-09-17	En ligne	Non pertinent	Courriel	3
Gautier	Céline	23-05-17	Bruxelles	00:18:19	Entretien	3
Gautier	Céline	09-08-18	Bruxelles	00:43:04	Entretien	3
GEN Summit	Your next recruit might be a robot	15-06-16	Vienne	Non enregistré	Conférence	2
Grégoire	Arnaud	11-09-17	Bruxelles	00:51:17	Entretien	1.3
Gutiérrez	Ricardo	11-01-16	Bruxelles	01:36:25	Entretien	2
Hamon	Benoît	24-07-18	Paris	00:36:15	Entretien	4
Koksal	Mehmet	06-02-18	Bruxelles	Non enregistré	Entretien	1
L'Echo	Réunion de travail 1	19-04-18	Bruxelles	00:57:54	Observation	4
L'Echo	Réunion de travail 2	26-04-18	Bruxelles	01:06:51	Observation	4
L'Echo	Réunion de travail 3	09-05-18	Bruxelles	01:01:02	Observation	4
L'Echo	Réunion de travail 4	16-05-18	Bruxelles	01:42:06	Observation	4
L'Echo	Réunion de travail 5	16-08-18	Bruxelles	01:57:49	Observation	4
L'Echo	Réunion de travail 6	26-02-19	Bruxelles	00:23:39	Observation	4
L'Echo/Syllabs	Call 1	02-05-18	En ligne	00:33:35	Appel	3
L'Echo/Syllabs	Call 2	04-05-18	En ligne	00:42:52	Appel	3
L'Echo/Syllabs	Call 3	17-06-18	En ligne	01:02:06	Appel	3
L'Echo/Syllabs	Call 4	21-09-18	En ligne	00:12:26	Appel	3
L'Echo/Syllabs	Call 5	26-11-18	En ligne	00:15:05	Appel	3
Lefèvre	Gabrielle	16-11-15	Bruxelles	00:31:27	Entretien	2
Legrand	Manon	26-09-17	Bruxelles	00:16:59	Entretien	3
Legrand	Manon	26-06-18	Bruxelles	00:13:51	Entretien	3
Lemaigre	Thomas	19-09-17	Bruxelles	00:45:38	Entretien	3
Matriche	Joël	mars-14	Liège	Non enregistré	Entretien	1
Matriche	Joël	08-09-17	En ligne	Non pertinent	Courriel	1
Mediafin	Kick off	31-01-18	Bruxelles	02:13:08	Observation	4
Mediafin	Kick off De Tijd	21-02-19	Bruxelles	00:48:51	Observation	4
Mormont	Marinette	26-09-17	Bruxelles	00:16:59	Entretien	3
Mormont	Marinette	26-06-18	Bruxelles	00:07:26	Entretien	3
Théâtre National	Les médias : renforts ou critiques des systèmes ?	28-09-17	Bruxelles	Non pertinent	Conférence	3

Nom	Prénom	Date	Lieu	Durée	Type	N° de chapitre
Vallet	Cédric	26-09-17	Bruxelles	00:10:50	Entretien	3
Vallet	Cédric	26-06-18	Bruxelles	00:22:48	Entretien	3
Verbeeren	Pierre	04-12-18	Bruxelles	00:31:13	Entretien	3
Warsztacki	Sandrine	08-06-17	Bruxelles	00:37:20	Entretien	3
Warsztacki	Sandrine	29-08-18	Bruxelles	00:09:13	Entretien	3
Wéry	Arnaud	10-11-17	Waremme	00:17:01	Entretien	1
Winkel	Julien	09-10-17	Bruxelles	00:12:24	Entretien	3
Winkel	Julien	26-06-18	Bruxelles	00:24:31	Entretien	3

Sources secondaires

L'ensemble sources secondaires (articles scientifiques et monographies) consultées dans le cadre de cette recherche ont été consignées dans une base de données, accessible en ligne dans une version interactive :

<https://www.ohmybox.info/these/dbthese/> (utilisateur + mot de passe : t2020ulb)

Celles-ci ne constituent pas l'ensemble de la bibliographie présentée dans cette thèse. Certaines sources y ont été privilégiées, tandis que d'autres ont permis d'apporter des éléments de contexte pour une compréhension générale du phénomène étudié.

Corpus pour l'analyse du discours médiatique des journalistes

Ce corpus des 300 articles est accessible en ligne, sous la forme d'une base de données interactive :

<https://www.ohmybox.info/these/discours/corpus.html> (utilisateur + mot de passe : t2020ulb)

Annexes

Annexe 1. Offre d'emploi pour un poste de datajournaliste au Guardian (2015)

<p>theguardian TheObserver</p> <p><i>Job Description</i></p> <p>Data Journalist</p>	
Department	Editorial
Reports to	Editor, Data Projects
Normal location	Kings Place, York Way, Kings Cross, London N1
Terms and conditions	12 month fixed term NUJ contract
Purpose of role	The Data Journalist will collaborate on data projects across the newsroom.
<p><u>Key responsibilities and accountabilities</u></p> <ul style="list-style-type: none"> • The Data Journalist will collaborate with reporters across the newsroom on a range of data-focused stories, from quick-hit pieces to long-term enterprise and investigations. • You must be able to do some writing, but the main focus is on enriching our journalism through data reporting and analysis. • You will be expected to generate story ideas and projects, but always with an eye toward collaboration across one of more desks. • You will work closely with investigations, the news desk, graphics, the interactive team, the Data Blog and data specialists elsewhere in the newsroom in order to identify opportunities at the earliest possible stage. • Because the Data Projects team is meant to be a force multiplier in the newsroom, the successful candidate must be aggressively collaborative. The ultimate goal is not just individual projects, but making stories that rely on (or should rely on) data better and data journalism itself less the purview of a handful of specialists. • Success means better, deeper, richer stories, and also a newsroom more willing to embrace the idea of data journalism generally and journalists who have the tools and training to work with data directly. 	

PERSON SPECIFICATION

	Essential	Desirable
<p>Knowledge & experience</p> <ul style="list-style-type: none"> • Track record of producing and editing significant, impactful pieces of data journalism. • Creative and forward-looking, always in search of new ways to use data in the reporting process. • Experience acquiring, cleaning and analysing datasets and a demonstrated ability to go from analysis to story, often under tight deadline. • Outstanding news judgement. Fastidious defender of accuracy and mindful of the ways data can be abused or misleading. • Knowledge of relevant FOI laws and how they apply to data. • Ability to collaborate with a range of newsroom data users, including reporters, editors, graphic artists, developers, etc. • Experience as a trainer, ideally in data journalism. • Excellent written and verbal communications skills. • The ability to juggle multiple competing priorities. 	ALL	
<p>Skills & behaviours</p> <ul style="list-style-type: none"> • Skilled with desktop spreadsheet and database software, such as MS Excel and MS Access. • Some experience with server-based database platforms like SQL Server, MySQL, Oracle or Postgres. • Knowledge of basic statistics and statistical software packages (SPSS, R, Stata, etc.) is a big plus. • Basic familiarity with web technologies (HTML, CSS, Javascript, Python, Ruby, etc.) and desire to learn code. • Experience with at least one programming language is a plus. • Flexible • Self Starter • Initiative • Calm under pressure 	ALL	

This job description is a guide to the work you will initially be required to undertake. It summarises the main aspects of the job but does not cover all the duties that the job holder may have to perform. It may be changed from time to time to meet changing circumstances. It does not form part of your contract of employment and as your experience grows, you will be expected to broaden your tasks, suggest improvements, solve problems and enhance the effectiveness of the role.

Annexe 2. Code pour l'analyse de sentiment du discours médiatique des journalistes

L'analyse de sentiment automatique a été réalisée avec le langage "R". Le code source a été adapté de "Text mining in R with Tyditext" (NICAR), et les ressources pour cette analyse proviennent de l'Université de l'Illinois à Chicago, "Opinion Mining, Sentiment Analysis, and Opinion Spam Detection. Feature-Based Opinion Mining and Summarization (or Aspect-Based Sentiment Analysis and Summarization)".

Sources : <https://paldhous.github.io/NICAR/2019/r-text-analysis.html> et <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>, consultés le 20/11/2018.

```
library("dplyr")
library("tidyverse")
library("tidytext")
library("ggplot2")

words <- titrefrd %>% unnest_tokens('word', 'Titre')
stop <- data.frame(word = stop_words)

def_words <- words %>%
anti_join (stop, by = "word")

def_words %>%
count(word, sort = TRUE)

def_words %>%
count(word, sort = TRUE) %>%
filter(n > 6) %>%
mutate(word = reorder(word, n)) %>%
ggplot(aes(word, n)) +
geom_col() +
xlab(NULL) +
coord_flip()

#bigrammes

bigrams <- def_words %>%
unnest_tokens(bigram, word, token = "ngrams", n = 2)

big <- data.frame(word = bigrams)
```

```

big %>%
count(bigram, sort = TRUE)

big %>%
count(bigram, sort = TRUE) %>%
filter(n > 5) %>%
mutate(word = reorder(bigram, n)) %>%
ggplot(aes(word, n)) +
geom_col(fill="dodgerblue4") +
xlab(NULL) +
coord_flip() +
labs(y = "Frequence") +
theme_minimal()

#sentiment

sentimentsf <- def_words %>%
inner_join(senti, by = "word")

sentimentsf_counts <- sentimentsf %>%
count(sentiment) %>%
arrange(-n)

negativf_freqs <- sentimentsf_counts %>%
filter(sentiment == "negatif")

sentimentsf %>%
count(sentiment, sort = TRUE) %>%
mutate(sentiment = reorder(sentiment, n)) %>%
ggplot(aes(sentiment, n)) +
geom_col() +
xlab(NULL) +
geom_col(fill="dodgerblue4") +
coord_flip() +
theme_minimal()

```

Annexe 3. The Critical Engineering Manifesto

The Critical Engineering Working Group. Berlin, October 2011-2016. Julian Oliver, Gordan Savičić, Danja Vasiliev. Traduction française : Benedicte Jacobs.

Source : <https://criticalengineering.org/fr>, consulté le 11/10/2016.

0. L'ingénieur critique considère l'ingénierie comme le langage le plus percutant de notre époque modélisant nos manières de nous déplacer, communiquer et penser. Le rôle de l'ingénieur critique est d'étudier et d'exploiter ce langage – de montrer son influence.

1. L'ingénieur critique considère notre dépendance à la technologie comme un défi autant qu'une menace. Plus la dépendance est grande plus il importe d'en analyser le fonctionnement interne indépendamment des droits de propriétés et dispositions légales.

2. L'ingénieur est conscient que chaque avancée technologique met à mal notre héritage technopolitique.

3. L'ingénieur critique déconstruit et incite à la méfiance face aux expériences utilisateurs luxuriantes.

4. L'ingénieur critique ne se laisse pas impressionner par les contraintes de l'implémentation. Il en explore les méthodes d'influence et leurs effets spécifiques.

5. L'ingénieur critique reconnaît que chaque travail d'ingénierie formate l'utilisateur proportionnellement à la dépendance de ce dernier au travail d'ingénierie.

6. En décrivant les interrelations des dispositifs, corps, agents et forces des réseaux, l'ingénieur critique élargit le concept de 'machine'.

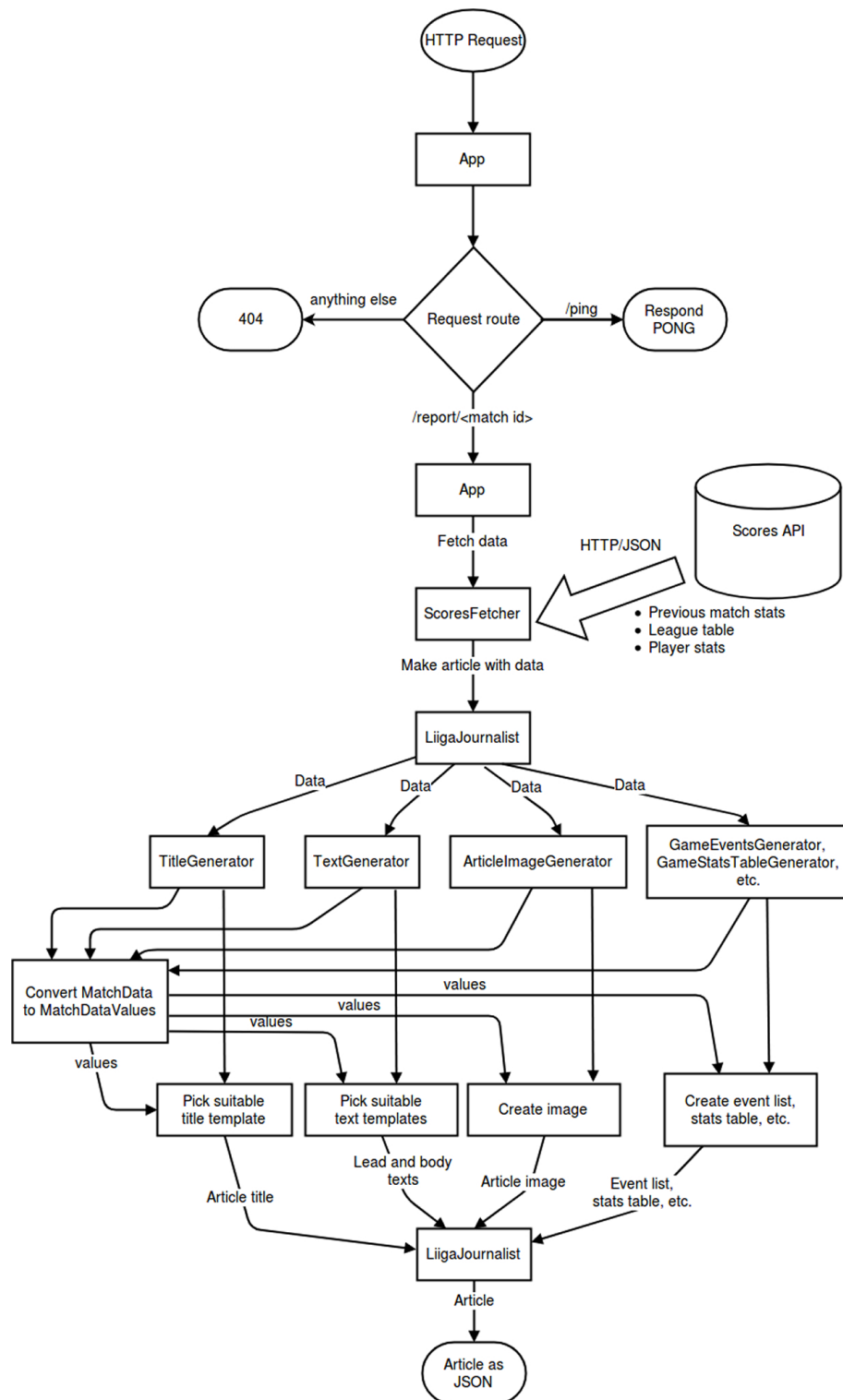
7. L'ingénieur critique observe l'espace entre la production et la consommation des technologies. En agissant rapidement, l'ingénieur critique peut déstabiliser cet espace provoquant des moments de déséquilibre et de déception.

8. L'ingénieur critique se tourne vers l'histoire de l'art, l'architecture, l'activisme, la philosophie et l'innovation pour y trouver des travaux exemplaires d'ingénierie critique. Les stratégies, idées, agendas, de ces disciplines seront adoptées, re-agencées et développées.

9. L'ingénieur critique constate que le code écrit s'immisce dans les secteurs du social et du psychologique régulant les comportements entre les personnes et les machines avec lesquelles elles interagissent. Prenant conscience de ce fait, l'ingénieur critique cherche à reconstruire les contraintes et actions sociales des utilisateurs en procédant à des fouilles numériques.

10. L'ingénieur critique considère l'utilisation de la vulnérabilité d'un système comme la forme la plus souhaitable de dénonciation.

Annexe 4. Modélisation du processus de Voitto, développé pour le radiodiffuseur Yle (Finlande)



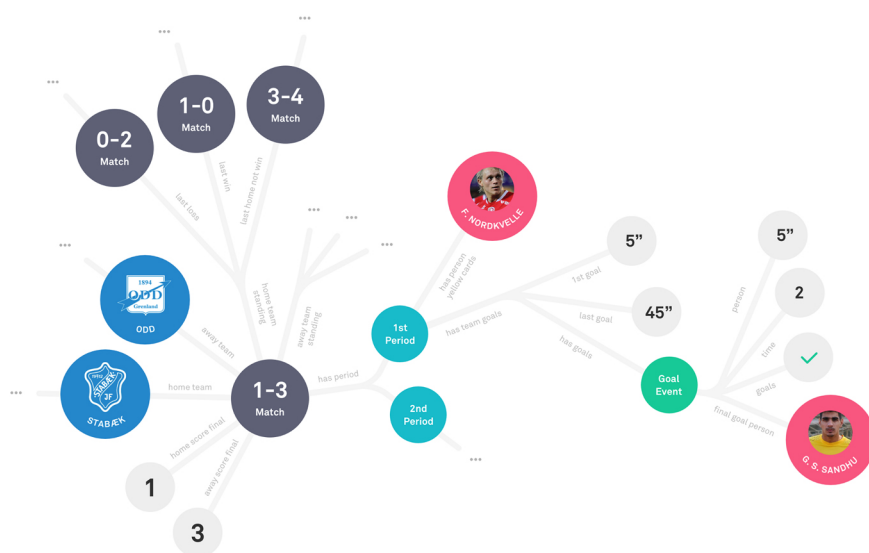
Annexe 5. Structure pour l'entraînement du "robot" de l'agence de presse norvégienne NTB

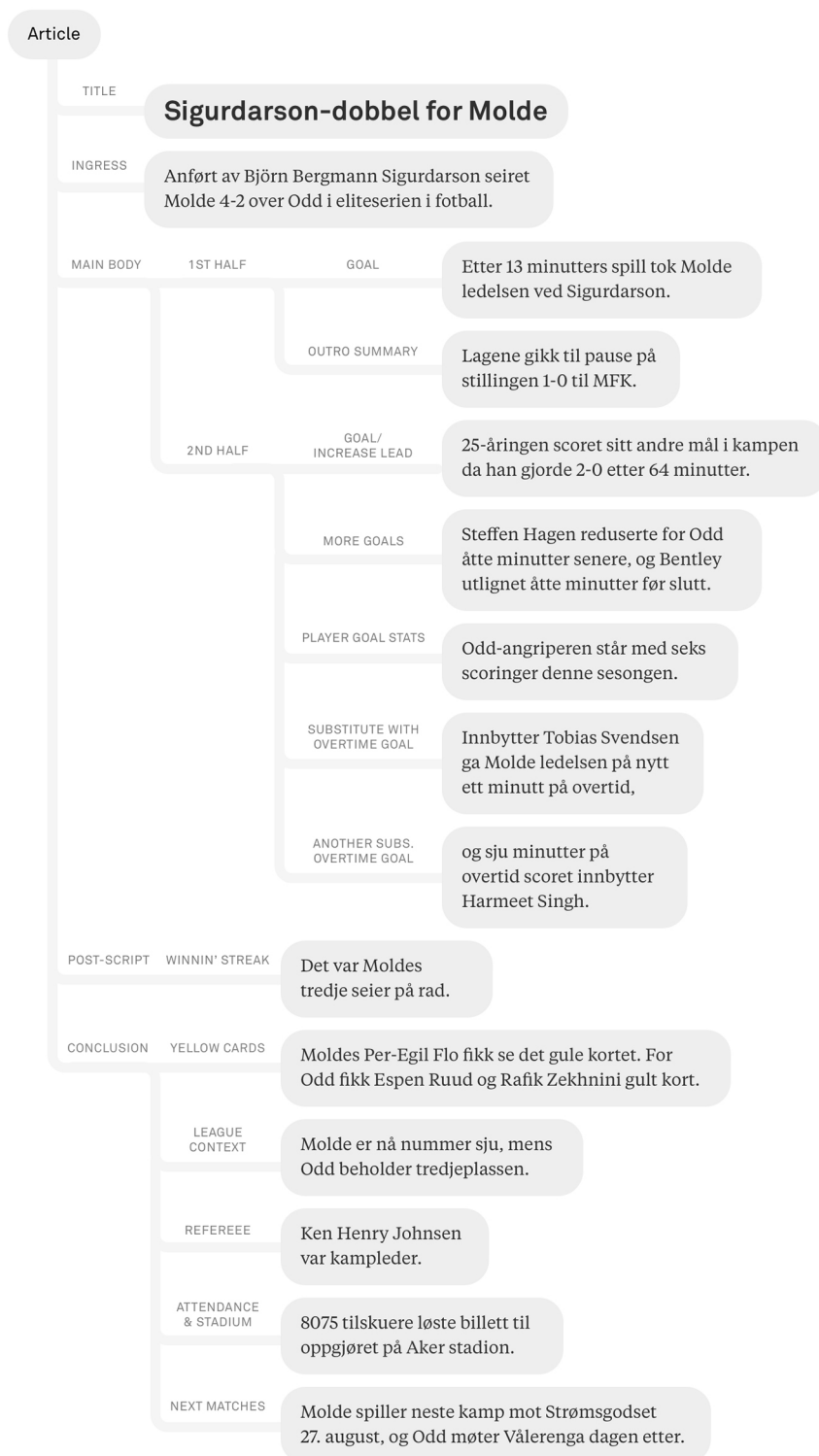
Sigurdarson-dobbel for Molde

Anført av Björn Bergmann Sigurdarson seiret Molde 4-2 over Odd i eliteserien i fotball.

Etter 13 minutters spill tok Molde ledelsen ved Sigurdarson. Lagene gikk til pause på stillingen 1-0 til MFK.

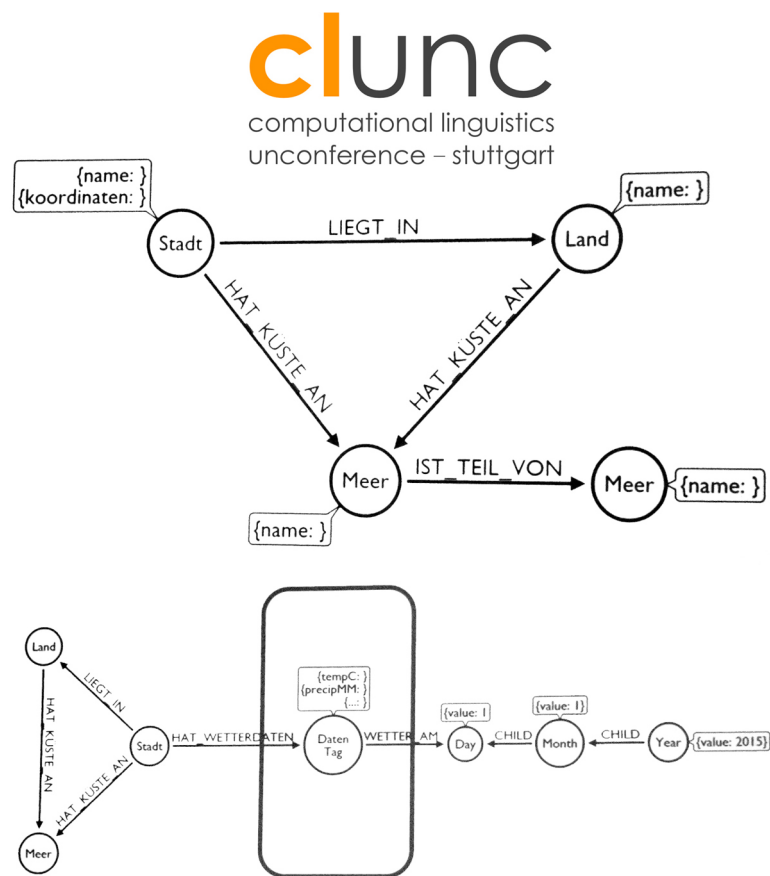
25-åringen scoret sitt andre mål i kampen da han gjorde 2-0 etter 64 minutter. Steffen Hagen reduserte for Odd åtte minutter senere, og Bentley utlignet åtte minutter før slutt.





Annexe 6. Présentation du processus de génération automatique de textes météorologiques relatifs à des destinations de vacances

Présentation du processus de génération automatique de textes météorologiques relatifs à des destinations de vacances par la société AEXEA/AX Semantics, les 11 et 12 mars 2016. Le projet présenté concerne la génération automatique de bulletins météorologiques, en langue allemande, sur des destinations de voyage. Cette présentation a mis en évidence le challenge lié à l'amélioration de la qualité des textes générés et de leur variabilité – une nécessité dès lors qu'un même jeu de données est susceptible de faire l'objet de plusieurs traitements.



DURCHSCHNITTliche WASSERTEMPERATUR IN ANTALYA IM FRÜHLING: 19°C

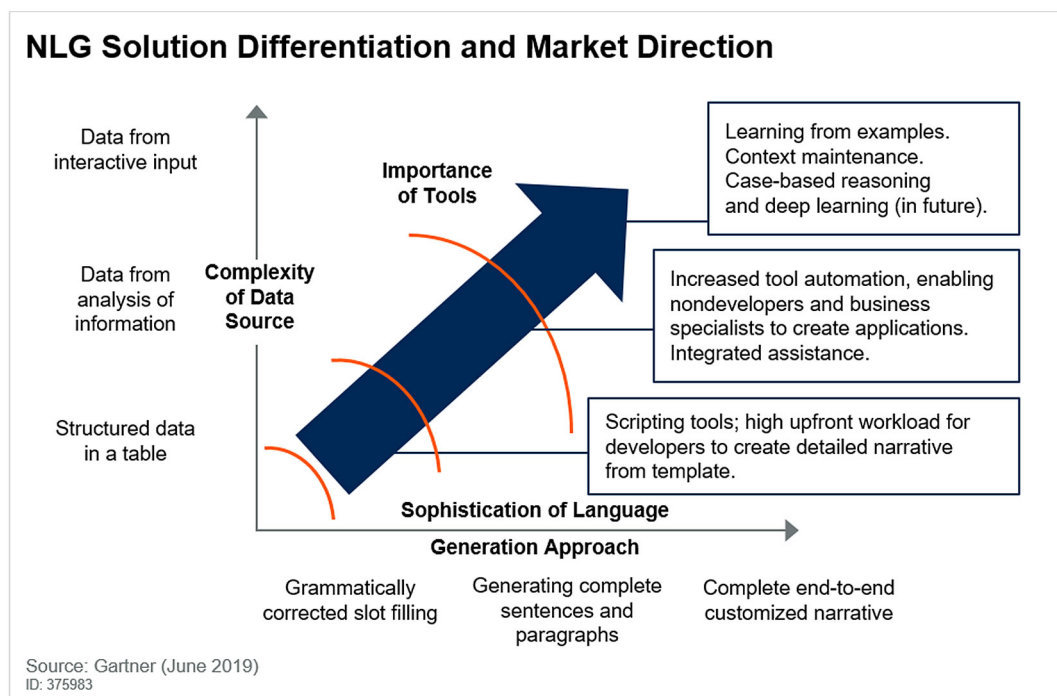
Antalya gehört damit, neben bspw. **Belek**, zu den Destinationen mit der höchsten Wassertemperatur am **Europäischen Mittelmeer**. Im Frühling schwanken die Temperaturen zwischen **16** und **20**°C. Bei diesen relativ kalten Temperaturen kommen vor allem die Abgehärteten unter uns auf ihre Kosten. Dabei sollten 10 - 20 Minuten ausreichen, um "erfrischt" aus dem Meer zu steigen. Für Tauchgänge empfehlen wir dir die Verwendung eines Halbtrockenanzugs mit 7mm Stärke.

DAS WETTER IM FRÜHLING

Ab dem Frühling ist das Klima in **Antalya** bereits für ausgedehnte Städtereisen geeignet. Im März bis Mai kannst du mit durchschnittlichen Lufttemperaturen von **21**°C rechnen. Manchmal klettert das Quecksilber sogar auf **27** Grad. Mit Regen musst du an durchschnittlich **sieben** Tagen im Monat rechnen. Diese Jahreszeit bietet dir damit bereits gute Voraussetzungen, um dir Sehenswürdigkeiten wie die **Altstadt**, das **Hadrianstor** oder die **Yivli-Minare-Moschee** anzusehen.

Annexe 7. Différenciation des solutions de génération automatique en langage naturel et directions du marché

En juin 2019, la société de consultance Gartner a publié son premier rapport sur le marché des logiciels spécialisés en génération automatique de textes en langage naturel. Le graphique ci-dessous montre qu'une solution logicielle devient importante plus le langage est sophistiqué, plus la source des données est complexe et lorsque des non-développeurs peuvent l'utiliser pour créer des applications spécifiques¹³¹.



¹³¹ Source; Gartner, via LinkedIn, consulté le 15/07/2019, URL : <https://www.linkedin.com/feed/update/urn:li:activity:6550675931621576705>

Annexe 8. Formulaire d'évaluation diffusés via Google Forms dans le cadre de l'expérience "Bxl'airbot"

1. Faire connaissance (juin 2017)

- Q1. Vos noms et prénom (réponse courte)
- Q2. Depuis combien de temps travaillez-vous pour *Alter Échos*? (réponse courte)
- Q3. Quel est votre statut? (réponse courte)
- Q4. Quelle est votre formation? (réponse longue)
- Q5. Quel est votre parcours professionnel? (réponse longue)

2. Perception et usages du newsbot (anonyme, juin 2017)

Q1. Avant l'arrivée du dispositif au sein de la rédaction, aviez-vous déjà entendu parler de robot-journalisme? (Modes d'évaluation booléen, oui/non + questions ouvertes)

Q1.1. À. Si "OUI", quelle opinion aviez-vous à ce propos? Si "NON", passez à la question 1.2.

Q1.1.B. Cette opinion a-t-elle été modifiée depuis le début de l'expérience?

Q1.1.C. Commentez 1.1.B.

Q1.2. Si la réponse 1.1 est "NON", quelle est votre opinion à ce propos?

Q2. Vous êtes-vous déjà rendu sur la plateforme? (Modes d'évaluation booléen + échelle de Likert de 1 à 5 – de "Pas d'accord" à "Tout à fait d'accord")

Q2.1.À. Si la réponse à la question précédente est "OUI", à quelle fréquence? Si la réponse est "NON", passez à la question 3.

Q2.1.B. Si plus d'une fois par semaine, précisez

Q2.1.C. Si moins d'une fois par mois, précisez

Q2.2. Avez-vous consulté la plateforme en vue de recueillir une information en particulier? Commentez.

Q3. Etes-vous d'accord avec les affirmations suivantes? (Mode d'évaluation : échelle de Likert de 1 à 5 – de "Pas d'accord" à "Tout à fait d'accord" – + question ouverte)

Q3.1. L'interface est claire

Q3.2. L'interface est lisible

Q3.3. Je trouve facilement ce que je cherche

Q3.4. Les données sont bien expliquées

Q3.5. Le traitement proposé des données est pertinent

Q3.6. La navigation est sans équivoques

Q3.7. Les infographies et tableaux sont compréhensibles

Q3.8. Le bulletin du jour est explicite et intelligible

Q3.9. Le bulletin du jour est plaisant à lire

Q3.10. Le bulletin du jour est précis

Q3.11. Le bulletin du jour est bien écrit

Q3.12. Le bulletin du jour est objectif

Q3.13. Avez-vous rencontré des difficultés en matière de navigation?

Si oui, précisez

Q4. À ce stade de l'expérience, quelles améliorations souhaiteriez-vous voir apportées à "Bxl'air bot"?

Q4.1. En quoi pensez-vous que ces améliorations vont aider/soutenir votre travail journalistique?

Q4.2. Commentaire libre sur le projet

Évaluation de la formation data (anonyme, juillet 2017)

Q1. Quelle a été votre motivation pour participer à cette formation? (réponse longue)

Q2. Aviez-vous des attentes particulières? (oui / non)

Q2.1. Si oui, lesquelles? Si non, passez à la question 3 (réponse longue)

Q2.2. Ont-elles été rencontrées? (oui / non)

Q3. Les explications étaient-elles suffisamment claires? (oui / non)

Q4. Le support écrit était-il suffisant? (oui / non)

Q5. Quels points travaillés dans cette formation allez-vous pouvoir appliquer dans votre activité professionnelle? (choix multiples)

- Requêtes dans une base de données
- Analyse de données dans un tableur
- Visualisation de données
- Je ne vais appliquer aucun de ces points

Q6. Ces points sont-ils en relation directe avec le projet Bxl'air bot? (oui / non)

Q7. Quel aspect spécifique souhaiteriez-vous approfondir par la suite? (choix multiples)

- Requêtes dans une base de données
- Analyse de données dans un tableur
- Visualisation de données
- Aucun de ces aspects

Q8. Le niveau de la formation vous a-t-il semblé...

- Faible
- Correct
- Trop élevé

Q9. Quelle est votre degré de satisfaction global?

- Pas du tout satisfait
- Moyennement satisfait
- Pas du tout satisfait

Q10. Etes-vous journaliste? (oui / non)

Q11. Souhaitez-vous ajouter un commentaire? (réponse longue)

Annexe 9. Guide d'entretien pour l'évaluation finale de l'expérience "Bxl'air bot"

Place du système d'information dans la rédaction

Q1. "Un an avec un robot", c'était le titre de l'article de Céline Gautier. Quelle place ce projet a-t-il occupé au sein d'*Alter Échos*?

Q2. Considérez-vous que cette application constituait une forme de journalisme?

Usages du dispositif

Q3. Vous êtes-vous sur le site du bot? Si oui, à quelle fréquence et pourquoi?

Q4. En quoi considérez-vous que cette application était utile?

Q5. En quoi considérez-vous que cette application était inutile?

Impact du dispositif et pratiques professionnelles

Q6. Pensez-vous que cette expérience a eu un impact plutôt positif, plutôt négatif?

Q7. En avez-vous retiré des bénéfices dans le cadre de ta pratique professionnelle?

Q8. Cela a-t-il modifié votre manière d'appréhender votre métier? Si oui, en quoi? Si non, pourquoi?

Q9. Cela a-t-il entraîné une réflexion particulière relative à l'évolution du métier?

Évaluation générale

Q10. Si c'était à refaire, seriez-vous partant? Si oui, pourquoi et sur quel type de données? Si oui, cela serait-il à refaire de la même manière? Si non, pourquoi?

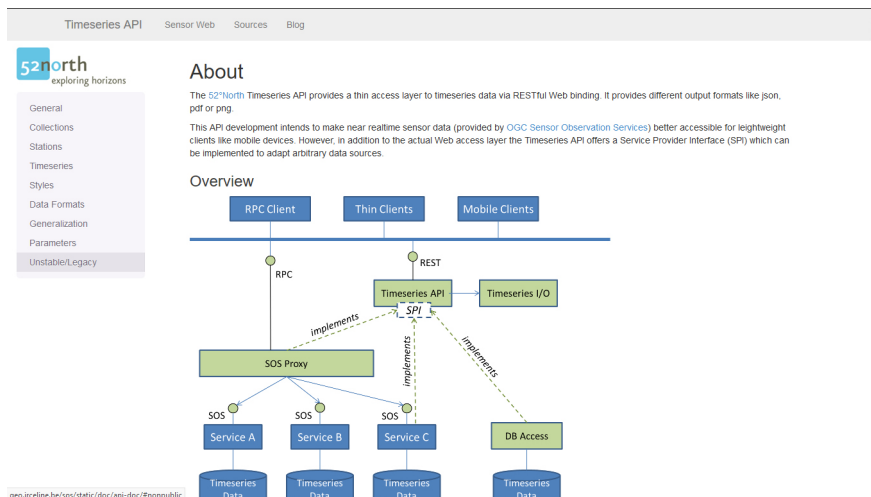
Annexe 10. Données en temps réel diffusées par CELINE (web service)

1. Données non validées en temps réel : <http://geo.ircline.be/sos/api/v1/>

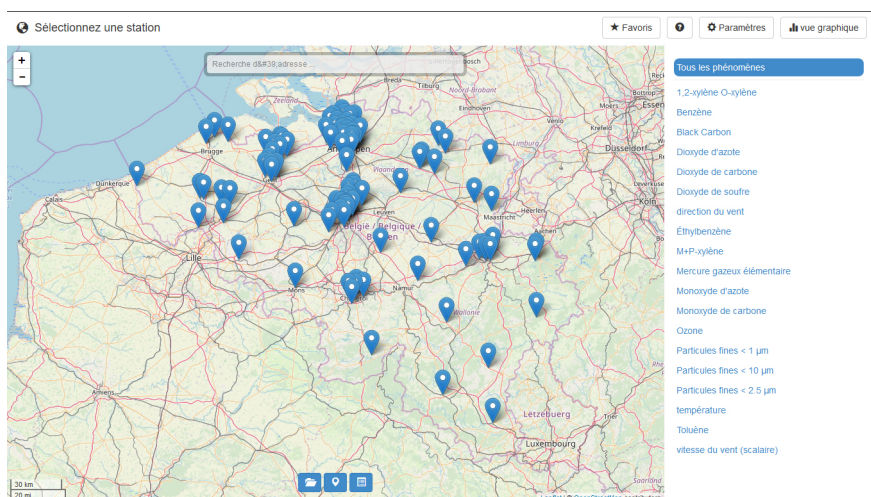
```
JSON Données brutes En-têtes
Enregistrer Copier Formater et indenter
{
  "id": "services",
  "label": "Service Provider",
  "description": "A service provider offers timeseries data."
},
{
  "id": "stations",
  "label": "Station",
  "description": "A station is the place where measurement takes place."
},
{
  "id": "timeseries",
  "label": "Timeseries",
  "description": "Represents a sequence of data values measured over time."
},
{
  "id": "categories",
  "label": "Category",
  "description": "A category group available timeseries."
},
{
  "id": "offerings",
  "label": "Offering",
  "description": "An organizing unit to filter resources."
},
{
  "id": "features",
  "label": "Feature",
  "description": "An organizing unit to filter resources."
},
{
  "id": "procedures",
  "label": "Procedure",
  "description": "An organizing unit to filter resources."
},
{
  "id": "phenomena",
  "label": "Phenomenon",
  "description": "An organizing unit to filter resources."
}
}

JSON Données brutes En-têtes
Enregistrer Copier Formater et indenter
[{"id":"services","label":"Service Provider","description":"A service provider offers timeseries data."}, {"id":"stations","label":"Station","description":"A station is the place where measurement takes place."}, {"id":"timeseries","label":"Timeseries","description":"Represents a sequence of data values measured over time."}, {"id":"categories","label":"Category","description":"A category group available timeseries."}, {"id":"offerings","label":"Offering","description":"An organizing unit to filter resources."}, {"id":"features","label":"Feature","description":"An organizing unit to filter resources."}, {"id":"procedures","label":"Procedure","description":"An organizing unit to filter resources."}, {"id":"phenomena","label":"Phenomenon","description":"An organizing unit to filter resources."}]
```

2. Documentation de l'API (Application Programming Interface) : <http://geo.ircline.be/sos/static/doc/api-doc/>



3. Données en ligne : <http://geo.ircline.be/sos/static/client/jsClient/#map>



4. Documentation sur les autres services disponibles : https://github.com/ircline/open_data

The screenshot shows the GitHub interface for the repository 'ircline/open_data'. At the top, there are navigation links for 'Why GitHub?', 'Business', 'Explore', 'Marketplace', and 'Pricing', along with a search bar and 'Sign in'/'Sign up' buttons. The repository name 'ircline / open_data' is displayed with 5 watches, 1 star, and 1 fork. Below this, there are tabs for 'Code', 'Issues', 'Pull requests', 'Projects', and 'Insights'. The current view is the 'Code' tab, showing the file structure 'open_data / datasets /'. A commit history table is visible, listing files like 'opeeters update datasets/rio.md', 'measurements.md', 'reporting.md', and 'rio.md' with their respective commit dates. Below the table, there are two sections of XML error messages. The first section shows an error related to a missing parameter value in an OWS request. The second section shows an error related to a NullPointerException in the OWSExceptionReport class.

File	Commit Message	Time
opeeters update datasets/rio.md		Latest commit dbed9bc on 28 Jun 2016
measurements.md	add initial version add link list of abbreviations	3 years ago
reporting.md	initial commit datasets/reporting.md	3 years ago
rio.md	update datasets/rio.md	3 years ago

```
-<ows:ExceptionReport version="2.0.0" xsi:schemaLocation="http://www.opengis.net/ows/1.1 http://geo.ircline.be/schemas/ows/1.1.0/owsAll.xsd">
  <ows:Exception exceptionCode="MissingParameterValue" locator="request">
    <ows:ExceptionText>
      Could not determine groserver request from http request org.groserver.platform.AdvancedDispatchFilter$AdvancedDispatchHttpRequest@1fbc6943
    </ows:ExceptionText>
  </ows:ExceptionText>
</ows:ExceptionReport>
```

```
-<ows:ExceptionReport version="1.0.0" xsi:schemaLocation="http://www.opengis.net/ows http://geo.ircline.be/schemas/ows/1.0.0/owsExceptionReport.xsd">
  <ows:Exception exceptionCode="NotApplicableCode">
    <ows:ExceptionText>java.lang.NullPointerException</ows:ExceptionText>
  </ows:ExceptionText>
</ows:ExceptionReport>
```

Annexe 11. Première version de l'interface de l'application "Bxl'air bot"

Bxl'air bot
i 🔔 🌐 📧

Données du jour
Statistiques
Rapports mensuels
Normes et dépassements


Bulletin du 13 mai 2017

[Twitter](#)
[Facebook](#)
[Partager 12](#)


Pas d'alerte pollution

11h04. L'indice de qualité de l'air est de 2 sur une échelle de 1 à 10, communique Bruxelles Environnement. Le taux moyen de particules fines de type PM10 est de 10.5 µg/m³. Ce taux est en-dessous de la recommandation de l'Organisation mondiale de la santé (OMS), ainsi que sous la norme européenne. Le taux moyen de particules fines de type PM2.5 est de 7.3 µg/m³. Ce taux se trouve sous la norme recommandée par l'OMS. Le taux moyen de particules fines de type black carbon (carbone suie) est de 1.4 µg/m³. Le ciel est très nuageux. La température observée est de 14,5 degrés.


Ce texte a été généré de manière automatique à partir de données extraites en temps réel.



Indice de la qualité de l'air
Source : Bruxelles Environnement (diffusion IRCEL-CELINE)



Bxl'air bot, le newsbot d'Alter Echos pour monitorer la qualité de l'air bruxellois



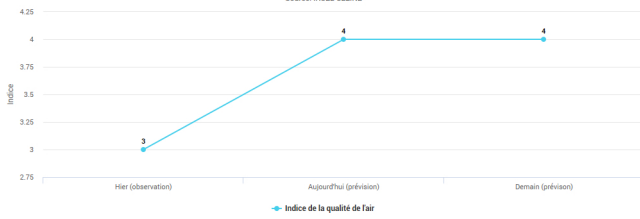
Taux de particules fines
Source : Cellule Interregionale d'Environnement (IRCEL-CELINE)

Indice de la qualité de l'air

L'indice de la qualité de l'air est calculé tout au long de la journée, à partir de données récoltées par la Cellule Interregionale d'environnement (CELINE). Cet indice est basé sur les taux de quatre polluants : ozone, dioxyde d'azote, dioxyde de soufre et particules fines. Il s'agit d'une appréciation qualitative (indice BeATMO), réalisée à partir de mesures techniques opérées en différents endroits tout au long de la journée. Sa valeur scientifique est très limitée, s'agissant avant tout d'un outil de communication.

Qualité de l'air : évolution des trois derniers jours

Source: IRCEL-CELINE



Day	Index
Hier (observation)	3
Aujourd'hui (prévision)	4
Demain (prévision)	4

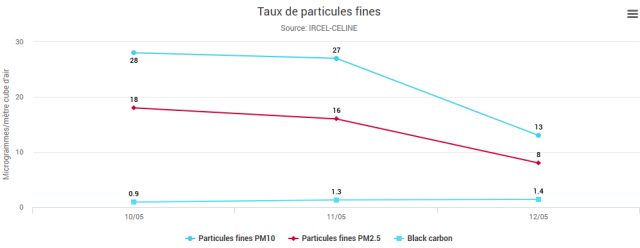
—•— Indice de la qualité de l'air

Taux moyen de particules fines

Trois types de particules fines font l'objet de relevés : les PM10, les PM2.5 et les Black carbon (carbone suie).
[i Voir les recommandations de l'OMS et les normes européennes](#)

Taux de particules fines



Source: IRCEL-CELINE



Date	PM10 (µg/m³)	PM2.5 (µg/m³)	Black carbon (µg/m³)
10/05	25	18	0.9
11/05	27	16	1.3
12/05	13	8	1.4

—•— Particules fines PM10 —•— Particules fines PM2.5 —•— Black carbon

Ce graphique a été généré à partir de données extraites et enregistrées quotidiennement de manière automatique.

Annexe 12. Bxl'air bot : génération automatique de textes en langue naturelle

À. Structure du résumé du texte en page d'accueil de l'application

Bulletin du [date] - [alerte] - [heure]. [indice][valeur] [Bruxelles], [source des données]. [PM 10] [valeur]. [comparaison] [valeur] [recommandation OMS]. [comparaison] [valeur] [norme européenne]. [comparaison] [valeur] [Belgique]. [PM 2.5] [valeur]. [comparaison] [recommandation OMS]. [comparaison] [Belgique]. [Black carbon] [valeur]. [comparaison] [Belgique]. [O3] [valeur]. [comparaison] [OMS]. [NO2] [valeur]. [comparaison] [Belgique]. [type de temps]. [température].

Exemple de textes générés le 15/07/2017.

Bulletin du 15 juillet 2017

Pas d'alerte pollution

13h50. L'indice de la qualité de l'air est très bon dans la capitale, indique Bruxelles Environnement. La teneur moyenne en particules fines de type PM10 est de 10,5 µg/m³. Elle se trouve sous la recommandation de l'Organisation mondiale de la santé (OMS), ainsi que sous la norme européenne. Le taux moyen de particules fines de type PM2.5 est de 8,3

µg/m³. Il est en-dessous de la norme recommandée par l'OMS, qui est de 25 µg/m³. Le taux moyen de black carbon (carbone suie) est de 1,1 µg/m³. La concentration d'ozone dans l'air est de 51 µg/m³. Le taux moyen de dioxyde d'azote est de 15 µg/m³. Le ciel est très nuageux, la température est de 19,3 degrés.

Ce texte a été généré de manière automatique à partir de données publiques extraites en temps réel.

Bulletin du 15 juillet 2017

Pas d'alerte pollution

14h30. L'indice de la qualité de l'air est très bon en région bruxelloise, indique Bruxelles Environnement. La teneur de l'air en particules fines de type PM10 est de 13 µg/m³. Elle est sous la norme européenne de 50 µg/m³, laquelle est identique à la norme recommandée par l'Organisation mondiale de la santé (OMS). Ce taux se trouve sous le taux moyen constaté à l'échelle nationale (15 µg/m³). Le taux moyen de particules fines de type PM2.5 est de 9 µg/m³.

Ce taux se trouve en-dessous de la norme recommandée par l'OMS. Il est plus élevé que le taux moyen observé à l'échelle nationale. Le taux moyen de particules fines de type black carbon (carbone suie) est de 0,8 µg/m³. La concentration d'ozone dans l'air est de 58 µg/m³. Le taux moyen de dioxyde d'azote est de 15,9 µg/m³. La couverture du ciel est très nuageuse, pour une température de 19,9 degrés.

Ce texte a été généré de manière automatique à partir de données publiques extraites en temps réel.

B. Structure du résumé du texte de la page "Statistiques"

L'indice de qualité de l'air a été [indice] pendant [nombre] jours. [nombre] phases d'avertissement 'fortes chaleurs et pic d'ozone' ont été enregistrées en Région bruxelloise depuis le 1er avril 2017. En moyenne, pour l'ensemble de la région, les recommandations de l'OMS ont été dépassées [nombre] fois en ce qui concerne le taux de particules fines de type PM10, et [nombre] fois en ce qui concerne celui des PM2.5. La recommandation de l'OMS relative au taux d'ozone a été dépassée à [nombre] reprises. La norme relative au taux de dioxyde d'azote [dépassement] (moyenne horaire). Le taux moyen de dioxyde d'azote est de [taux]. Il [évaluation] de [taux] la norme relative à la moyenne annuelle (soit [nombre] jours de dépassement). Le taux de [taux] de Black carbon, à partir duquel ses effets sont considérés comme plus nocifs pour la santé que les PM10 et PM2.5, a été franchi [nombre] fois.

Exemple de texte généré le 15/07/2017.



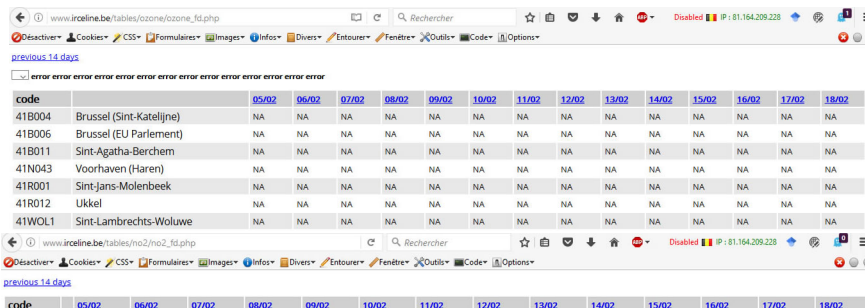
En bref

L'indice de qualité de l'air a été moyen pendant 14 jours. Il a été médiocre pendant 10 jours. Il a très médiocre pendant 2 jours. Il a été mauvais pendant 2 jours. Il n'a pas atteint le seuil maximal de 10 (exécration). 12 phases d'avertissement 'fortes chaleurs et pic d'ozone' ont été enregistrées en Région bruxelloise depuis le 1^{er} avril 2017. En moyenne, pour l'ensemble de la région, les recommandations de l'OMS ont été dépassées 0 fois en ce qui concerne le taux de particules fines de type PM10, et 2 fois en ce qui concerne celui des PM2.5. La recommandation de l'OMS relative au taux d'ozone a été dépassée à 22 reprises. La norme relative au taux de dioxyde d'azote n'a pas été dépassée (moyenne horaire). Le taux moyen de dioxyde d'azote est de 60µg/m³. Il dépasse de 20µg/m³ la norme relative à la moyenne annuelle (soit 88 jours de dépassement). Le taux de 1 µg/m³ de Black carbon, à partir duquel ses effets sont considérés comme plus nocifs pour la santé que les PM10 et PM2.5, a été franchi 34 fois.

Ce texte a été généré de manière automatique à partir de données publiques enregistrées de manière quotidienne, depuis le 01/04/2017.

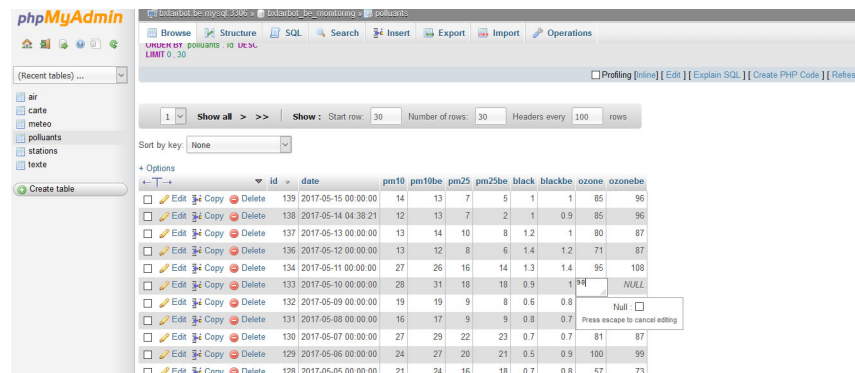
Annexe 13. Échantillon d'anomalies détectées dans le cadre du monitoring des données

1. Mesures relatives aux taux d'ozone et de dioxyde d'azote : valeurs manquantes (18/02/2017)



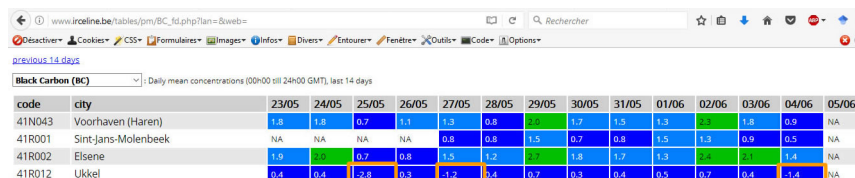
code		05/02	06/02	07/02	08/02	09/02	10/02	11/02	12/02	13/02	14/02	15/02	16/02	17/02	18/02
41B004	Brussel (Sint-Katelijne)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
41B006	Brussel (EU Parlement)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
41B011	Sint-Agatha-Berchem	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
41N043	Voorhaven (Haren)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
41R001	Sint-jans-Molenbeek	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
41R012	Ukkel	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
41W011	Sint-Lambrechts-Woluwe	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

2. Mesures relatives à l'ensemble des taux de polluants : valeurs modifiées a posteriori (17/05/2017)



id	date	pm10	pm10be	pm25	pm25be	black	blackbe	ozone	ozonebe
139	2017-05-15 00:00:00	14	13	7	5	1	1	85	96
138	2017-05-14 04:38:21	12	13	7	2	1	0.9	85	96
137	2017-05-13 00:00:00	13	14	10	8	1.2	1	80	87
136	2017-05-12 00:00:00	13	12	8	6	1.4	1.2	71	87
134	2017-05-11 00:00:00	27	26	16	14	1.3	1.4	95	109
133	2017-05-10 00:00:00	28	31	18	18	0.9	1	NULL	NULL
132	2017-05-09 00:00:00	19	19	9	8	0.6	0.8	81	87
131	2017-05-08 00:00:00	16	17	9	9	0.8	0.7	81	87
130	2017-05-07 00:00:00	27	29	22	23	0.7	0.7	81	87
129	2017-05-06 00:00:00	24	27	20	21	0.5	0.9	100	99
128	2017-05-05 00:00:00	21	24	16	18	0.7	0.8	57	73

3. Mesures relatives au taux de black carbon : valeurs négatives (05/06/2017)



code	city	23/05	24/05	25/05	26/05	27/05	28/05	29/05	30/05	31/05	01/06	02/06	03/06	04/06	05/06
41N043	Voorhaven (Haren)	1.8	1.8	0.7	1.1	1.3	0.8	0.5	1.7	1.5	1.3	0.5	1.8	0.9	NA
41R001	Sint-jans-Molenbeek	NA	NA	NA	NA	0.8	0.8	1.5	0.7	0.8	1.5	1.3	0.9	0.5	NA
41R002	Elsene	1.9	0.7	0.7	0.8	1.5	1.2	0.5	1.8	1.7	1.3	0.5	0.5	1.4	NA
41R012	Ukkel	0.4	0.4	-2.8	0.3	-1.2	0.4	0.7	0.3	0.4	0.5	0.7	0.4	-1.4	NA

4. Valeurs manquantes : panne d'électricité sur le réseau (02/07/2017)

Dimanche 02/07/2017 : 06h00 mise à jour (heure locale)
Cliquez sur les codes stations pour avoir plus d'information.

code	commune	NO2	CO	SO2	PM10	PM2.5	COH6
		mooyenne 1 heure	mooyenne 8 heures	mooyenne glissante 24 heures	mooyenne glissante 24 heures	mooyenne glissante 24 heures	mooyenne glissante 24 heures
		µg/m³	µg/m³	µg/m³	µg/m³	µg/m³	µg/m³
Région de Bruxelles-Capitale							
41B001	Bruxelles (Arts-Loi)		NA	NA	NA		
41B004	Bruxelles (Sainte-Catherine)		NA	NA	NA		
41B006	Bruxelles (Parlement UE)		NA	NA	NA		NA
41B008	Bruxelles (Rue Belliard)		NA	NA	NA		
41B011	Berchem-Sainte-Agathe		NA		NA	NA	
41MEU1	Neder-Over-Heembeek		NA		NA	NA	NA
41N043	Avant-port (Haren)		NA	NA	NA	NA	NA
41R001	Molenbeek-Saint-jean		NA	NA	NA	NA	NA
41R002	Ixelles		NA	NA	NA	NA	NA
41R012	Ukkel		NA	NA	NA	NA	NA
41W011	Woluwe-Saint-Lambert		NA	NA	NA	NA	NA
41X013	Forest		8				

Bulletin du 3 juillet 2017

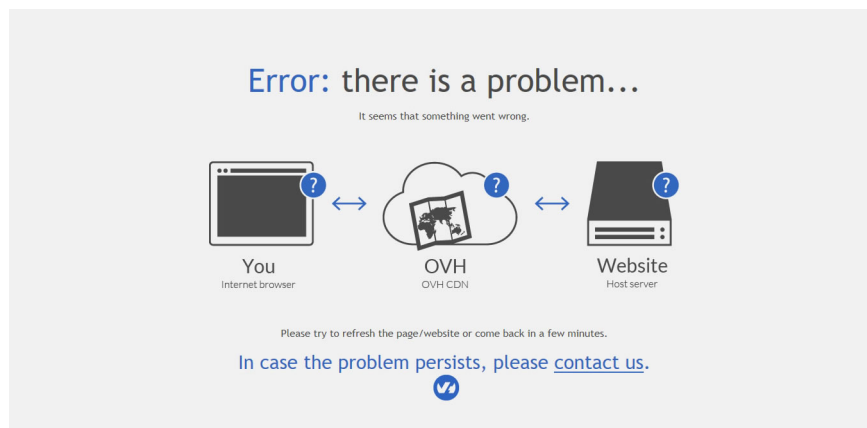
Pas d'alerte pollution

10h33. Plusieurs données ne sont pas disponibles aujourd'hui, l'indice de qualité de l'air y compris. La teneur moyenne en black carbon (carbone suie) est de 1,8 µg/m³. La concentration

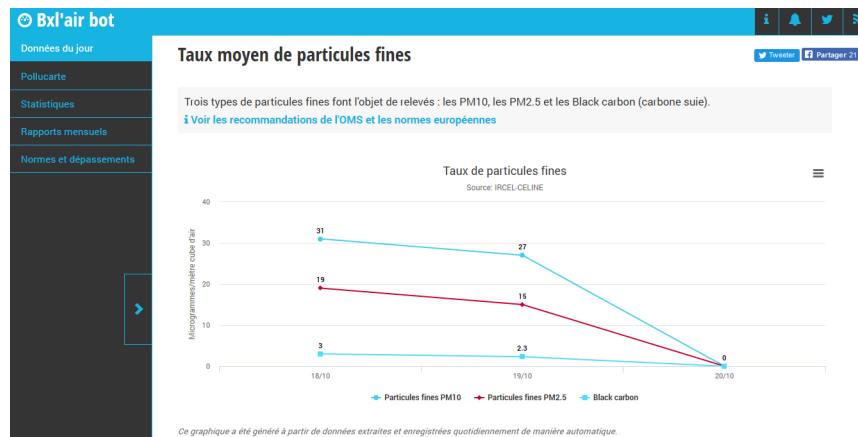
d'ozone dans l'air est de 39 µg/m³. Le taux moyen de dioxyde d'azote est de 38,6 µg/m³. Le ciel est peu nuageux, pour une température de 18,1 degrés.

Ce texte a été généré de manière automatique à partir de données extraites en temps réel.

5. Observations de l'IRM : panne du serveur web (05/10/2017)



6. Anomalie dans l'enregistrement des données (21/10/2017)



Annexe 14. "Bxl'air bot" : tableau de bord (modules d'analyses statistiques)

Bxl'air bot

[i](#)
[Twitter](#)
[Partager](#)

Données du jour

Pollucarte

Statistiques

Rapports mensuels

Téléchargements

Normes et dépassements

Statistiques (du 01/04/2017 au 31/03/2018)

En bref

L'indice de qualité de l'air a été moyen pendant 28 jours. Il a été médiocre pendant 14 jours. Il a très médiocre pendant 3 jours. Il a été mauvais pendant 3 jours. Il n'a pas atteint le seuil maximal de 10 (exécutable). 14 phases d'information (niveau d'alerte 0) ont été enregistrées en Région bruxelloise depuis le 1er avril 2017. En moyenne, pour l'ensemble de la région, les recommandations de l'OMS ont été dépassées 2 fois en ce qui concerne le taux de particules fines de type PM10, et 20 fois en ce qui concerne celui des PM2.5. La recommandation de l'OMS relative au taux d'ozone a été dépassée à 28 reprises. La norme relative au taux de dioxyde d'azote n'a pas été dépassée (moyenne horaire). Le taux moyen de dioxyde d'azote est de 58,54 µg/m³. Il dépasse de 18.4µg/m³ la norme relative à la moyenne annuelle. Le seuil de 1 µg/m³ de black carbon (carbone suie), à partir duquel ses effets sont considérés comme plus nocifs pour la santé que les PM10 et PM2.5, a été franchi 157 fois. Le taux de 3 µg/m³ de black carbon a été dépassé 8 fois.

Ce texte a été généré de manière automatique à partir de données publiques (open data) non validées enregistrées de manière quotidienne pendant un an, du 1er avril 2017 au 30 mars 2018. Celles-ci donnent donc une indication sur la qualité de l'air bruxellois sur une période de douze mois. Retrouvez un dossier spécial « qualité de l'air à Bruxelles » dans le numéro d'Alter Echos publié en mai 2018. Source : CELINE, Cellule interrégionale d'Environnement.

Indice de la qualité de l'air

Septembre 2017

Septembre 2017

Très bon : 5 jours
Bon : 15 jours
Assez bon : 8 jours
Moyen : 2 jours

Niveau d'alerte

Septembre 2017

Septembre 2017

Aucune phase d'information : 30 jours

Moyennes par station

Sélectionner une station

Arts Loi

Sélectionner une période

Septembre 2017

OK

Polluants atmosphériques

Septembre 2017

Septembre 2017

Moyenne (région bruxelloise)*

PM10
Aucun dépassement n'a été constaté.

PM2.5
La recommandation de l'OMS a été dépassée 2 fois.

Ozone
Aucun dépassement n'a été constaté.

Dioxyde d'azote
La recommandation de l'OMS et la norme européenne ont été dépassées 25 fois (moyenne annuelle). Aucun dépassement n'a été constaté (moyenne horaire). La moyenne mensuelle est de 59µg/m³. Cette moyenne dépasse de 19µg/m³ les norme et recommandation (moyenne annuelle).
Black carbon (carbone suie) : si ces très fines particules ne font l'objet d'aucune recommandation ou norme, le seuil de 1 µg/m³ a été dépassé 12 fois.

*Le mode de calcul s'appuie sur une moyenne par 24 heures non-validée. Une marge d'erreur est donc possible.

Maximums par polluant

PM10
65 µg/m³ à Haren, le 28/08

PM25
57 µg/m³ à Berchem-Sainte-Agathe, le 03/03

Black carbon
9.3 µg/m³ à Haren, le 07/11

Ozone
190 µg/m³ à Berchem-Sainte-Agathe, le 20/06

Dioxyde d'azote
168 µg/m³ à Arts Loi, le 28/08

Moyennes par station

Sélectionner une station

Haren

Sélectionner une période

Juin 2017

OK

Juin 2017

Haren

PM10 : 24
PM25 : 12
Black carbon : 1.4
Ozone : 87
Dioxyde d'azote : **59**

Annexe 15. "Bxl'air bot" : modèle de système d'alerte automatisé

Bxl'air Bot

alter échos

Alter Échos vous alerte

La teneur de l'air en particules fines de type PM10 est de 10,3 microgrammes/mètre cube. Le taux moyen de PM25 est de 8 µg/m³. Le taux moyen de black carbon (carbone suie) est de 0,7 µg/m³. La concentration d'ozone dans l'air est de 41 µg/m³. Le taux moyen de dioxyde d'azote est de 11 µg/m³.

Bruxelles Environnement indique que l'indice de la qualité de l'air est très bon.

*Ce bulletin a été généré de manière automatique à partir de données publiques extraites en temps réel.
Source : IRCELINE*

[En savoir plus sur www.bxlairbot.be](http://www.bxlairbot.be), le newsbot qui surveille la qualité de l'air bruxellois

À découvrir sur Alter Échos

Crise politique: ça va trancher?

Le cdH a donc décidé de tirer la prise des trois gouvernements auxquels il participait. En Région wallonne, à Bruxelles et en Communauté française, les humanistes ont préféré se séparer du PS, englué jusqu'au cou dans diverses «affaires». Avec quelles conséquences pour le secteur associatif?

[Lire plus](#)

Lendemain de crise

L'étendue de la casse... C'est ce que nous avons tenté de mesurer dans ce numéro d'Alter Échos. Quelles conséquences le geste de Benoit Lutgen a-t-il eu sur l'avancée des grands chantiers que nous suivons depuis le début de la législature dans les pages de ce

Structure (phrase à trous)

Sans dépassement :

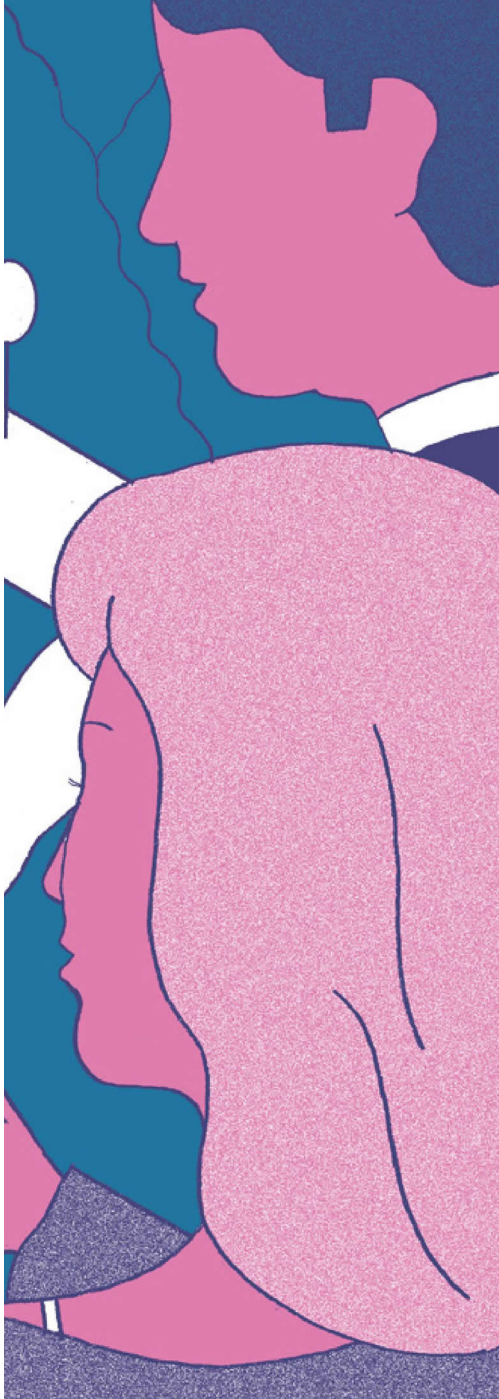
La teneur de l'air en particules fines de type PM10 est de [valeur]. Le taux moyen de PM25 est de [valeur]. Le taux moyen de black carbon (carbone suie) est de [valeur]. La concentration d'ozone dans l'air est de [valeur]. Le taux moyen de dioxyde d'azote est de [valeur]. Bruxelles Environnement indique que l'indice de la qualité de l'air est [valeur].

Exemple avec dépassement pour l'ozone :

La teneur de l'air en particules fines de type PM10 est de [valeur]. Le taux moyen de PM25 est de [valeur]. Le taux moyen de black carbon (carbone suie) est de [valeur]. La concentration d'ozone dans l'air est de [valeur]. Ce taux dépasse de [valeur - norme] la norme recommandée par l'Organisation mondiale de la santé. Le taux moyen de dioxyde d'azote est de [valeur]. Bruxelles Environnement indique que l'indice de la qualité de l'air est [valeur].

Annexe 16. Article publié dans le dossier "Pollution, l'air de rien"





Pollution, l'air de rien

Il y a l'air wallon, que l'on pourrait croire bon, et l'air bruxellois, qui ne l'est sûrement pas. L'air officiel, qui sent le printemps, et l'air « alternatif », qui voit tout en noir. Celui des riches, qui peuvent en changer à chaque voyage, et celui que les pauvres se coltinent toute une vie. Ce sont, en somme, des airs variés pour une réalité atmosphérique au cœur de nombreuses batailles juridiques, citoyennes et politiques.

En Wallonie, on a longtemps cru qu'il n'y avait pas de quoi s'inquiéter de la pollution. Aujourd'hui, Greenpeace attaque la Région wallonne en justice : elle n'effectuerait pas correctement les mesures de la qualité de l'air, ce qui empêcherait les citoyens d'être conscients du problème (lire « La pollution est sous-estimée en Wallonie » p. 23).

Or, la connaissance, c'est la base du changement. « *Le but, c'est de constater ou de s'enfuir ? Ou changer les choses ?* », interpelle une Molenbeekoise engagée dans un processus de « science citoyenne ». Grâce à un petit capteur, elle pourra bientôt connaître « son » air et celui de ses enfants. Et, à son niveau, en appeler au changement (lire « Souffle citoyen sur Molenbeek » p. 28). C'est avec cette même envie de connaissance que nous avons accueilli à la rédaction d'*Alter Échos* un robot de collecte et d'analyse de données liées à la qualité de l'air. Il nous permet aujourd'hui de poser un œil critique sur les informations communiquées par les autorités bruxelloises (lire « Un an avec un robot » p. 16).

À Bruxelles comme en Wallonie, la lutte politique contre la pollution a pris la forme d'une chasse au vieux diesel – une initiative peu populaire mais nécessaire (lire « Air wallon : Carlo Di Antonio 'prend des dispositions' » p. 24 et « LEZ : la zone qui cache la forêt » p. 26). Suffira-t-elle à ce que l'air de demain soit de bonne qualité, dans tous les quartiers et pour tous les citoyens (lire « Sommes-nous tous égaux face à la pollution de l'air ? » p. 20) ? Il est encore permis d'en douter... ♦

Un dossier illustré par Adrien Herda.

Un an avec un robot

Pendant douze mois, *Alter Échos* a accueilli le Bxl'air bot à la rédaction. Cette application de datajournalisme nous a aidés à enregistrer, compiler et compter des données sur la qualité de l'air. Le résultat ? Pour vous, de l'info inédite sur la pollution à Bruxelles. Pour nous, une première expérience humano-robotique.

PAR CÉLINE GAUTIER

Il a mis le pied dans la porte de la rédaction en mars 2017 sans qu'on l'ait vraiment invité... Le Bxl'air bot est arrivé avec l'enthousiasme de sa conceptrice, la journaliste et développeuse Laurence Dierickx, dans le cadre de son doctorat en information et communication à l'ULB. Son idée : tester, pendant un an et pour la première fois en Belgique, l'immersion d'un robot d'information (ou « newsbot ») dans un média.

Le Bxl'air bot ne prend pas beaucoup de place et ne sert pas le café. Il s'agit d'une simple application qui a fait son nid sur notre site internet. Du 1^{er} avril 2017 au 31 mars 2018, ce « baby bot » a audité, chaque jour, la qualité de l'air dans la capitale, sur la base des données publiées par CELINE, la Cellule interrégionale de l'environnement. Il a produit, minutieusement, son petit rapport quotidien, encore disponible sur <http://bxlairbot.be/>

Jusqu'ici, les données brutes fournies par CELINE étaient surtout utilisées par les autorités régionales bruxelloises pour communiquer au grand public un indice de la qualité de l'air (voir qualitedelair.brussels/) et pour rendre des comptes à l'Europe sur les niveaux de pollution. L'intérêt du robot, c'est qu'il peut automatiser des calculs que les journalistes pourraient faire manuellement avec les données de CELINE mais qu'ils n'auraient – pour être honnêtes – jamais le temps et la patience de faire.

DE L'OBJECTIVITÉ DU ROBOT-JOURNALISTE

Tout comme les autorités, qui interprètent les données d'une façon rassurante pour le public (voir plus

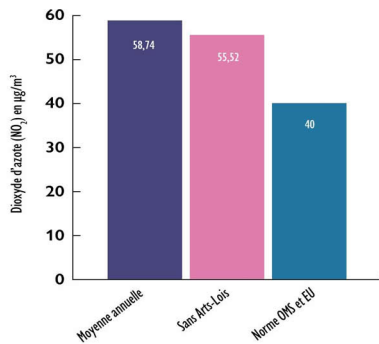
loin notre graphique « La communication des autorités »), le robot assume la part de subjectivité que comporte toute interprétation d'informations. « *Le robot a une approche journalistique*, défend Laurence Dierickx. *Il tient compte de la santé publique.* » En clair, cela signifie que, pour chaque résultat publié, Bxl'air bot le met en relation avec les normes – non pas celles fixées par l'Europe et qui tiennent compte de réalités économiques (et du poids de la bagnole dans notre économie) mais celles proposées par l'Organisation mondiale de la santé (OMS) qui visent, plutôt, à protéger nos cœurs et nos poumons.

Par exemple, quand CELINE enregistre une moyenne de 27 microgrammes par mètre cube ($\mu\text{g}/\text{m}^3$) de particules fines PM10, l'Europe inspire à fond (elle tolère une moyenne annuelle de $40 \mu\text{g}/\text{m}^3$), alors que l'OMS se tord de douleur (elle recommande de ne pas dépasser $20 \mu\text{g}/\text{m}^3$). C'est donc en se basant sur des normes strictes et plus respectueuses de notre santé que le robot conclut, après un an d'enregistrement : « *En moyenne, pour l'ensemble de la Région, les recommandations de l'OMS ont été dépassées deux fois en ce qui concerne le taux de particules fines de type PM10, et 20 fois en ce qui concerne celui des PM 2,5. La recommandation de l'OMS relative au taux d'ozone a été dépassée à 28 reprises.* » Pour rappel, la mauvaise qualité de l'air serait responsable, selon l'Agence européenne de l'environnement, de 12.000 décès prématurés en Belgique par an (pneumonies, cancers, accidents cardiovasculaires, etc.). Ces chiffres valaient bien un robot.

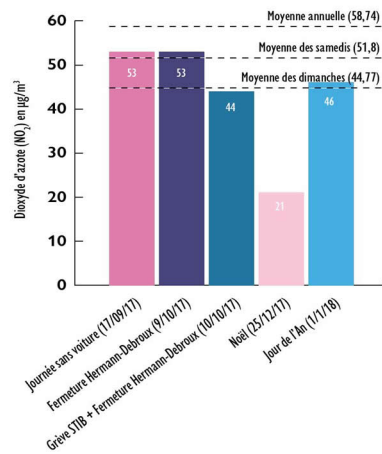
GRAPHIQUE 1*
BRUXELLES EN INFRACTION

Moyennes annuelles de dioxyde d'azote (NO₂) en µg/m³

CELINE, la Cellule interrégionale de l'environnement qui gère les stations de mesure, enregistre des données à Arts-Loi. Sur ce carrefour très embouteillé, les concentrations de dioxyde d'azote (NO₂), un gaz lié au transport routier et au chauffage, sont les plus élevées de la Région bruxelloise. Ces données sont bien disponibles sur le site de CELINE mais ne sont pas prises en compte dans les moyennes, ni pour communiquer au public l'indice de la qualité de l'air ni pour rendre des comptes à l'Europe sur les niveaux de pollution en Belgique. Argument de la Région : Arts-Loi n'est pas représentatif de Bruxelles car trop pollué (la station de mesure est trop proche des voitures). Cette interprétation a été dénoncée (jusqu'au tribunal), depuis des années, par des militants pour la qualité de l'air, qui estiment au contraire qu'il faut tenir compte des pires situations, comme des meilleures, pour avoir un tableau complet de l'enfumage bruxellois.



Frondeur dans l'âme, notre petit robot nous fournit donc une moyenne annuelle de la concentration en dioxyde d'azote à Bruxelles en tenant compte des mesures d'Arts-Loi. Le résultat obtenu (58,74 µg/m³) est bien plus élevé que la moyenne calculée sans Arts-Loi (55,52). Mais dans les deux cas, on dépasse – et de loin – la limite de 40 µg/m³ fixée tant par l'Organisation mondiale de la santé (OMS) que par l'Union européenne (UE). Si le dioxyde d'azote (NO₂) était un gaz hilarant, à Bruxelles, on serait déjà tous morts de rire.

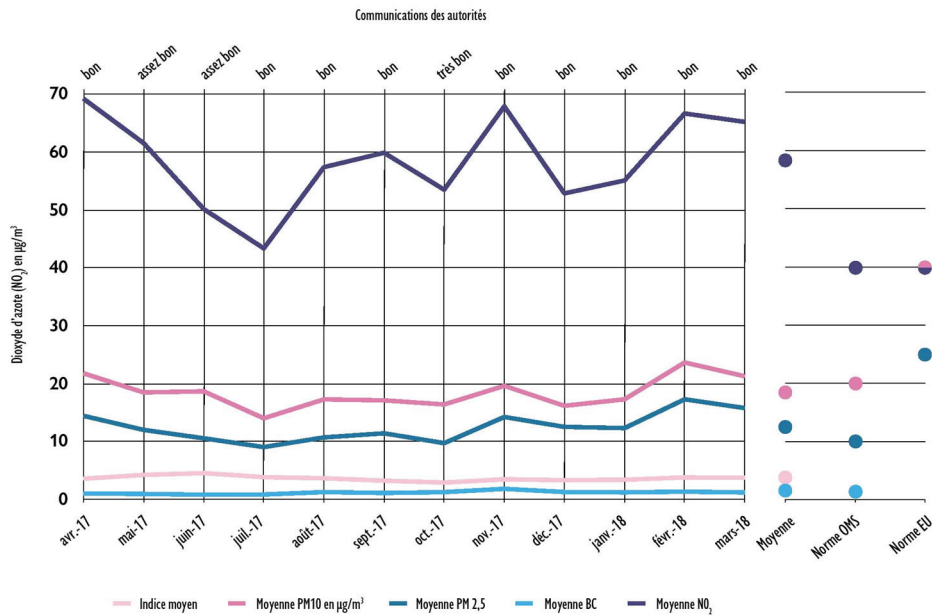


GRAPHIQUE 2
L'EFFET DU DIMANCHE
Dioxyde d'azote (NO₂) en µg/m³

Ce graphique montre l'effet du week-end sur les émissions de dioxyde d'azote. Par rapport aux jours de semaine, la pollution diminue le samedi et baisse encore notablement le dimanche. Ces jours-là, il y a en effet moins de voitures et moins de chauffage dans les bureaux et les collectivités.

Nous avons également repris dans le tableau les résultats enregistrés certains jours de l'année. Pour pouvoir faire de réelles comparaisons, il faudrait tenir compte des conditions climatiques et répéter l'expérience plusieurs années d'affilée. Mais notons, à titre indicatif, que Noël était le jour où nous pouvions respirer le plus à l'aise. Et que les jours de grève ou de fermeture d'un viaduc, les résultats semblent se rapprocher de ceux d'un week-end. Parce qu'en cas de force majeure, on finit par prendre moins sa voiture, privilégier le covoiturage ou travailler de chez soi? ➔

Les graphiques tiennent compte des données enregistrées par le robot entre le 1^{er} avril 2017 et le 31 mars 2018 sur le territoire de Bruxelles-Capitale.



➔ **GRAPHIQUE 3 – LA COMMUNICATION DES AUTORITÉS: TOUT VA TRÈS BIEN**

Si l'on fait la moyenne par mois des indices de la qualité de l'air tels qu'ils sont communiqués par Bruxelles Environnement, on se dit que la pollution, il n'y a vraiment pas de quoi en faire un dossier spécial. Le pire du pire à Bruxelles, en moyenne mensuelle, sur la période d'enregistrement, c'est que l'air soit « assez bon ». Voyons voir...

PM 10. On ne dépasse en effet qu'une seule fois (en avril 2017) le seuil fixé par l'OMS (20 µg/m³ de moyenne annuelle) et jamais le seuil fixé par l'Union européenne (40 µg/m³ de moyenne annuelle).

PM 2,5. Bruxelles fait déjà moins la maligne : il n'y a qu'au mois de juillet et octobre 2017 qu'elle est en dessous du seuil fixé par l'OMS (10 µg/m³ de moyenne annuelle), même si elle ne dépasse jamais celui de l'UE (25 µg/m³ de moyenne annuelle).

Black carbon. Il n'existe pas de seuil légal, mais l'OMS estime que 1 µg/m³, c'est déjà trop – soit la situation que vivent les Bruxellois 10 mois sur 12.

Dioxyde d'azote (NO₂). Sur une année, il n'y a pas un mois où l'on ne dépasse pas la norme OMS/UE fixée à 40 µg/m³ de moyenne annuelle. Pour rappel, le robot tient compte des résultats enregistrés à Arts-Loi, ce que l'IBGE ne fait pas.

Ozone (O₃). Non repris dans ce graphique. Mais le robot nous informe que « la recommandation de l'OMS relative au taux d'ozone a été dépassée à 28 reprises » sur l'année écoulée.

En résumé, sur l'année de mesures, « l'indice de qualité de l'air a été moyen pendant 28 jours. Il a été médiocre pendant 14 jours. Il a très médiocre pendant 3 jours. Il a été mauvais pendant 3 jours. Il n'a pas atteint le seuil maximal de 10 (exécutable). » Paroles de robot.

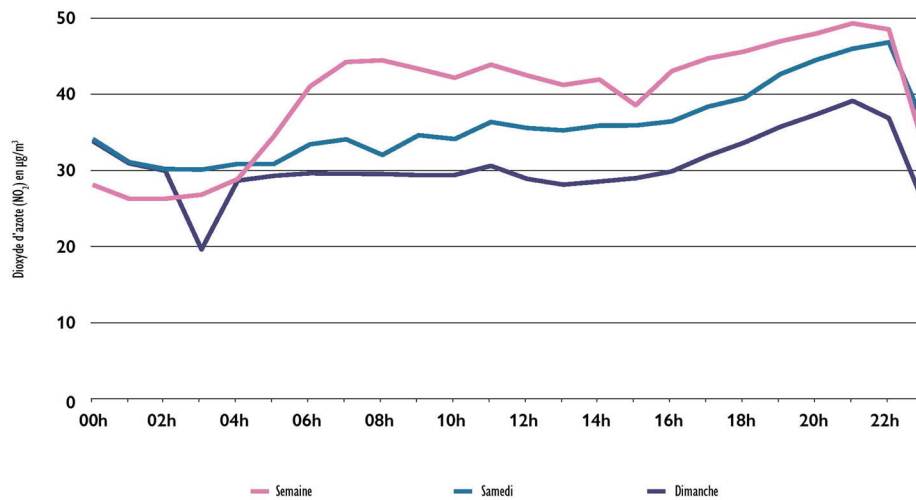
L'indice de la qualité de l'air n'est jamais qu'un outil de communication envers le grand public. Il repose forcément sur une simplification des résultats de mesure et sur des choix arbitraires. De toute évidence, les autorités bruxelloises et leur « bon » air ont à cœur de ne pas inquiéter les foules. « Assez troublant », n'est-ce pas ?

GRAPHIQUE 4
LA JOURNÉE TYPE D'UN RESPIRANT BRUXELLOIS
 Moyenne annuelle de dioxyde d'azote (NO₂),
 heure par heure.

Quel est le meilleur moment pour promener votre enfant asthmatique dans le quartier ou pour grimper le Mont-des-Arts à vélo? Le dimanche à 3 heures du matin. En semaine, les concentrations de dioxyde d'azote grimpent progressivement en fin de nuit jusqu'à atteindre des seuils inquiétants dès 7 heures du matin. Aux heures où vous conduisez les enfants à l'école, l'air est déjà bien vicié. Si l'on se réfère plutôt aux particules fines (PM 10 et PM 2,5), non reprises dans ce tableau, c'est alors vers minuit qu'il faut envisager de respirer.

Concernant les émissions de dioxyde d'azote, liées notamment au trafic automobile, il n'existe pas – ou plus – d'heure de pointe et d'heure creuse dans une journée type à Bruxelles. Parce que ça bouchonne à toute heure? La pollution stagne toute la journée. Il faut attendre 22 heures pour retrouver une situation plus acceptable.

Ce graphique est réalisé grâce aux moyennes de toutes les stations. Il ne reflète donc pas la situation des Bruxellois les plus pollués ni celle des moins pollués, mais celle du respirant «moyen» dans un quartier «moyen». Le robot tient compte de deux enregistrements par heure (et non d'un enregistrement en continu) avec, parfois, des résultats inutilisables. Il s'agit donc d'une information précieuse, qui met en lien la pollution et l'activité humaine, mais qui mériterait d'être nuancée par d'autres types de mesures. ♦



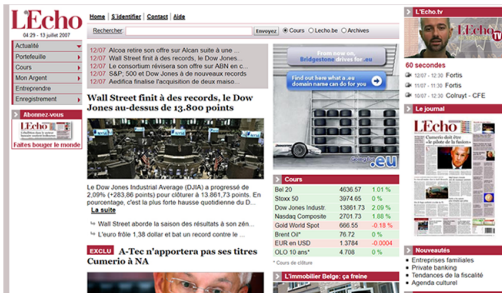
Annexe 17. Évolution de l'interface du site web de *L'Echo*, de 2002 à 2019



2002



2006



2006



2006



2006



2008



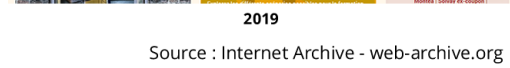
2011



2014



2018



2019

Source : Internet Archive - web-archive.org

Annexe 18. Modes de calcul des indicateurs automatiques

La mise en œuvre de ces formules a été réalisée via une interface conçue et développée spécifiquement : <https://www.ohmybox.info/linguistics/fr>

Score de Flesch-Kincaid

Formule : $206.835 - 1.015 \times (\text{nbre_mots}/\text{nbre_phrases}) - 84.6 \times (\text{nbre_syllabes}/\text{nbre_mots})$

Index (indice) Coleman-Yau

Formule : $5.89 \times (\text{nbre_caractères}/\text{nbre_mots}) - 0.3 \times (\text{nbre_phrases}/\text{nbre_mots}) - 15.8$

Grade Flesch-Kincaid

Formule : $0.39 \times (\text{nbre_mots}/\text{nbre_phrases}) + 11.8 \times (\text{nbre_syllabes}/\text{nbre_mots}) - 15.59$

Index (indice) de lisibilité automatisé (ARI)

Formule : $4.71 \times (\text{nbre_caractères}/\text{nbre_mots}) + 0.5 \times (\text{nbre_mots}/\text{nbre_phrases}) - 21.43$

Distance de Levenshtein

Algorithme en langage de programmation Javascript

```
function distance(a, b) {
  var n = a.length, m = b.length, matrice = [];
  for(var i=-1; i < n; i++) {
    matrice[i]=[];
    matrice[i][-1]=i+1;
  }
  for(var j=-1; j < m; j++) {
    matrice[-1][j]=j+1;
  }
  for(var i=0; i < n; i++) {
    for(var j=0; j < m; j++) {
      var cout = (a.charAt(i) == b.charAt(j)) ? 0 : 1;
      matrice[i][j] = minimum(1+matrice[i][j-1], 1+matrice[i-1][j], cout+matrice[i-1][j-1]);
    }
  }
  return matrice[n-1][m-1];
}
```

Fonction native de PHP

levenshtein (string \$str1 , string \$str2) : int

Taux de similarité

similar_text(\$source, \$cible, \$similar); echo round(\$similar,2). "%" ;

Annexe 19. Formulaire d'évaluation diffusés via Google Forms dans le cadre de l'expérience "Quotebot"

Formulaire d'évaluation en ligne (première évaluation)

Q1. Avez-vous participé à la correspond de "Quotebot" ?

(Oui / Non)

Q1.a. Si oui, est-ce parce que :

vous étiez curieux

vous vous sentiez contraints

vous étiez intéressé par le projet

vous trouviez ce projet amusant

autre (précisez)

Q1.b. Si non, est-ce parce que :

vous n'en aviez pas d'intérêt

vous n'aviez pas le temps

vous n'étiez pas présent à la rédaction

vous ne trouviez pas le projet amusant

autre (si autre, précisez)

Q2. "Quotebot" correspond-t-il à vos attentes ?

(Oui / Non)

Q.2.a. Si non, pourquoi ?

les textes nécessitent des retraitements et cela va plus vite d'écrire soi-même

cela prend plus de temps de travailler avec "Quotebot"

autre (si autre, précisez)

Q2.b. Si oui, pourquoi ?

les textes peuvent être publiés tels quels

"Quotebot" vous fait gagner du temps

autre (si autre, précisez)

Q3. Trouvez-vous "Quotebot" écrit-il comme vous ? (Oui / Non)

Q3a. Si non, est-ce parce que :

il est trop standardisé

il est trop répétitif

il commet des erreurs

autre (si autre, précisez)

Etes-vous d'accord avec les affirmations suivantes?

(Echelle de Likert de 1 à 5)

Q4. Les productions de "Quotebot" sont lisibles.

Q5. Les productions de "Quotebot" ont du sens.

Q6. Les productions de "Quotebot" sont claires.

Q7. Les productions de "Quotebot" sont exactes.

Q8. Les productions de "Quotebot" sont fiables.

Q9. Les productions de "Quotebot" sont utilisables.

Q10. Les productions de "Quotebot" sont utiles.

Q11. Les productions de "Quotebot" sont cohérentes.

Q12. Pensez-vous que "Quotebot" doit être amélioré?

Q.12 a. Si oui, pourquoi?

Q13. Souhaitez-vous déposer un commentaire libre à propos de votre expérience de "Quotebot"?

Formulaire d'évaluation en ligne (deuxième évaluation)

Q1. Avez-vous pris connaissance de la dernière version de "Quotebot"?

(Oui / Non)

Q1.a. Si non, pourquoi :

je n'ai pas eu le temps

je ne suis pas intéressé/e par ce projet

je ne suis plus intéressé/e par ce projet

je n'avais plus ce projet en tête

Q1.b. Si oui, je pense que... (plusieurs réponses possibles)

Les productions de Quotebot ont été améliorées

Les productions de Quotebot sont utilisables

Les productions de Quotebot devraient encore être améliorées

Les productions de Quotebot ne sont pas encore utilisables

Autre (si autre, précisez)

Q2. "Quotebot" correspond-t-il à vos attentes?

(Oui / Non)

Q.2.a. Si non, pourquoi?

les textes nécessitent des retraitements et cela va plus vite d'écrire soi-même

cela prend plus de temps de travailler avec "Quotebot"

autre (si autre, précisez)

Q2.b. Si oui, pourquoi?

les textes peuvent être publiés tels quels

"Quotebot" vous fait gagner du temps

autre (si autre, précisez)

Q3. Trouvez-vous "Quotebot" écrit-il comme vous? (Oui / Non)

Q3a. Si non, est-ce parce que :

il est trop standardisé

il est trop répétitif

il commet des erreurs

autre (si autre, précisez)

Etes-vous d'accord avec les affirmations suivantes?

(Echelle de Likert de 1 à 5)

Q4. Les productions de "Quotebot" sont lisibles.

Q5. Les productions de "Quotebot" ont du sens.

Q6. Les productions de "Quotebot" sont claires.

Q7. Les productions de "Quotebot" sont exactes.

Q8. Les productions de "Quotebot" sont fiables.

Q9. Les productions de "Quotebot" sont utilisables.

Q10. Les productions de "Quotebot" sont utiles.

Q11. Les productions de "Quotebot" sont cohérentes.

Q12. Pensez-vous que "Quotebot" doit être amélioré?

Q.12 a. Si oui, pourquoi?

Q13. Souhaitez-vous déposer un commentaire libre à propos de votre expérience de "Quotebot"?

Formulaire "Présentation du service 'Investir' de *L'Echo*"

Q1. Vos nom et prénom *

Q2. En quelle année êtes-vous entré(e) à *L'Echo* et quand y avez-vous rejoint le service boursier? *

Q3. Quel est votre parcours académique? *

Q4. Quel est votre parcours professionnel en dehors de *L'Echo*?

Q5. Quel âge avez-vous? *

* *champs obligatoires*

Annexe 20. Échantillon de textes rédigés par les journalistes du service "Investir" de *L'Echo*

Les bilans

OUVERTURE - priorité 1

- Bruxelles : 9h
- Indices européens : Bel20, Dax, Cac40, AEX, Footsie, Stoxx 600 : 9h
- Wall Street : 15h30 (ou 14h30 changement d'heure!)

Échantillon de texte : ouverture à la hausse à la bourse de Bruxelles

La Bourse de Bruxelles était bien orientée dans les premiers échanges, avec un indice Bel20 qui gagnait XX%. Au sein de l'indicateur, XX valeurs étaient en hausse, XX inchangée(s) et XX en baisse. Les meilleures progressions étaient à mettre sur le compte de/d' Valeurs 1 (+XX%), de/d' Valeurs 2 (+XX%) et de/d' Valeurs 3 (+XX%). À l'autre bout du spectre, on retrouvait Valeurs 4 (-XX%), Valeurs 5 (-XX%) et Valeurs 6 (-XX%). Hors Bel20, on retiendra les bonds de/d' Valeurs 7 (XX%), Valeurs 8 (XX%) et Valeurs 9 (XX%), ou encore les pertes encaissées par Valeurs 10 (XX%), Valeurs 11 (XX%) et Valeurs 12 (XX%).

La Bourse de Bruxelles était bien orientée dans les premiers échanges, avec un indice Bel20 qui gagnait 0,72% à 3.913,72 points. Au sein de l'indicateur, 5 valeurs étaient en hausse, 2 inchangée(s) et 13 en baisse. Les meilleures progressions étaient à mettre sur le compte d'AB InBev (+2%), de Solvay (+1,7%) et d'Umicore (+1%). À l'autre bout du spectre, on retrouvait Aegas (-3%), Proximus (-1,9%) et Cofinimmo (-0,82%). Hors Bel20, on retiendra les gains de Mithra (+14%), d'Aedifica (+3%) et de Sapec (+1,7%). À l'opposé, Montea cédait 2%, Ter Beke 1,3% et Curetis 1,2%.

OU

Emmenée par Valeurs 1 (+XX%), Valeurs 2 (+XX%) et Valeurs 3 (+XX%), la Bourse de Bruxelles progressait dans les premières minutes de la séance, comme le reflétait le gain de XX% de l'indice Bel20. En queue de peloton, Valeurs 4, Valeurs 5 et Valeurs 6 pesaient sur la tendance et rétrogradaient de respectivement XX%, XX% et XX%.

CLOTURE - priorité 1

- Bruxelles : 17h40 + graphiques
- Indices européens : Bel20, Dax, Cac40, AEX, Footsie : 17h40
- Stoxx 600 : tendance + Top 3 indices sectoriels & Flop 3 : 18h
- Texte supplémentaire en effet. Il n'est pas nécessaire d'avoir un top/flop Stoxx 600 à l'ouverture
- Wall Street : 22h ou 22h20 (S&P 500 & Nasdaq)

TOP/FLOP - priorité 2

- Bel20 à 17h40
- Stoxx 600 : tendance + Top 3 indices sectoriels & Flop 3 : 18h
- Bonus / eFast soiriste

- Top/flop Dow Jones - après 22h
- Top/flop Nasdaq 100 - après 22h

APERCU À 18h (commentaire 3) + graphiques - priorité 3

- Euro/dollar
- Pétrole Brent
- Or
- Taux Olo / Bund / Treasury

HIT PARADE : classement - 4

- Ensemble du Bel 20 / Classement par performance
- Top 10 / Flop 10 : Euronext Bruxelles/| Classement par performance
- Bel20, Dax, Cac40, AEX, Footsie, Stoxx 600, Tokyo, Wall Street / Classement par performance

MI-SEANCE - priorité 5

- Bruxelles : 12h-13h + graphiques
- Indices européens : Bel20, Dax, Cac40, AEX, Footsie, Stoxx 600 : 12h-13h
- Wall Street : 19h

Les alertes

1. "Flash" : cotation d'une valeur (action, obligation, indice, devise) qui sur ou sous-performe : envoyer un message.

Bourse de "XXXX" : "VALEUR X" chute/bondit de XX% dans les premiers échanges (situation où la variation se produit pour le premier cours enregistré du jour à l'ouverture par rapport à la clôture de la veille).

Bourse de "XXXX" : "VALEUR X" chute/s'envole de XX% en cours de séance

Bourse de "XXXX" : "VALEUR X" accélère ses gains/sa chute en cours de séance dans un marché globalement en hausse/baisse de XX%

Bourse de "XXXX" : la "VALEUR X" a désormais effacé ses gains/pertes de l'année (Ref. évolution en pourcentage de l'action entre le cours actuel et le cours de clôture de l'action de la dernière séance de l'année précédente).

Bourse de "XXXX" : l'indice X a désormais effacé ses gains/pertes de l'année (Ref. évolution en pourcentage de l'action entre le cours actuel et le cours de clôture de l'action de la dernière séance de l'année précédente).

2. Valeurs reprises dans les principaux indices européens : Bel20, Bel Mid, Bel Small, Dax, Cac40, AEX, EuroStoxx50, Footsie (Temps réel?), Stoxx 600.

Si une valeur fluctue entre X et Y, envoyer une alerte : quelles actions, dans quel volume (par rapport à la moyenne des 30 derniers jours).

Bourse de "XXXX" : L'action XX se négocie à la baisse/hausse | monte/baisse dans un volume supérieur à la moyenne des trente derniers jours.

OU

Bourse de "XXXX" : la "VALEUR X" se négocie à la baisse/hausse dans des volumes particulièrement nourris, supérieurs de 30% à la moyenne quotidienne de ses six derniers mois.

Valeurs reprises dans les principaux indices européens : Bel20, Bel Mid, Bel Small, Dax, Cac40, AEX, EuroStoxx50, Footsie (Temps réel?), Stoxx 600.

3. Dépassement de plus haut ou plus bas historique et/ou annuel des indices européens.

Bourse de "XXXX" : l'indice XX" dépasse son plafond/plus haut/sommet/pic annuel/depuis le début de l'année.

Bourse de "XXXX" : l'indice XX" enfonce/"descend sous" son plus bas annuel / plancher depuis le début de l'année.

Bourse de "XXXX" : l'indice XX" dépasse son /plus haut/plafond/sommet/pic historique.

Bourse de "XXXX" : l'indice XX" enfonce son plus bas historique

Base la plus lointaine possible.

Les principaux indices européens : Bel20, Bel Mid, Bel Small, Dax, Cac40, AEX, EuroStoxx50, Footsie (Temps réel?), Stoxx 600.

4. Dépassement de plus haut ou plus bas historique et/ou annuel dans Euronext Bruxelles.

Bourse de "XXXX" : L'action/Le titre XX" dépasse son plafond/plus haut/sommet/pic annuel/- depuis le début de l'année.

Bourse de "XXXX" : la valeur XX" enfonce/"descend sous" son plus bas annuel / plancher depuis le début de l'année.

Bourse de "XXXX" : la valeur XX" dépasse son /plus haut/plafond/sommet/pic historique.

Bourse de "XXXX" : la valeur XX" enfonce son plus bas historique

Base la plus lointaine possible. Toutes les valeurs au sein d'Euronext BXL.

Annexe 21. Interface de démonstration de Syllabs

Captures d'écran réalisées dans le cadre de la troisième vague d'évaluation, le 23/12/2018 ¹³².

The screenshot shows the Syllabs interface with three main sections: 'syllabs Turn data into text', 'Demo: Quotebot - FR', and 'Generated text'. The 'Quotebot - FR' section contains a dropdown menu with 'top flop Bruxelles' selected. Below the dropdown is a list of examples: 'Select Example', 'ouverture Europe (part1)', 'ouverture Europe (part2)', 'clôture Europe', 'mi-séance Europe', 'ouverture US', 'clôture Belgique', 'top flop Nasdaq 100', 'ouverture Belgique', 'clôture US', 'top flop Dow Jones', 'mi-séance US', 'mi-séance Belgique', 'aperçu à 18h', and 'ouverture Belgique (BEL20 seul)'. The 'Generated text' section contains the following text: 'L'indice Bel20 a clôturé sur une note stable ce jour: il a offert un léger accroissement de 0,04% à 3.506,48. Des 20 actions qui le composent, 10 sont en hausse et 10 sont en baisse.'

The screenshot shows the Syllabs interface with three main sections: 'syllabs Turn data into text', 'Demo: Quotebot - FR', and 'Generated text'. The 'Quotebot - FR' section contains a dropdown menu with 'ouverture Europe (part1)' selected. Below the dropdown is a text input field with 'ue' and a 'Generate' button. The 'Generated text' section contains the following text: 'À l'aube de la séance, les principaux marchés du continent européen se distinguaient par des indices globalement en hausse. Le CAC 40 parisien se adjugeait 1,03% à 5.180,23 points. L'indice AEX, quant à lui, augmentait de 0,86% à 522,82 points. À 3.537,12 points, le Bel20 se octroyait 0,99% ce dimanche.'

¹³² URL en accès protégé : http://demo.sweeny.syllabs.com/demo_quotebot/

Annexe 22. Échantillons de textes sources et de textes cibles

Evaluations 1			Evaluations 2			Evaluations 3		
Type	Trend	N	Type	Trend	N	Type	Trend	N
Ouverture WS	OD	2	Ouverture BXL	H	3	Ouverture Europe	H	3
Ouverture WS	B	1	Ouverture Europe	B	1	Ouverture Europe	OD	3
Ouverture Europe	B	1	Ouverture Europe	B	1	Ouverture WS	B	3
Ouverture Europe	OD	2	Ouverture Europe	OD	1	Ouverture Bruxelles	B	3
Ouverture Bruxelles	B	2	Ouverture Europe	S	3	Ouverture Bruxelles Bel 20	LB	6
Ouverture Bruxelles	H	5	Ouverture Europe	H	1	Mi-séance Europe	B	5
Clôture Wall Street	OD	3	Ouverture Wall Street	OD	2	Mi-séance Wall Street	B	3
Clôture Wall Street	B	3	Clôture Bruxelles	H	3	Mi-séance Bruxelles	B	6
Clôture Europe	OD	1	Clôture Europe	H	2	Clôture Europe	B	4
Clôture Bruxelles	S	1	Clôture Europe	OD	1	Clôture Belgique	H	3
			Clôture Wall Street	H	2	Clôture Wall Street	B	3
			Clôture Wall Street	B	2	Aperçu 18h Bruxelles	B	3
						Top-Flop Dow Jones		3
						Top-Flop Nasdaq 100		3
						Top-Flop Bruxelles		4
TOTAL		21			22			55

Trend (tendance)

OD = ordre dispersé

B = baisse

N = nombre de textes prélevés

S = stable/stagnation

H = hausse

LB = légère baisse

Échantillon de textes cibles

QBT001B. Les marchés du continent européen présentaient des taux globalement en baisse, avec un Footsie britannique qui restait totalement stable à 7.704,4 points. L'indice AEX cédait 0,71% à 558,04 points, le CAC 40 parisien rétrogradait de 0,53% à 5.419,71 points tandis que le DAX allemand perdait 1,11% à 12.668,77 points. De son côté, l'indice action Bel20 de Bruxelles a ouvert avec une évolution de 20,63% à 3 771,34 points.

QTB002B. Wall Street pointait en ordre dispersé à l'ouverture de la séance ce jour, avec un S&P 500 qui était en légère hausse de 0,09% à 2.774,84 points. Le Dow Jones cédait du terrain, à hauteur de 0,13% à 25.209,29 points.

QTB005B. Ouverture en hausse ce mercredi avec un indice Bel20 qui affichait une progression de 0,17% à 3.838,65 points. Parmi les meilleures progressions on retrouvait des valeurs comme bpost (0,53%), Solvay (0,37%) et Aperam (0,21%). De l'autre côté du spectre, il était question des groupes comme arGEN-X (-0,72%), Engie (0,3%) et Colruyt (0,32%). Hors Bel20, les variations les plus significatives étaient à mettre sur le compte de valeurs telles que Celyad (8,87%), Dexia (8,89%) mais aussi ThromboGenics (0,16%).

QTB006B. Bpost (0,53%), Solvay (0,37%) ou encore Aperam (0,21%) illustrent bien la hausse de l'indice Bel20, qui augmentait de 0,17% à 3.838,65 points. Hors Bel20, les variations les plus significatives étaient à mettre sur le compte de valeurs telles que Celyad (8,87%), Dexia (8,89%) ou encore ThromboGenics (0,16%).

QTB011B. L'indécision était de mise pour les marchés européens à l'aube de la séance, avec un Footsie britannique qui restait inchangé à 7.655,79 points. L'indice AEX augmentait de 0,35% à 572,46 points, le CAC 40, quant à lui, montait de 0,39% à 5.399,19 points et l'indice DAX, quant à lui, cédait 0,31% à 12.510,23 points. L'indice action Bel20 ouvert quant à lui la séance de ce mercredi avec 0,17% à 3 838,65 points.

QTB015B. Les marchés New Yorkais ont clôturé en ordre dispersé, avec un indice Dow Jones Industrial Average qui progressait de 0,57% à 25.187,22 points. L'indice élargi S&P 500 était en légère baisse de 0,18% à 2.806,98 points et le Nasdaq composite montait de 0,43% à 7.403,292 points.

QTB019B. La bourse de Wall Street a clôturé globalement en baisse : avec un indice Dow Jones Industrial Average qui rétrogradait de 0,26% à 25.459,49 points, le S&P 500 cédait du terrain, à hauteur de 0,3% à 2.837,44 points tandis que le Nasdaq composite piquait du nez de 1,6% à 7.726,7 points.

QV4_01. À l'aube de la séance, les principaux marchés du continent européen se distinguaient par des indices globalement en hausse, le CAC 40 parisien se adjugeait 1,03% à 5.160,23 points. L'indice AEX, quant à lui, augmentait de 0,86% à 522,82 points. À 3.537,12 points, le Bel20 se octroyait 0,99% ce dimanche.

QV4_15. Trois heures après l'ouverture, l'indice Bel20 se effondrait de 1,76% à 3.512,55 points. Sur les marchés européens, le DAX allemand, quant à lui, se délestait de 1,39% à 11.549,89 points, l'indice AEX perdait 1,62% aujourd'hui, à 519,45 points tandis que le Footsie britannique rétrogradait de 1,75% à 7.020,67 points. À 3.222,01 points, l'indice paneuropéen Euro Stoxx 50 reculait de 1,37% ce jour.

QV4_18. Wall Street présentait des indices majoritairement en baisse à l'ouverture de la séance, l'indice Dow Jones Industrial Average cédait du terrain, à hauteur de 0,35% à 26.705,25 points. L'indice élargi S&P 500, pour sa part, était en légère baisse de 0,04% à 2.936,76 points et l'indice Nasdaq perdait 0,39% à 7.477,45 points.

QV4_19. L'indice Bel20 a clôturé en hausse ce dimanche : ce dernier a progressé de 0,1% à 3.750,81.

QV4_22. Sur les 100 actions qui composent le Nasdaq 100, 58 sont en hausse et 42 sont en baisse. En tête, on retrouvait Seagate Technology PLC (2,45%), Facebook (2,23%) mais aussi KLA-Tencor Corp (1,78%). Hasbro (-2,01%), Microchip Technology (-2,25%) et Tesla (-2,75%) fermaient la marche.

QV4_23. Sur les 100 actions qui composent l'indice Nasdaq, 58 sont en hausse et 42 sont en baisse. Parmi les meilleures progressions on retrouvait des valeurs comme Seagate Technology PLC (2,45%), Facebook (2,23%) et KLA-Tencor Corp (1,78%). Hasbro (-2,01%), Microchip Technology (-2,25%) ou encore Tesla (-2,75%) se situaient, quant à eux, à l'autre bout du spectre.

QV4_32. Des 30 actions qui composent le Dow Jones, 22 sont en hausse et 8 sont en baisse. L'indice était donc emmené par Walgreens Boots Alliance (2,38%), The Procter & Gamble Company (1,82%) et The Walt Disney Company (1,6%). Microsoft Corp (-0,62%), Nike (-0,62%) ou encore Visa Inc (-0,91%) fermaient la marche.

QV4_44. À Bruxelles, l'indice Bel20 s'est délesté de 0,39% à 3.508,34 points. L'indice paneuropéen Euro Stoxx 50, quant à lui, qui a augmenté de 0,29% à 3.223,77 points.

QV4_47. En baisse à l'ouverture de la séance, l'indice Bel20 affichait un repli de 0,11% à 3.737,15 points. Les reculs les plus importants étaient à mettre sur le compte des actions Ontex (-0,38%), Engie (-0,44%) ou encore Aperam (-0,51%). Colruyt (3,35%), Telenet Group Holding (1,14%) et bpost (0,35%), quant à eux, limitaient les pertes.

Échantillon de textes sources et de textes cibles

TEST 1

Source

Les marchés européens ont entamé la séance en baisse ce lundi. Dès les premières minutes de transactions, le Cac 40 parisien se repliait de 1,88%, le Dax allemand de 2,77% et le Footsie britannique de 3,78%. À la Bourse de Bruxelles, le Bel 20 cédait 1,8%.

Cible

Les marchés du continent européen présentaient des taux globalement en baisse, avec un Footsie britannique qui restait totalement stable à 7.704,4 points. L'indice AEX cédait 0,71% à 558,04 points, le CAC 40 parisien rétrogradait de 0,53% à 5.419,71 points tandis que le DAX allemand perdait 1,11% à 12.668,77 points. De son côté, l'indice action Bel20 de Bruxelles a ouvert avec une évolution de 20,63% à 3 771,34 points.

TEST 4

Source

La Bourse de Bruxelles était orientée en baisse dans les premiers échanges, avec un indice Bel20 qui perdait 2,8% à 7615,89 points. Au sein de l'indicateur, 8 valeurs étaient en baisse, 2 inchangée(s) et 8 en hausse. Les reculs les plus importants étaient à mettre sur le compte des actions Valeurs 1 (+2,88%), Valeurs 2 (+2,78%) et Valeurs 3 (+6,56%). En queue de peloton, on retrouvait Valeurs 4 (-1%), Valeurs 5 (-12,8%) et Valeurs 6 (-2,76%). Hors Bel20, on retiendra les pertes encaissées par Valeurs 10 (2,98%), Valeurs 11 (5,88%) et Valeurs 12 (2,77%) tandis Valeurs 7 (12,7%), Valeurs 8 (10,88%) et de Valeurs 9 (1,8%) progressaient.

Cible

Ageas (-0,32%), Ackermans & van Haaren (-0,4%) et GBL (-0,43%) font décliner l'indice Bel20, qui perdait 0,54% à 3.771,34 points. Hors Bel20, on retiendra les meilleurs résultats de différents groupes comme Hamon Groep (4,12%), IBA (2,13%) ou encore 2Valorise (2,04%). À l'inverse, les pertes encaissées concernent des valeurs comme Picanol (0,22%), ABO-Group (0%) et Accentis (0%).

TEST 19

Source

Les marchés européens pointaient en ordre dispersé à l'entame de la séance ce lundi/mardi/mercredi/jeudi/vendredi : le Cac 40 parisien cédait 1,1%, tandis que le Dax allemand reculait de 3,32% et le Footsie britannique grappillait 2,22%. À la Bourse de Bruxelles, le Bel 20 grignotait pour sa part 1

Cible

L'indécision était de mise pour les marchés européens à l'aube de la séance, avec un Footsie britannique qui restait inchangé à 7.655,79 points. L'indice AEX augmentait de 0,35% à 572,46 points, le CAC 40, quant à lui, montait de 0,39% à 5.399,19 points et l'indice DAX, quant à lui, cédait 0,31% à 12.510,23 points. L'indice action Bel20 ouvert quant à lui la séance de ce mercredi avec 0,17% à 3 838,65 points.

TEST 20

Source

L'indécision est de mise en Europe à l'aube de la séance. Dans les premiers échanges, Paris gagnait 1,22%, mais Francfort et Londres corrigeaient de 1%. Amsterdam fondait de 2,2% et son équivalent à Bruxelles de 3,3%.

Cible

L'indécision était de mise pour les marchés européens à l'aube de la séance, avec un Footsie britannique qui restait inchangé à 7.655,79 points. L'indice AEX augmentait de 0,35% à 572,46 points, le CAC 40, quant à lui, montait de 0,39% à 5.399,19 points et l'indice DAX, quant à lui, cédait 0,31% à 12.510,23 points. L'indice action Bel20 ouvert quant à lui la séance de ce mercredi avec 0,17% à 3 838,65 points.

TEST 9b

Source

Les marchés européens évoluaient sur une note stable à l'entame de la séance ce lundi : le Cac 40 parisien progressait de 1,2%, tandis que le Dax allemand reculait de 1,2% et le Footsie britannique de 1,2%. À la Bourse de Bruxelles, le Bel 20 grimpait de 1,2%.

Cible

À l'ouverture, l'hésitation dominait pour les principaux marchés d'Europe, avec un indice Footsie qui progressait de 0,23% à 7.655,79 points. Du côté de l'indice AEX, ce dernier augmentait de 0,12% à 572,46 points, le CAC 40 se renforçait de 0,35% à 5.399,19 points, l'indice DAX parfaitement stable, n'enregistrant ni hausse, ni baisse tandis que l'indice BEL20, quant à lui, montait de 0,22% à 3.838,65 points. De son côté, l'indice action Bel20 des principaux marchés du continent européen ont ouvert avec une évolution de 0,22% à 3.838,65 points.

TEST 10b

Source

L'attentisme domine à l'ouverture, les indices actions évoluant dans des marges étroites autour de leur point d'équilibre. Dans les premiers échanges, Paris gagnait 1%, mais Francfort et Londres se délestaient de 1,2%. Amsterdam pliait de 1,8% et son équivalent à Bruxelles de 1%.

Cible

À l'ouverture, l'hésitation dominait pour les principaux marchés d'Europe, avec un indice Footsie qui progressait de 0,23% à 7.655,79 points. Du côté de l'indice AEX, ce dernier augmentait de 0,12% à 572,46 points, le CAC 40 se renforçait de 0,35% à 5.399,19 points, l'indice DAX parfaitement stable, n'enregistrant ni hausse, ni baisse tandis que l'indice BEL20, quant à lui, montait de 0,22% à 3.838,65 points. De son côté, l'indice action Bel20 des principaux marchés du continent européen ont ouvert avec une évolution de 0,22% à 3.838,65 points.

TEST 18c

Source

Wall Street marque le pas dans les premiers échanges. L'indice Dow Jones perd 0,35% à 26.705,25 points, l'indice S&P 500 0,04% à 2.936,76 points et l'indice composite du Nasdaq 0,39% à 7.477,45 points.

Cible

Wall Street présentait des indices majoritairement en baisse à l'ouverture de la séance, l'indice Dow Jones Industrial Average cédait du terrain, à hauteur de 0,35% à 26.705,25 points. L'indice élargi S&P 500, pour sa part, était en légère baisse de 0,04% à 2.936,76 points et l'indice Nasdaq perdait 0,39% à 7.477,45 points.

TEST 28c

Source

Wall Street a clôturé en hausse ce dimanche : le Dow Jones a progressé de 0,14%, tandis que le S&P 500 reculé de 2,15% et le Nasdaq de 0,03%.

Cible

À la clôture de la séance, l'hésitation dominait à la Bourse de Wall Street, l'indice Dow Jones a perdu 0,14% à 25.762,54 points. Le S&P 500 a gagné 2,15% à 2.809,92 points tandis que l'indice composite du Nasdaq était en légère hausse de 0,03% ce dimanche, à 7.648,02 points.

TEST 44c

Source

La Bourse de Bruxelles a fini en baisse. L'indice Bel 20 a perdu 0,39% à 3.508,34 points.

Cible

À Bruxelles, l'indice Bel20 s'est délesté de 0,39% à 3.508,34 points. L'indice paneuropéen Euro Stoxx 50, quant à lui, qui a augmenté de 0,29% à 3.223,77 points.

TEST 52c

Source

La Bourse de Bruxelles a clôturé sur une note stable ce lundi : son indice de référence, le Bel 20, a progressé de 0,44% à 3.506,48 points.

Cible

L'indice Bel20 a clôturé sur une note stable ce dimanche : il a connu un léger accroissement de 0,04% à 3.506,48. Des 20 actions qui le composent, 10 sont en hausse et 10 sont en baisse.

Annexe 23. Liste des templates de "Quotebot" et indications sur la récupération des données

zone	heure de la période	début des calls	fin des calls	indice(s) concerné(s)	données à récupérer et à calculer	valeurs obligatoires	champs utilisés	sélection à faire	tableau
BXL	ouverture	9h	09h01	09h05	BeI20	BeI20	- openPrice (du jour) - ChangePercentage - le timestamp est utilisé pour détecter la validité (mise à jour) de la donnée	TOP et FLOPS (3 dans texte, 5 dans tableau) SLOW : valeurs déjà triées, mais avec des bizarries FAST : valeurs non triées --> c'est Syllabs qui trie	- 1 tableau TOP/FLOP BEL20
BXL	ouverture	9h	09h01	09h05	BeI20 & ses composants (issues) à calculer : - Top 3 BeI20 - Flop 3 BeI20 --> valeurs négatives !!!	BeI20			
BXL	ouverture	9h	09h15	09h30	BeI20 & ses composants (issues) - HorsBeI20 : BelMid & BelSmall et leurs composants à calculer : - Top 3 et Flop 3 BeI20 - Top 3 et Flop 3 BelMid et BelSmall	BeI20, BelMid, BelSmall		TOP et FLOPS (3 dans texte, 5 dans tableau) SLOW : valeurs déjà triées, mais avec des bizarries FAST : valeurs non triées --> c'est Syllabs qui trie --> trie plus complexe, car 2 indices à fusionner (BelMid et BelSmall)	- 1 tableau TOP/FLOP BEL20 - 1 tableau TOP/FLOP HorsBEL20
EU	ouverture Europe1	9h	09h01	09h05	3 indices EU - BeI20 - CAC40 - AEX --> uniquement les indices	Les 3 indices sont obligatoires			néant
EU	ouverture Europe2	9h	09h20	09h30	5 indices EU - BeI20 - CAC40 - AEX - Footsie - Dax --> uniquement les indices	- les 5 indices sont obligatoires - le template génère le texte si BeI20 et 1 autre indice, mais on va rester sur les 5 valeurs pour des questions de simplicité côté gestion de la data		néant	néant
NY	ouverture US	15h30	15h31	15h40	2 indices WS - Dow Jones - Nasdaq --> uniquement les indices	- les 2 indices		néant	néant
EU	clôture Europe1	17h40	17h41	17h45	3 indices EU - BeI20 - CAC40 - AEX --> uniquement les indices - Or - OLO Belgique - Taux d'échange dollar/euro	- les 3 indices sont obligatoires	openPrice dayChangePercentage	néant	- 1tableau avec les indices EU - classement par ordre décroissant - tableau chiffres clés prévoir champs vides pour bund pétrole brent

période	zone géo	article	heure de la période	début des calls	fin des calls	indice(s) concerné(s)	données à récupérer et à calculer	valeurs obligatoires	champs utilisés	sélection à faire	tableau
18h		aperçu 18h EU (tableau seulement) --> HIT PARADE TOP FLOP Eurostoxx				- Eurostoxx 50	- Eurostoxx 50 avec composants - top 5 et flop 5 Eurostoxx 50				
hit-parade hebdo (après clôture) --> que des tableaux !	BXL	hit-parade hebdo BXL BeI20 composite	17h50	17h50		BeI20	BeI20 & ses composants		LastPrice & LastPrice J-7 (attention jours fériés)		- 1 tableau avec tous les composants par ordre décroissant
hit-parade hebdo BXL (all BeI20 Top/Flop)	BXL	hit-parade hebdo BXL (all BeI20 Top/Flop)	17h50	17h50		BeI20 & ses composants (issues) à calculer : - Top 5 BeI20 - Flop 5 BeI20	BeI20 & ses composants à calculer : - Top 5 BeI20 - Flop 5 BeI20				- 1 tableau avec les Top 5 et les Flop 5 du BeI20
hit-parade hebdo BXL (all i)	BXL	hit-parade hebdo BXL (all i)	17h50	17h50		- BeI20 & ses composants (issues) à calculer : - Top 3 BeI20 - Flop 3 BeI20 --> valeurs négatives !!!	- BeI20 & ses composants (issues) à calculer : - Top 3 BeI20 - Flop 3 BeI20 --> valeurs négatives !!!				- 1 tableau avec les Top 5 et les Flop 5 de la Bourse de BXL (fusion BeI20, BelMid et BelSmall)
hit-parade hebdo EU	EU	hit-parade hebdo EU	17h50	17h50		- BeI20 - CAC40 - AEX - Footsie - Dax - Eurostoxx 50 --> uniquement les indices	- BeI20 - CAC40 - AEX - Footsie - Dax - Eurostoxx 50 --> uniquement les indices				- 1 tableau avec les indices
hit-parade hebdo monde	EU & NY	hit-parade hebdo monde	17h50	17h50		- BeI20 - CAC40 - AEX - Footsie - Dax - Eurostoxx 50 - Nasdaq --> uniquement les indices	- BeI20 - CAC40 - AEX - Footsie - Dax - Eurostoxx 50 - Nasdaq --> uniquement les indices				- 1 tableau avec les indices
hit parade mensuel									LastPrice & LastPrice j-30/31 (attention jours fériés)		
hit parade trimestriel (1 fois par trimestre)				toutes les minutes					LastPrice & LastPrice -T(attention jours fériés)		
hit parade annuel (1 fois en fin d'année)									LastPrice & LastPrice -an (attention jours fériés)		
hit parade year-to-date									LastPrice & LastPrice du 29/12/2017		