

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'ex- périence "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 in- telligence" 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