STANDARDISED MODELLING AND INTERCHANGE OF LEXICAL DATA IN SPECIALISED LANGUAGE[1]

Marc Van Campenhoudt
Termisti research group
Tradital research centre
Université libre de Bruxelles

ABSTRACT

In order for their content to be exploited to its fullest, true electronic dictionaries must be built on a very strictly defined structure with highly specific data categories. Specialised language dictionaries, especially multilingual ones, strongly benefit from a structure where data are identified with a high degree of precision, in a monosemic 'hub and spoke' model. A representation that meets the ISO 16642 *Terminological Markup Framework* (TMF) and ISO 30042 *TermBase eXchange* (TBX) standards is ideally suited to implement such a model. However, using an XML markup language that meets these standards defined by the International Standards Organization does not guarantee the data's quality or interoperability. In addition, the description and maintenance of standardised categories of terminological data (ISO 12620) still pose a number of difficult challenges.

Keywords: computational lexicography, terminology, standardisation, languages for special purposes

[1]. This contribution is a revised version of an article published in French under the title 'Modélisation et échange normalisé des données lexicales en langue spécialisée' in journal *Revue française de linguistique appliquée* (XXII-1, June 2017), devoted to the topic '*Diversité des ressources lexicales : de leur élaboration à leur utilisation*' ['Diversity of lexical resources: from production to utilization'].

INTRODUCTION

The idea of identifying data categories in dictionaries is not recent. It appears in the earliest versions of the *Text Encoding Initiative* (TEI[2]), released around the same time as the emergence of SGML (*Standard Generalized Markup Language*) at the end of the 1980s, replaced in the 1990s by XML (*Extensible Markup Language*), which is especially simple to learn. The goal is to share works in a standard future-proof format that is not reliant on a specific piece of software, while also facilitating the exploitation of content from all available lexical resources (Romary, Salmon-Alt and Francopoulo 2004, 22). By changing a simple word processing file into a structured document, its content is made much more valuable, with markup making it possible to answer highly specific questions such as 'what French verbs were borrowed from English between 1801 and 1900?'. This requires a methodical collation of the information contained in the work, which no reasonable human mind could be expected to achieve by reading pages alone.

Up until version P4 (Burnard and Sperberg-McQueen, 2004), the TEI devoted a chapter to dictionaries and one to terminological databases. However, in the introduction to the latter, it notes its obsolescence: *'since its first publication, this chapter has been rendered obsolete in several respects, chiefly as a result of the publication of ISO[3] 12200, and a variant of it (TBX)'* (Burnard and Sperberg-McQueen 2004, 323). Since the beginning of the TEI, this chapter has evolved from TIF (*Terminology Interchange Format*) to MARTIF (*Machine-Readable Terminology Interchange Format*) (ISO 12200 1999). Later becoming TBX (*TermBase eXchange*, ISO 30042 2008), it still carries in its markup traces that are characteristic of the TEI. However, TBX has strongly distanced itself from the TEI as its promoters gradually understood that a surface description intended for lexicographies—often polysemic

[2]. www.tei-c.org: 'The Text Encoding Initiative (TEI) Guidelines are addressed to anyone who wants to interchange information stored in an electronic form. They emphasize the interchange of textual information, but other forms of information such as images and sound are also addressed. The Guidelines are equally applicable in the creation of new resources and in the interchange of existing ones.' (Burnard & Sperberg-McQueen 2012, 1) This reference can be consulted for an introduction to the principles of markup.

[3]. The International Standards Organization (ISO) gathers national standards organisations such as the Association française de normalisation (AFNOR). 'Through its members, it brings together experts to share knowledge and develop voluntary, consensus-based, market relevant International Standards that support innovation and provide solutions to global challenges.' (www.iso.org/about-us.html). To this end, it relies on the opinion of experts from a technical committee. Technical Committee 37 (TC 37) is responsible for standards related to language resources.

ones—made it difficult to manage equivalencies. Later we will offer a reminder of why a simple surface description is not satisfactory, and even venture a critical perspective on the choices made. Besides, the diversity of data categories used in terminology soon became difficult to manage within the boundaries of the TEI if one wished to describe the huge variety of dictionaries available beyond major references dedicated to general language[4]. This again calls for a critical analysis, as the lists of data categories and the methods used to build these lists can be problematic. Lastly, based on our practical experience, converting a specialised dictionary—whether printed or electronic—into a standardised interchange format often results in identifying many inconsistencies both in its microstructure and macrostructure.

The identification of data categories poses other problems, as definitions of the categories must be agreed upon, which necessarily brings up epistemological questions. In the context of a standardisation initiative, a compromise must be reached between a minimal level of stability and a required level of enrichment. The transition to a model involving a well-defined structure and data categories can challenge a number of strongly held beliefs, such as the idea of a fundamental opposition between initial macrostructural choices (homonymy vs. polysemy), which is now rendered moot by the possibilities of digital modelling. As a result, the very techniques used to compile dictionaries are called into question, including those used to create their microstructure. The challenge seems to lie in compiling dictionaries in this new paradigm.

The point of view that will be promoted here is based on our personal experience as an expert in the digitisation of specialised dictionaries who has not been involved in international standardisation processes since his participation in the IST-SALT[5] project and who works with what information is available. This means that some critiques might be answered in current developments, especially with the new ISO/DIS 12620 (2018) and ISO/DIS 30042 (2018) standards.

[4]. These observations do not negate the wish for a return of a chapter in the TEI dedicated to terminology (Romary 2014). This is all the more true that the ISO 24613 (2008) standard, *Lexical Mark-up Framework* (LMF), devoted to lexicographies is largely compatible with the chapter in the TEI devoted to dictionaries (Romary 2015).

[5]. www.ttt.org/salt/description.html, see 2.1.

1 SPECIALISED LEXICOGRAPHY VS. TERMINOGRAPHY

The world of specialised dictionaries is highly diverse, on the same level as—and perhaps even more than—general language dictionaries. A traditional distinction has long been made, based on various criteria, between works that follow the methods of terminology and works that are specialised lexicographies[6].

The lexicography of general language is often thought to be focused on single words, typically rejecting phrases. It is viewed as a descriptive and semasiological discipline that results in a polysemic approach, with synonyms presented in alphabetical order. For bilingual dictionaries, this necessarily implies creating two volumes; all meanings listed under a given entry in one language do not match the sum of meanings of all their equivalents. This can easily be checked by comparing the meanings of *banane* in French and *banana* in English—or *timbre* and *stamp*—in the *Grand Robert & Collins* (2017). At the same time, it should also be mentioned that certain monolingual lexicographies in general language adopt a monosemic—and therefore homonymic—approach, without however turning to onomasiology. These are rare, date back to the early 1960s for French, and have not been very popular. A typical example is the excellent *Lexis* (Dubois 2014), which is still in print, and in which the monosemic approach is still not quite pushed to its limits.

The terminographical method, which would be 'conceptual' and prescriptive, would favour an onomasiological, monosemic, and homonymic approach that groups synonyms together not in alphabetical order, but based on meaning. This approach 'naturally' implies recognising phrases as terms (for instance, *quarantine buoy, azimuthal equidistant projection, electronic chart display and information system,* etc.).

The opposition between these two worlds is largely disproven by an observation of what specialised dictionaries and terminological databases actually are. In practice, many specialised

[6]. For finer differences, see e.g. Bergenholtz and Kaufmann (1997), Humbley (1997), Béjoint (2007), and Fuertes-Olivera and Tarp (2014).

monolingual lexicographies are encyclopaedic works whose purpose—and therefore whose content—changes depending on the model reader. Polysemy can very well belong in such works, as specialised terms are also subject to a diversification of meaning, especially in diachrony (Dury and Picton 2010) and as our own understanding evolves (Temmerman 2000, 153-154). Terms that are phrases naturally appear as distinct entries, possibly camouflaged as simple lexias to follow the practice of 'proper' dictionaries[7]: for instance, *projection: azimuthal equidistant* (OHI 1994-1998) or *cuirasse (~ ferrugineuse)* (Lozet & Mathieu 1990). In addition, electronic dictionaries used in linguistic engineering—first and foremost lexicon-grammars based on object classes (Le Pesant and Mathieu-Colas 1998, 20-22)—naturally favour monosemy.

As we have previously shown (Van Campenhoudt 2000 and Van Campenhoudt 2001a), the transition from a lexicographical approach to a terminographical one is largely the result of a need to create multilingual dictionaries.. A bilingual dictionary requires only two volumes. For more languages, if all permutations of language pairs are to be included, the number of volumes required becomes unmanageable, as it is defined by the formula L^2-L where L is the number of languages (3 languages: 6 volumes; 4 languages: 12 volumes; etc.). Obviously, the sum of the three reference bilingual dictionaries published by Robert, French-English, French-Italian, and French-Dutch, will never cover English-Italian, English-Dutch, or Dutch-Italian.

Regarding language permutations, it should also be noted that current bilingual dictionaries for general language are often encoding dictionaries (used to translate towards a foreign language). Terminographies, on the other hand, are largely used in translation, i.e. for decoding, but they do claim to guarantee that all language pairs are available in both directions. Once again, the facts disprove the theory: outside of exceptional cases like *The Machine Tool* dictionary by Eugen Wüster (1968) or, long before it, *From Keel to Truck* by Heinrich Paasch (1894), most multilingual terminographies are 'false multilinguals' (Quemada 1967) in which one language takes precedence over others that are provided as an addition for those wishing to write in a second language. This issue, of course, carries over to online

[7]. Without dwelling on this question, which will be covered in section 4.2, we feel that the position traditionally adopted by lexicographers who favour graphical words is increasingly challenged by the requirements of interfacing between texts and electronic dictionaries.

versions of these works, and even to large terminological databases where a language like English can also serve as a pivot language, often implicitly.

Many specialised lexicographies can purportedly be used for translation. However, upon closer scrutiny, it seems that one language often prevails, either as a source language or a pivot language, and there is little information to be found in the other languages. These multilingual specialised dictionaries are most often encoding dictionaries, which allows users to translate only in a single direction[8]. They offer definitions in the main language that inform a model reader—and native speaker of this language—, and equivalents in the second language that help the reader express themself in a foreign language. It is, therefore, paradoxical to see translators—who are professional decoders—use these dictionaries to such an extent. It is true that online versions of these works make it possible to search directly in the second language, whereas the original printed version is clearly an encoding dictionary (decoding requires using the index in the final pages). Terminological databases used as multilingual translation tools sometimes remain ambiguous in that regard, starting with those that were created for as part of linguistic policy (Cissé *et al.* 2009).

A number of original techniques were tested throughout history in order to solve the problem of switching between source and target language, as truly multilingual specialised dictionaries are in much higher demand than general language dictionaries, owing to the needs of commercial and industrial exchanges. Terminology as a scientific discipline was undoubtedly first theorised by Eugen Wüster (1979), who was also a pioneer in the development of linguistic standards for the ISO. We should not forget, however, that multilingual dictionaries—often specialised—appeared before the first monolingual general language dictionaries (Quemada 1967, 37-73 and 567 ff.), and have probably been published in greater numbers. This enabled Wüster to draw from terminographical practices that were already quite advanced in the 19th century in the maritime domain, and that reached an apex with the second edition of Heinrich Paasch's *From Keel to Truck* (Van Campenhoudt 2003).

[8]. There are, of course, notable exceptions, such as the *Dictionnaire de la comptabilité et de la gestion financière* (Ménard, Arsenault and Joly 1994) and the *Lexique français-anglais, anglais-français de la bourse et des marchés financiers* (Villeneuve 2001).

Still, it seems to us that most terminographies are lacking—not to say paltry—in terms of content, compared to many specialised lexicographies. They typically feature much fewer categories, and many fields are hardly used at all. The more a database is geared towards translation needs, the less semantic, lexical, or grammatical content it seems to have (Van Campenhoudt 2000), even though such data would promote quality translations. Terminographical dictionaries that do not offer this type of data are often presented as a table where the leftmost column contains entries in the source language. The most striking fact may be that many 'terminological' glossaries are created using a spreadsheet, with one column per language and each cell containing one or several equivalent terms. These flat files do not structure information, and do not allow for advanced data processing.

2 DATA CATEGORISATION

The use of 'records' is common between lexicologists/lexicographers and terminologists/terminographers. The former traditionally use them in general language, to observe usage and store attested forms. The latter have been using records for a long time, as evidenced by the oldest ISO standards (such as ISO /R 919 1969), to present their dictionaries. Before the computer age, such dictionaries were not published, but they could be accessed as a series of index cards in organisations or translation firms.

2.1 ISO categories

Drawing up an inventory of data categories has appeared as an increasingly urgent need as organisations producing terminology for translation have started using computer-aided translation (CAT) tools. The first such large-scale inventory is the ISO 12620 (1999). standard. Written in English and French, it included a large collection of data presented in a logical structure, with each category systematically described.

This standard was limited by its rigid nature: any changes made to the inventory had to go through a very slow updating process that had little chance of ever coming to an end: changes had to be submitted to ISO experts, who then had to reach a consensus. The new version of the standard (ISO 12620 2009) was a radical change, as a list of categories was no longer hard-coded in the standard itself. In line with the results of the EU's IST-SALT project, whose purpose was to promote sharing data categories and enable more precise descriptions (Budin and Melby 2000), the next 12620 standard had included the creation of an online server, *ISOcat*[9]. Users could negotiate, describe, and share all data categories used in developing language resources, far beyond the scope of terminography (Wright *et al.* 2013). The server was online until 2015, at which point its content stopped receiving updates[10]. Since then, the data were transferred on a *TermWeb* server and later converted to the TBX format[11]. At the time of writing, the former *ISOcat* website redirects to the *Data Category Repository* (DCR)[12], maintained by '*LTAC/TerminOrgs, a liaison organization to ISO Technical Committee 37*' (ISO/DIS 12620 2018, 4). In addition to its strong relationship with TBX[13], the website points to a '*DatCatInfo database*' that clearly contains *ISOcat* data converted to *TermWeb* with no substantial changes in content. *ISOcat*'s *Persistent IDentifiers* (*PID)* remain valid and shall be used throughout this text[14]. In addition, the data available seem to broadly correspond to the inventory made in 1999 (table 1), which TBX uses to a great extent. This means that the ISO's former categories (from 1999), which were historically used by TBX and have served as a basis for *ISOcat*, appear to have now become a reference DCR. It should be noted that the project for an ISO/DIS 12620 (2018) standard provides for the coexistence of multiple DCRs. It also specifies the description requirements that this implies, as well as rules on update and maintenance that allow for a dialogue with users. This dialogue was not successful on the *ISOcat* website, where an increasingly unmanageable mass of unvalidated data was stored (see Warburton 2015).

[9]. www.isocat.org.

[10]. The reasons for this decision are provided on the website's home page.

[11]. demo.termweb.se/termweb/app.

[12]. www.datcatinfo.net.

[13]. The *TBX Master Data Category Selection* (www.tbxinfo.net/tbx-datcats) page seems to largely corroborate this observation.

[14]. The URLs for *ISOcat*'s PIDs dynamically redirect to *TermWeb*.

ISO 12620 EN (1999)	ISOCAT
A.10.2.2.1 created by / originator	DC-162 created by
A.10.14 concept identifier	DC-139 concept identifier
A.5.3 context	DC-149 context
A.10.2.1.2 input date	DC-274 input date
A.5.1 definition	DC-168 definition
A.4 subject field	DC-489 subject field
A.1 term	DC-508 term
A.2.2 grammar	DC-250 grammaticalInfo
A.10.7.1 language identifier	DC-2482 languageID

Table 1. Comparison of selected old and new data categories

2.2 Flaws in the content published on *ISOcat*

Although we had high hopes for *ISOcat*'s novel approach, it eventually became clear that our reservations regarding the old standard—which had led us to taking part in the SALT project—were still relevant to the current data categories. In our own experience, a number of flaws remain:

- A bias towards European languages, with grammar that is essentially inherited from Latin. As a result, for the research project entitled *Conception d'un dictionnaire électronique unilingue wolof et bilingue wolof-français*[15] (2007-2008) ['Development of a electronic dictionary for Wolof (monolingual) and Wolof-French (bilingual)'], there was no category available for noun class, which is required to describe languages that use this concept, many of which exist in sub-Saharan Africa[16]. As a result, major transnational languages such as Wolof and Swahili could not be described. Other questions could be raised regarding the baggage of certain *ISOcat* contributors, or—*nolens volens*—their cultural imperialism, upon reading certain definitions that claim to be of universal value: *(feminine) 'Definition: Of, relating to, or constituting the gender that ordinarily includes most words*

---

[15]. Project funded by the *Lexicologie, terminologie, traduction* network, See flsh-dico-wolof.ucad.sn.

[16]. Generally speaking, few developing countries take part in the work sessions of ISO Technical Committee 37, which explains the low consideration given to features of their languages.

or grammatical forms referring to females.[17]'

- A lack of data categories suited to innovative projects[18]. For instance, there is no data category that refers to the level of readability of a term in a database intended to offer simple equivalents of specialised terms used in hospitals (Babeliris project[19]). Similarly, we have recently noted the absence of an *ad hoc* category for the feminine form of positions in institutions that are concerned with gender issues.

- A categorisation of data that can be problematic. For instance, ISO category 'term type' (DC-2677) oddly includes data based on semantics (the term is a synonym), morphology (the term is an initialism, an acronym, a transliteration, etc.), or even statistics (the term is a collocation). Only in certain cases is it possible to establish a morphological relationship between two abbreviated forms.

- An inadequate ability to account for semantic relationships. The representation of hyponymic relationships seems mishandled (*related concept, related concept broader, related concept narrower, broader concept generic*, etc.), there is no distinction between the different types of meronymic relationship, and other relationships (cause, agent, chronological sequence, etc.) are bound to a single category: *sequentially related concept*. We had already drawn attention to this in our contribution to the SALT project (Van Campenhoudt 2001b). It would certainly be desirable for the categorisation of semantic relationships to take into account the new opportunities offered by descriptive languages that are specific to the field of ontologies (Roche 2013).

- An underlying doctrinal orientation, influenced by Eugen Wüster's legacy. While ISO Technical Committee 37 (see note 3) has considerably evolved over time, traces of its originating figure remain very present in certain standards and various category names and definitions that do not take into

[17]. www.isocat.org/rest/dc/247. For the sake of honesty, we must point out that the records in their current fiexed state include duplicate entries that were obviously created in order to call attention to issues of grammatical gender (see e.g. www.isocat.org/rest/dc/1558).

[18]. We will not mention here the proliferation of new categories that have appeared in recent major terminographical projects involving Igor Mel'čuk's lexical functions, *Framenet*'s valences, or other innovative models (Faber & L'Homme 2014).

[19]. The research project dealt with communication situations in which the medical and administrative staff in Brussels hospitals is in contact with people whose native language is neither French nor Dutch. Among its ambitions were the development of a terminological database that would take this issue into account (Van de Velde 2014).

account new approaches developed in the past three decades[20]. For instance, the 'definition' category (DC-168 and DC-1972) is still defined as '*[A] representation of a concept by a descriptive statement which serves to differentiate it from related concepts*.' This seems inevitable when other ISO standards, including the vocabulary of terminology (ISO 1087-1 2000), are not amended in *ISOcat*. In addition, many categories from the old 1999 standard, ISO 12620, have been copied as-is into *ISOcat*, with no changes made to their definition. The categorisation of types of definition (based on DC-169) and of types of contexts (based on DC-150) is a good example of this. Using a deprecated standard as a basis for the reintroduction of controversial categories without changes seems almost harmful, especially since the intention of *ISOcat*'s proponents was to no longer keep things static (see our next point).

- Too slow of an update process, due to the ISO' s procedures. The goal of the new ISO 12620 (2009) standard was, as we understood it, to introduce opportunities for sharing and openness. When the server was still online, it enabled users to declare new categories and create custom subsets. Unfortunately, when its content was 'locked', a period of major instability—perhaps even confusion —began in the field of terminology. For instance, there are multiple cases of duplicate records whose content does not match. We have already mentioned grammatical genders in note 17. There are also three categories for 'concept relation' (DC-88, DC-142, and DC-242), leaving the user unable to tell which one to use.

- Difficulties in arbitrating between the wide range of new proposals: *ISOcat* enabled custom categories, but the arbitration and validation processes remained obscure. Our exchanges with those in charge of data categories related to terminology as part of our research projects have not been fruitful. It should be acknowledged, however, that the arbitrator's role is an especially difficult one on a platform that is open to groups with such diverse interests, while industries that require interoperable linguistic data are seeking stability above all (Wright *et al.* 2014).

All these considerations raise the question of what a standard is. In practice, a standard is not an optimal solution so much as a consensus, within a group of humans with diverse interests, that does not

[20]. See the summary by Fuertes-Olivera and Tarp (2014, 104-128). For a more in-depth look, read Kockaert and Steurs (2015).

necessarily reflect a diversity of points of view. Practices that are presented as ideal can be influenced by a pressure group, or constitute an imperfect compromise between a number of approaches.

Data categories should not be 'locked' in an obsolete state of knowledge, and there is no point in regretting the obvious progress that was made with the *ISOcat* server. Some have argued that businesses require stability, but this is in contradiction to the creative dynamism of linguistic engineering. Does ending ISO category updates not also mean ending updates to TMLs such as TBX, by making it more difficult to include new data categories or freely redesigned data categories? Rather than returning to the *status quo ante*, we believe it would be preferable to solve the issues related to arbitration and adaptation to rapid progress. Otherwise, one might assume that ISO standards prevail over personal reflection and over the need to rethink terminographical practices in light of major advances in the translation industry and in linguistic engineering. The possibility for all users to create their own data category repository—as planned in the ISO/DIS 12620 (2018) project—seems a promising idea, and those who do so will most likely do it out of necessity, after careful thought, and following the new standard.

3 DATA MODELLING

The author of a structured document must not only identify data categories, but also provide a hierarchical structure that organises the data categories, and provide an interpretation of the structure. The TEI's structure still closely resembles a printed dictionary's layout, which was quickly deemed inadequate to create true electronic dictionaries. There is no point in adhering to a surface structure that involves many redirects using pointers to manage synonyms and equivalents, despite the fact that XML markup can—using XSL (*eXtensible Stylesheet Language*) stylesheets—transform the organisation of data in a document at will. Figures 1 and 2 illustrate the potential of this process: starting with the same markup, in the late 1990s, collaborators of the MLIS-DHYDRO project were able to produce a lexicographical (polysemic) version and a terminographical (monosemic and homonymic) version of the *Hydrographic Dictionary* (OHI 1994-1998)[21].

[21]. Other layout models were also produced, such as a monolingual homonymic dictionary.

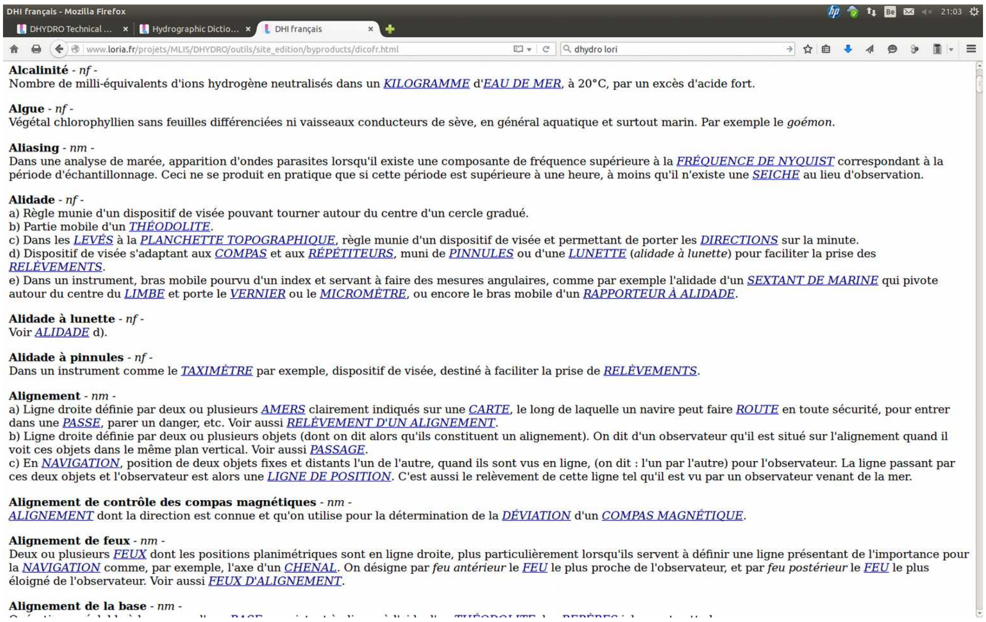

Figure 1. Transforming data with XML markup into a polysemic lexicography

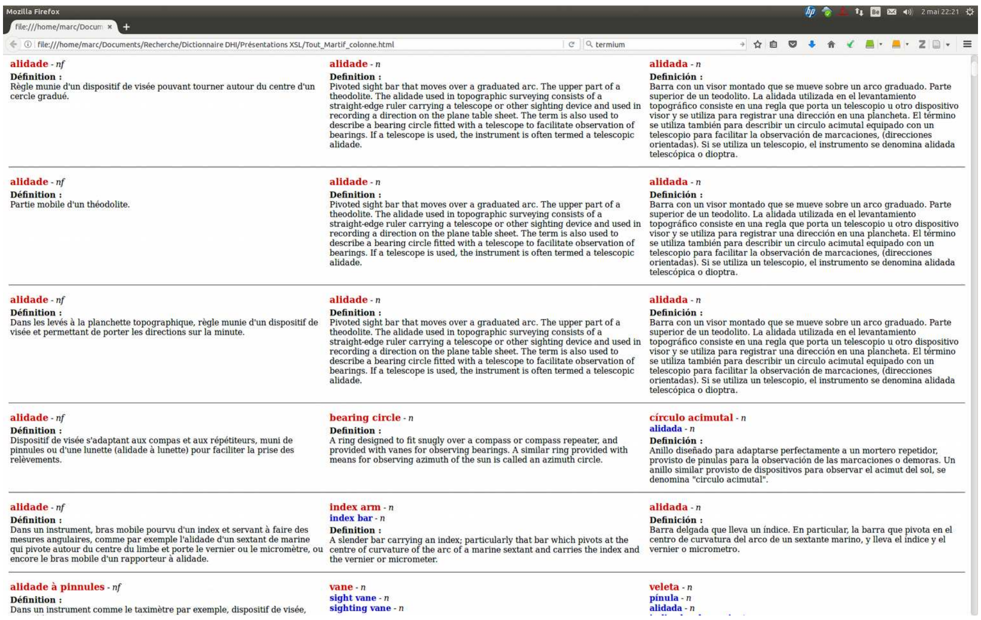

Figure 2. Transforming the same data into a homonymic terminography

3.1 The hub and spoke model

We shall not go into greater detail on the deeper theoretical implications of this demonstration (Van Campenhoudt 2002), which was largely ignored by the *doxa*. The purpose of this article is rather to examine the data modelling that enables such ‘tricks’, and that resulted in the ISO 16642 (2003) standard. The underlying model is the one we had presented at the university of Rennes in 1992, during the *1re Université d’automne en terminologie* [‘1st autumn university in terminology’] (Merten, Mertens, and Van Campenhoudt 1993), and that had already been implemented in 1990 in the *Termisti* software application (Blampain, Petrussa, and Van Campenhoudt 1992). This application had been built on the work of Jean-Michel Henning (1989a and 1989b) and his *MC4* software application*,* which used Eurodicautom’s format. The core idea—which we have largely contributed to theorising—is to manage multilingual and monolingual terminological data following a hub and spoke model where the semantic space that guarantees equivalence between languages serves as a pivot (figure 3).

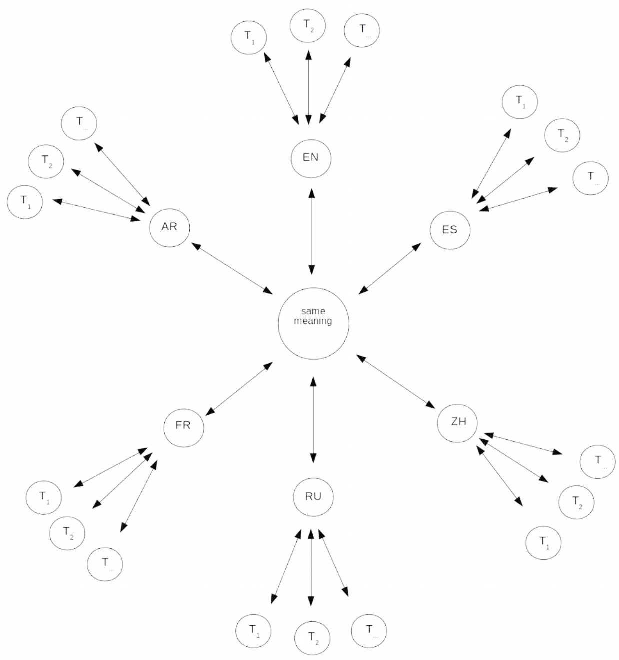

Figure 3. Data model of the *Termisti* software application (1990) for the UN’s six languages

In the middle, acting as a pivot, is the information common to all languages. This can consist in administrative or semantic data, an illustration, a documentary classification, etc., as long as it is independent from expression in a specific language. It is essential that this pivot correspond to the semantic space that demarcates equivalence between languages. Each branch is a language, containing the lexical units described. An unlimited number of languages can be linked to the central pivot, and an unlimited number of terms (T_1, T_2, T_3, etc.) meeting this definition can be included for all languages. In this model, properties are inherited from the higher level and semantic relationships are built between pivots (Romary and Van Campenhoudt 2001).

This hub and spoke model enables translation in any language combination and in any direction. It is free from the constraints of paper mentioned in section 1: there is no need for volumes upon volumes for various language pairs in specific directions (e.g. English to French). For instance, using the L^2-L formula, the UN's six official languages add up to thirty language pairs, which would require thirty different dictionaries stored as a cumbersome network architecture that experts in computer science or transportation refer to as a 'point to point' structure (figure 4).

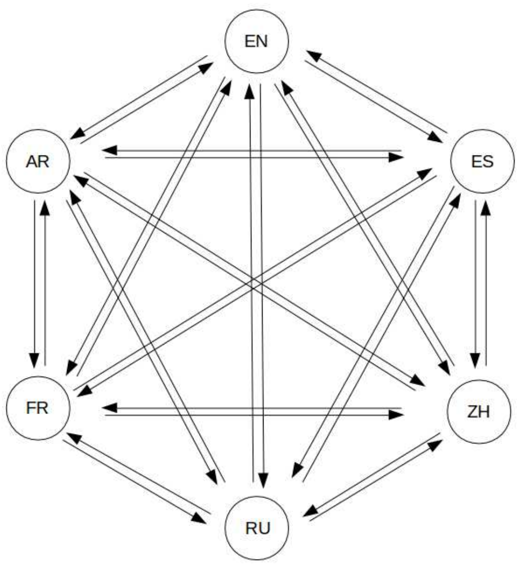

Figure 4. 'Point to point' model for the UN's official languages: 30 language pairs

Adding the European Union's 24 official languages to the pivot could cover as many as 552 language pairs for translation (24^2-24). This is the model that was chosen for the basic architecture of the EU's terminological database, IATE[22] (Fontenelle and Rummel 2014), as can be seen from its surface structure (definitions precede terms).

The implementation of hub and spoke models was analysed in detail in volume 17-2 of the *International Journal of Lexicography*, with a discussion of the *Simullda* by Maarten Janssen (2002). *Simullda* opened promising new perspectives compared to similar solutions like our own (Van Campenhoudt 2004), *Hub-and-Spoke* (Martin 2004), which is based on a pivot language, and *Eurowordnet* (Vossen 2004), which uses an 'inter-lingual-index'.

The hub and spoke model implies strict monosemy. It may seem like an onomasiological model, but it should be noted that we are describing here a computer model that does not determine the dictionary's final appearance, which may vary depending on specific needs (see figures 1 and 2). The question of delimiting meaning does not depend on the model chosen: it is possible to follow Wüster's logic and take a prescriptive conceptual approach where a term in each language is attached to a standardised concept. But one can also opt for a fully descriptive approach where meaning is identified using the traditional tools of lexical semantics (Van Campenhoudt 2001a) or by validating semantic attributes, like in *Simullda*. This means the model is only onomasiological in appearance, and the choice to enable a multilingual framework is what implies the domination of a restrictive monosemic approach.

As soon as it aims to be monosemic, the model implies homonymy. Of course, one could deny that signifiers such as *banana* or *stamp* have multiple homonyms, by calling upon the tradition—which is more a typographical one than a scientific one (see section 4.2)—of our monolingual dictionaries. However, monosemy is required in translation, as we have seen with the cases of *banane – banana* and *timbre – stamp* in the *Grand Robert & Collins* (2017). Unadventurous lexicologists who stick to descriptions of their native language may conjure various tropes—and, if required, engage in some intellectual acrobatics—to establish semantic links between the various acceptations in a given entry from

[22]. iate.europa.eu.

the *Oxford English Dictionary* or the *Petit Robert*, but they will eventually have to acknowledge that semantic relationships such as synonymy, hyponymy, and meronymy are only valid in a monosemic context. Also, by prioritising monosemy, the hub and spoke model is more similar to the electronic dictionaries used in automatic language processing (see section 1). It directly integrates the perspectives of artificial intelligence, as each pivot can easily correspond to a node in a semantic network. This will never be possible with polysemic entries in a traditional lexicography.

3.2 The challenge of ISO markup

The ISO 16642 standard (2003, then 2017), *Terminological Markup Framework* (TMF), was mainly drafted by Laurent Romary at the conclusion of the DHYDRO project, which he coordinated at Loria (Lorraine Research Laboratory in Computer Science and its Applications). This standard proposes adopting the hub and spoke model, and distinguishes several hierarchical levels (figure 5): pivot, language, term, and term component[23].

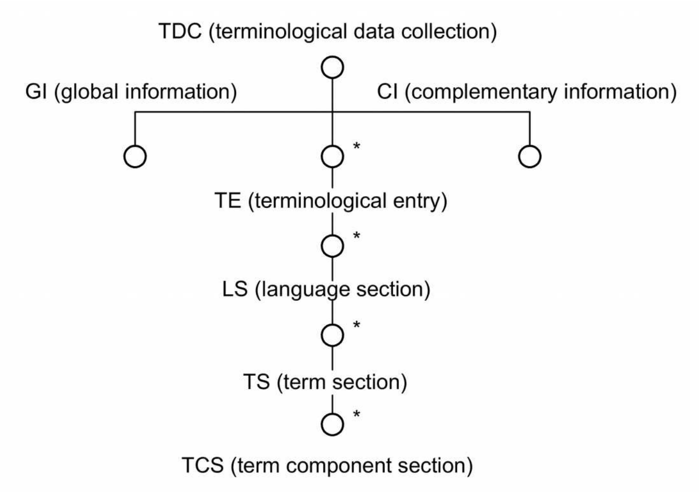

Figure 5. Excerpt from the ISO 16 642 (2017, 10) standard

[23]. As far as we know, the latter is seldom used, and we will not discuss it here.

All document type definitions (DTDs) that adhere to these details are *Terminological Markup Languages* (TMLs) and, as such, should be interoperable. The most well known TML is TBX, which is the ISO standard used in the localisation industry that we have already mentioned in the introduction. Still, as TMF originally had a DTD, it can be used without going through another TML, which was done —masterfully—for the *Termsciences* database[24]. TMF's DTD, which is strangely no longer a 'generic mapping tool' in the standard's second edition (ISO 16642 2017, IV), is only a few lines long and more an outline than an in-depth description. Based on our extensive experience in teaching, we can state that students generally find it particularly easy to use, whereas TBX requires that certain data categories be placed at specific levels and linked to specific tags (*descrip, descripNote, admin, adminNote, term, termNote,* etc.). The concept of such restrictions is certainly a good idea, which we will discuss in section 4, but we feel that the specific choices made for TBX are not always appropriate: the definition can be placed under the term, which is incompatible with the hub and spoke model that inspired TMF, and the context must necessarily—and logically—be placed at the level of the term, yet the type of context may be placed at three different levels.

9.3.1 Properties and descriptions of concepts

These data-categories describe properties of concepts. Some are allowed at the term level.

Data-category name	Data type	Target	Meta data-category	Level
audio	plainText	binaryData	<descrip>	langSet, termEntry, term
characteristic	plainText	none	<descrip>	term
conceptOrigin	plainText	none	<admin>	langset, termEntry, term
context	noteText	none	<descrip>	term
contextType	picklist	none	<descripNote>	langset, termEntry, term
definition	noteText	none	<descrip>	langSet, termEntry, term

Table 2: excerpt from the TBX standard (ISO 30042 2008, 17)

In reality, the purpose of a TML used for data interchange, such as TBX or possibly TMF, is to account for all the practices that exist in electronic dictionaries, including practices that might seem absurd[25]. This means that a single XML file that strictly adheres to the relevant DTDs can contain a number of records with a variety of structures, giving the appearance of a file that is attractive in terms of

[24]. www.termsciences.fr.

[25]. As a reminder, until its designers adapted it to the requirements of TMF and TBX, the industry-standard software application *Multiterm* strangely enforced a lexicographical perspective, as the definition was required to appear under the term.

interchange: but a TBX or TMF file that is valid with regards to the DTD may very well be barely usable in practice[26].

It should also be noted that the hub and spoke model is not a perfect guarantee against monosemy: it can easily be circumvented, with TBX as well as TMF, by declaring the same language multiple times in order to include multiple definitions, as we have observed in our students' work. This is a limitation of the markup language and of its *xml:lang* attribute. The workaround could be blocked by limiting the use of this attribute, or by limiting the use of the *language identifier* data category (DC-279), in a schema or DTD defining an original TML.

Besides, from an epistemological perspective, universal markup standards do not solve all theoretical problems, and some even contain underlying biases. As we have seen above with the definition, the level at which data categories are placed can be controversial. Definitions may reasonably be placed at the level of the inter-lingual pivot, or at the level of the language in a hub and spoke model. We have always believed that the hub and spoke model, considering it involves inheriting properties, requires placing the definition at the level of the language. But if we disregard this argument, it seems logical to place it at a higher level, since the pivot is a common semantic space. This, however, means using an *xml:lang* attribute for each definition.

An analysis of this point reveals that TBX's DTD[27] requires declaring in root element <martif> a default language for all records[28]. This seems especially problematic when promoters of the format provide examples of proper markup that include a definition with no *xml:lang* attribute at the highest level. In doing so, they contribute to spreading the idea of a dominating pivot language. Perhaps the language of 'concepts'?

---

[26]. At a time when some are considering taking part in ISO certification for terminological resources, it seems important to highlight this reality that is largely underestimated. For a practical demonstration, please see our examples online: http://difusion.ulb.ac.be/vufind/Record/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/275367/Holdings (for TBX) and http://difusion.ulb.ac.be/vufind/Record/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/275368/Holdings (for TMF).

[27]. Based on the DTD available at www.ttt.org/oscarstandards/tbx/TBXcoreStructV02.dtd.

[28]. This seems a little excessive considering the W3C's requirements (www.w3.org/TR/REC-xml, section 2.12, May 4, 2017)

```
<?xml version='1.0'?> <!DOCTYPE martif SYSTEM "TBXcoreStructV02.dtd">
<martif type="TBX" xml:lang="en">
        <martifHeader>
            <fileDesc>
                <sourceDesc>
                    <p>From an Oracle corporation termbase</p>
                </sourceDesc>
            </fileDesc>
            <encodingDesc>
                <p type="XCSURI">http://www.lisa.org/fileadmin/standards/tbx/TBXXCSV02.XCS</p>
            </encodingDesc>
        </martifHeader>
        <text>
            <body>
                <termEntry id="eid-Oracle-67">
                <descrip type="subjectField">manufacturing</descrip>
                <descrip type="definition">A value between 0 and 1 used in ...</descrip>
                <langSet xml:lang="en">
                    <tig>
                        <term id="tid-Oracle-67-en1">alpha smoothing factor</term>
                        <termNote type="partOfSpeech">noun</termNote>
                    </tig>
                </langSet>
                etc.
```

(ISO 30042 2008, 21)

This practice is also found in files provided as examples on TBX's official websites[29]. The use of this initial declaration even results in proposing a file in which the French definition given at the level of the pivot inherits an *xml:lang="EN"* attribute.

```
<martif type="TBX" xml:lang="EN">
        <martifHeader>
            [...]
        </martifHeader>
        <text>
            <body>
                <termEntry id="eid-VocCod-211.01">
                    <descrip type="subjectField">personnel</descrip>
                    <descrip type="definition">personne qui accomplit un travail copie ou
                    d'écriture</descrip>
                    <langSet XML:lang="fr">
                        <ntig>
                            <termGrp>
                                <term id="tid-voccod-211.01-fr1">copiste</term>
                                <termNote type="termType">entryTerm</termNote>
                            </termGrp>
                        </ntig>
                        [...]
                    </langSet>
                    etc.
```

medtronic_TBX.tbx

One final point should be mentioned: standards can conceal economic considerations, as the translation and localisation industries have a vast potential market for interchanges of data and services. Adding TBX import and export features to a piece of software dedicated to managing specialised

[29]. See www.tbxinfo.net/wp-content/uploads/2014/07/TBX-Default-Sample-Files.zip and www.ttt.org/oscarstandards/tbx/TBX-resources.zip.

dictionaries is a strong selling point. Still, after their purchase, users will need not only training on the software itself, but knowledge of ISO standards and the linguistic foundations of the TMF standard. Failing that, they might have to purchase additional services, e.g. to design the data model, or they could use their new tool in such a basic way that they will soon revert to a simple spreadsheet, as we have mentioned at the end of section 1. In the field of specialised translation, the well-documented mediocrity of terminographical practices unfortunately appears to prevail over standards when, as a response do the creation of non-ISO tab-delimited standard UTX[30], a *TBX-min*[31] format was suggested in order to ensure a transition from tab-delimited dictionaries to databases (Lommel *et al.* 2014). This reality certainly contributes to a perception amongst some that terminology is a 'poor man's lexicology' (Van Campenhoudt 2016, 592-593) even though, paradoxically, the terminologist's methodological reflection and modern tools have significant potential to develop lexicographical practices.

4 THE CHALLENGES OF A TRUE DIGITAL TRANSITION

As we have seen, interchange standards are necessarily vulnerable to absurd practices. This means we need to be stricter than the standard when it comes to managing our own data. *Termsciences* uses TMF tags, but a different TML appears to lie behind the design of the terminological records available online—one that is much more restrictive than TMF's DTD. Based on our experience in digitising multilingual specialised dictionaries, it is indeed essential to model data using a DTD that—as a true TML—adheres to TMF's hub and spoke model but is more restrictive than universal interchange formats that are, by nature, open to the most absurd structures (see section 3.2 and note 26). This implies acquiring a skill set that is currently quite uncommon.

[30]. www.aamt.info/english/utx.

[31]. www.tbxinfo.net/tbx-min-resources-and-tools.

4.1 From paper to screen

Many ‘electronic’ dictionaries simply display on screen the same content as the paper version, and have the exact same problems in terms of digital modelling. The main benefit of digitising this data is the possibility to quickly browse through entries based on electronically-stored text. What seemed like incredible progress two decades ago now seems barely significant.

The more traditional limitations are related to the fact that in a lexicographical approach, synonyms are scattered in alphabetical order and sometimes defined in different ways. For instance, the *Dictionnaire de l’océan* (CILF 1989), available online[32], offers very similar definitions of the terms *anadrome* and *potamotoque* and their antonyms *catadrome* and *thalassotoque*. A vague link such as ‘see…’ is provided between potential synonyms. Of course, these records can be input as distinct entries in a TML, but reflecting on their meaning, reference, and semantic relationship might result in deciding to merge them.

The DHYDRO project, which dealt with a very rich specialised lexicography (OHI 1994-1998), had already enabled the identification of a series of typical problems that should be solved as part of a conversion to a markup framework such as TMF: entry format, subdivision of acceptations, synonyms declared as part of the definitions, confusion between synonyms and hypernyms, vague semantic relationships between entries, competing equivalents, presence of various data categories (related to spelling, history, usage, etc.) in the definitions, etc. Administrators of large accumulated terminological databases currently have a very good idea of the strategies to implement in order to algorithmically identify content inconsistencies, which inevitable appear when many users are authorised to input data.

The considerable efforts required by the ‘retroconversion’ in a digital format of a dictionary available as a text processor file are largely ignored. Beyond the potential for a proliferation of data categories that can be identified in a rich lexicography, another problem is the diversity of meanings and synonyms. Even if the hub and spoke model should prove unsatisfactory *a priori* for some authors of

[32]. www.cilf.fr/unepage-terminologie-terminologie-1-1-0-1.html.

dictionaries, asking them to make an effort to adhere to it for a few entries would certainly help them understand the importance of strictly following standards when compiling a dictionary.

4.2 Towards new dictionary writing standards

Considering the cost involved in converting an existing dictionary to a TMF-compatible TML, it would make more sense to promote creating new dictionaries directly in a pre-structured format. This can be done by inserting tags while writing unstructured text, following a loose DTD; still, given the points raised above, it seems far more reasonable to make the effort to input information following a TMF-compatible structured model—ideally in a database—, making it possible to convert the data into any kind of layout and to automatically generate running text.

Here are a few examples of tips on methodology that we had given at the end of the DHYDRO project, regarding dictionary entries (table 3). Presentation practices for 'dictionary style' entries are often the result of text contraction processes, which used to be necessary to keep printing costs down. Some still attempt to justify this from a theoretical perspective (Svensén 2009, 102-104), but these practices have now become a major obstacle to an efficient use of electronic dictionaries. Dictionary shorthand is intelligible to human readers, but computers have difficulty interpreting it. By following strict input procedures into data fields, it will still be possible to generate dictionary-style entries that make use of such contractions if necessary.

Traditional dictionary		**Database**
labour (or labor)	must be stored as	**main entry:** labour **grammatical data:** n **synonym:** labor **grammatical data:** n
high frequency (HF)	must be stored as	**main entry:** high frequency **grammatical data:** n **synonym:** HF **grammatical data:** n **abbreviated form of:** high frequency
high water**:** double	must be stored as	**main entry:** double high water

		grammatical data: n
etc.		

Table 3: examples of writing instructions for inputting entries (DHYDRO project)

The idea to ensure maximum data granularity is now referred to as the '*principle of data elementarity*' in the ISO/DIS 12620 (2018: 6) standard. As we had once mentioned in the manual for DHYDRO[33], this principle implies that definitions should not contain information intended for other fields (see table 4).

In astronomy.	must be stored as	**domain:** astronomy
also called *astro-navigation*…	must be stored as	**synonym:** *astro-navigation*
wavemeter. …Also written as two words.	must be stored as	**main entry:** *wavemeter* **spelling variant:** *wave meter*
The name 'racon' is derived from the words radar beacon	must be stored as	**etymology**: *radar beacon*
etc.		

Table 4: examples of writing instructions for inputting definitions (DHYDRO project)

Still, in order to produce a rich electronic dictionary, one must go through the difficult process of strictly following markup standards. The experience of large terminological databases shows that it is very difficult to ensure that users systematically fill all data fields. From the point of view of ergonomics, it is much more comfortable to freely input text than to feel constrained by a structure—even one that was very well thought out. Conversely, writing unstructured text to produce a dictionary is especially tedious when one has little content to offer. This certainly explains the success of tab-delimited formats with translators, as we mentioned, as well as the need to offer simplified versions of interchange standards in order to ensure they are actually used.

CONCLUSION

[33]. termisti.ulb.ac.be/archive/dhydro/dhydro.html, "Data model".

Traditional paper dictionaries are written directly as text that is accessible to the reader. The main feature of electronic dictionaries is that their underlying structure is different from the surface structure that appears on screen. With the proper design, this underlying structure can allow the data to be exploited at its fullest. A monosemic hub and spoke model most likely offers the most potential: it can deal with multilingualism, it is compatible with lexicon-grammars, it creates a network of semantic relationships, it can determine equivalence, etc. Based on a tree structure, it enables inherited properties and makes it easy to use XML markup without the need for specific software. One can prefer a point to point model, but in any case there is little benefit to creating an electronic dictionary without identifying highly specific data categories.

Specialised dictionaries, especially multilingual ones that have rich content, are high value-added products for the fields of science and industry, who wish to use and share all the data they contain. This implies, of course, using ISO interchange standards that include a specific architecture and specific data categories, but this should be done knowingly. These standards are designed to represent all dictionaries, meaning they let users create content structures with little credibility. The specifications contained in the standard may also raise epistemological questions that are not insignificant. In any case, adherence to an ISO standard, e.g. to TBX's DTD, by no means guarantees the quality of a dictionary, nor of the software used to create it.

This can be due to the difficulty of improving standards within the ISO's decision-making processes, without a national mandate, and facing certain influential experts. At the same time, ignoring the content of these standards could mean eventually having to use them regardless, due to technological developments, without having contributed to their evolution. The lack of critical studies done on standards seems unfortunate, especially since most scientific publications about these standards are written by their very designers.

It is true that the publishing industry still seems largely unaware of the existence of interchange formats. Most specialised dictionaries are not written by authors who are up to date on the latest developments in linguistic engineering, and publishers only vaguely realise the potential for exploiting and adding value to their content using linguistic engineering. Should the situation persist, we might remain dependent for years to come upon pseudo electronic dictionaries that are read on a computer, tablet, or e-reader, almost like reading a paper dictionary. The challenge of automated dictionary interpretation in order to convert them into databases, while difficult, most likely remains a reasonable possibility given the reality of how human beings work.

BIBLIOGRAPHY

Béjoint, H. 2007. « Nouvelle lexicographie et nouvelles terminologies ». In L'Homme, M.-Cl. and Vandaele, S. (dir.) *Lexicographie et terminologie : compatibilité des modèles et des méthodes*. Ottawa : Les Presses de l'Université d'Ottawa, 29-78.

Bergenholtz, H. & Kaufmann, U. 1997. "Terminography and Lexicography. A Critical Survey of Dictionaries from a Single Specialised Field". *HERMES - Journal of Language and Communication in Business* 18: 91-125.

Blampain, D., Petrussa, Ph. & Van Campenhoudt, M. 1992. À la recherche d'écosystèmes terminologiques. In Clas, A. & Safar, A. (éd.), *L'environnement traductionnel. La station de travail du traducteur de l'an 2001. Journées scientifiques du réseau thématique de recherche Lexicologie, terminologie et traduction. Mons, 25-27 avril 1991.* Sillery et Montréal : Presses de l'Université du Québec et AUPELF-UREF, 273-282 (Universités francophones. Actualité scientifique) http://hdl.handle.net/2013/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/237290.

Budin, G. & Melby, A. 2000. "Accessibility of Multilingual Terminological Resources - Current Problems and Prospects for the Future". In *Second International Conference on Language Resourcesand Evaluation, Athens, Greece, 31 May - 2 June 2000*. Athens: ELREC, no pages. http://www.lrec-conf.org/proceedings/lrec2000/pdf/283.pdf.

Burnard, L. & Sperberg-McQueen, M. 2004. *TEI P4: Guidelines for Electronic Text Encoding and Interchange.* Charlottesville*:* University of Virginia Press. http://www.tei-c.org/Vault/P4/index.html.

Burnard, L. & Sperberg-McQueen, M. 2012. *TEI Lite: Encoding for Interchange: an introduction to the TEI. Final revised edition for TEI P5.* http://www.tei-c.org/release/doc/tei-p5-exemplars/pdf/tei_lite.doc.pdf

CILF. 1989. *Dictionnaire de l'océan : index anglais-français, allemand-français, espagnol-français.* Version en ligne. Paris : Conseil international de la langue française. www.cilf.fr/unepage-terminologie-terminologie-1-1-0-1.html.

Cissé, M.Th., Diagne, A.M., Muraille, P. & Van Campenhoudt, M. 2009. « Repenser le dictionnaire électronique multilingue dans un contexte d'aménagement linguistique ». In Depecker, L. & Dubois, V. (éd.), *L'idée de politique linguistique. Actes du colloque de la Société française de terminologie (Paris, 18 décembre 2008).* Paris: Société française de terminologie (Le savoir des mots, n° 6). 47-70. https://dipot.ulb.ac.be/dspace/bitstream/2013/237301/3/sft2008.pdf.

Dubois, J. 2014. *Le Lexis - le dictionnaire érudit de la langue française.* Paris : Larousse.

Dury, P. & Pincton, A. 2010. « Terminologie et diachronie : vers une réconciliation théorique et méthodologique ? ». *Revue française de linguistique appliquée* XIV (2) : 31-41.

Faber, P. & L'Homme, M.-Cl. 2014. "Lexical semantic approaches to terminology. An introduction." *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 20 (2): 143 –150.

Fontenelle, Th. & Rummel, D. 2014. "Term Banks". In Hanks, P. & Schryver, G.-M. de (eds). *International Handbook of Modern Lexis and Lexicography.* Berlin Heidelberg: Springer, 1-12. http://link.springer.com/referenceworkentry/10.1007/978-3-642-45369-4_21-1.

Fuertes-Olivera, P.A. & Tarp, S. 2014. *Theory and Practice of Specialised Online Dictionaries, Lexicography versus Terminography*. Berlin & Boston: De Gruyter Mouton. https://www.degruyter.com/view/product/248469.

Grand Robert & Collins. 2017. *Le Grand Robert & Collins : dictionnaire français-anglais, anglais-français*. Version en ligne 3.0. Paris : Dictionnaires Le Robert et Glasgow: HarperCollins. http://www.lerobert.com/espace-numerique/enligne/le-grand-robert-collins-en-ligne-12-mois.html.

Henning, J.-M. 1989a. « Exemples de continuité dans la gestion automatique de données terminologiques ». *Meta : Journal des traducteurs / Meta : Translators' Journal* 34-3 : 493-495.

doi:10.7202/003046ar.

Henning, J.-M. 1989b. "MC4: A Package for the Management of Terminological Data". In *Conference on Terminology ın Advanced Microcomputer Applications. Vienna, 22-23 November 1989.*

Humbley, J. 1997. "Is Terminology Specialized Lexicography? The Experience of French-Speaking Countries". *HERMES - Journal of Language and Communication in Business* 18: 13-31. doi:10.7146/hjlcb.v10i18.25410.

ISO 12200. 1999. *Computer Applications in Terminology -- Machine-Readable Terminology Interchange Format (MARTIF) – Negotiated Interchange = Applications informatiques en terminologie – Format de transfert de données terminologiques exploitables par la machine (MARTIF) – Transfert négocié*. Geneva: International Organization for Standardization.

ISO 12620. 1999. *ISO 12620. Computer Applications in Terminology – Data Categories = Aides informatiques en terminologie – Catégories de données*. Geneva: International Organization for Standardization.

ISO 12620. 2009. *ISO 12620. Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources*. Geneva: International Organization for Standardization.

ISO/DIS 12620. 2018. *ISO/DIS12620. Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources*. Draft international standard. Geneva: International Organization for Standardization.

ISO 16642. 2003. *ISO 16642. Computers Applications in Terminology – Terminological Markup Framework*. Geneva: International Organization for Standardization.

ISO 16642. 2017. *ISO 16642. Computers Applications in Terminology – Terminological Markup Framework*. 2nd ed. Geneva: International Organization for Standardization.

ISO 24613. 2008. *ISO 24613. Language Resource Management – Lexical Markup Framework (LMF)*. Geneva: International Organization for Standardization.

ISO 1087-1. 2000. *ISO 1087-1. Terminology Work – Vocabulary – Part 1: Theory and Application = Travaux terminologiques – Vocabulaire – Partie 1 : Théorie et application*. Geneva: International Organization for Standardization.

ISO 30042. 2008. *ISO 30042. Systems to Manage Terminology, Knowledge and Content – TermBase eXchange (TBX)*. Geneva: International Organization for Standardization.

ISO /DIS 30042. 2018. *ISO 30042. Systems to Manage Terminology, Knowledge and Content – TermBase eXchange (TBX)*. Geneva: International Organization for Standardization.

ISO /R 919. 1969. *ISO /R 919 : Guide pour l'élaboration des vocabulaires systématiques (exemple de méthode)*. Genève : Organisation internationale de normalisation.

Janssen, M. 2002. SIMuLLDA: *A Multilingual Lexical Database Application Using a Structured Interlingua.* Utrecht: Universiteit Utrecht. http://dspace.library.uu.nl/handle/1874/615.

Kockaert, H.J. & Steurs, F. (eds). 2015. *Handbook of Terminology*. Vol. 1. Amsterdam & Philadelphia: John Benjamins Publishing. http://www.jbe-platform.com/content/books/9789027269560.

Le Pesant, D. & Mathieu-Colas, M. 1998. « Introduction aux classes d'objets ». *Langages* 32 (131) : 6-33. doi:10.3406/lgge.1998.2164.

Lommel, A., Melby, A., Glenn, N., Hayes, J. & Snow, T. 2014. TBX-Min: A Simplified TBX-Based Approach to Representing Bilingual Glossaries. In *Terminology and Knowledge Engineering 2014*. Berlin: Gesellschaft für Terminologie und Wissenstransfer & Deutsches Institut für Normung, 147-156. https://tke2014.sciencesconf.org/conference/tke2014/eda_en.pdf.

Lozet, J. & Mathieu, Cl. 1990. *Dictionnaires de science du sol. Avec index anglais-français.* 2e éd. Paris: Technique et documentation – Lavoisier.

Martin, W. 2004. Simullda, the Hub-And-Spoke Model and Frames or How to Make the Best of Three Worlds? *International Journal of Lexicography* 17 (2), 175-187. doi:10.1093/ijl/17.2.175.

Ménard, L., Arsenault, M. & Joly, J.-Fr. 1994. *Dictionnaire de la comptabilité et de la gestion financière : anglais-français avec index français-anglais*. Toronto, Institut canadien des comptables agréés, Paris, Ordre des experts comptables et Bruxelles, Institut des reviseurs d'entreprises. http://catalogue.bnf.fr/ark:/12148/cb37668243c.

Merten, P., Mertens, J. & Van Campenhoudt, M. 1993. « Microglossaire, réseau notionnel et gestion informatique. Une expérience de recherche en Communauté française de Belgique ». In Gouadec, D. (éd.). *Terminologie & terminotique : outils, modèles & méthodes: Actes de la première université d'automne en terminologie.* Paris : Maison du dictionnaire., 277-293. https://

dipot.ulb.ac.be/dspace/bitstream/2013/237389/3/rennes93.pdf.

OHI. 1994-1998. *Hydrographic Dictionary, Dictionnaire hydrographique, Diccionario Hidrográfico*. 5th ed. 3 vol. Monaco: International Hydrographic Organization. Special Publication No. 32

Paasch, H. 1894. *From keel to truck. De la quille à la pomme de mât. Vom Kiel zum Flaggenknopf. Marine dictionary in English, French and German, amply illustrated by explanatory diagrams of the most important details.* 2nd ed. Anvers: H. Paasch, Hamburg: Eckardt & Messtorff, London,: Fisher. http://catalogue.bnf.fr/ark:/12148/cb437830563.

Quemada, B. 1967. *Les Dictionnaires du français moderne, 1539-1863, étude sur leur histoire, leurs types et leurs méthodes*. Paris, Bruxelles et Montréal, Didier. http://catalogue.bnf.fr/ark:/12148/cb35937768g.

Roche, Chr. 2013. « Représentations formelles en terminologie ». In *Verbal and Nonverbal Representation in Terminology. Proceedings of the TOTh Workshop 2013.* Copenhagen: Centre for Textile Research - University of Copenhagen, 23-35. https://hal.archives-ouvertes.fr/hal-01354676.

Romary, L. 2014. "TBX goes TEI – Implementing a TBX basic extension for the Text Encoding Initiative guidelines". In *Terminology and Knowledge Engineering 2014. Berlin, Gesellschaft für Terminologie und Wissenstransfer & Deutsches Institut für Normung* 116-126. https://hal.inria.fr/hal-00950862.

Romary, L. 2015. "TEI and LMF Crosswalks". *JLCL - Journal for Language Technology and Computational Linguistics* 30 (1): 47-70. http://www.jlcl.org/2015_Heft1/3Romary.pdf.

Romary, L., Salmon-Alt, S. & Francopoulo, G. 2004. Standards going concrete: from LMF to Morphalou. In *Proceedings of the Workshop on Enhancing and Using Electronic Dictionaries*. Association for Computational Linguistics, 22-28. http://dl.acm.org/citation.cfm?id=1610042.1610047.

Romary, L. & Van Campenhoudt, M. 2001. « Normalisation des échanges de données en terminologie : le cas des relations dites "conceptuelles" ». In *Actes des 4es Rencontres terminologie et intelligence artificielle*. Nancy: INIST-CNRS 77-86. https://dipot.ulb.ac.be/dspace/bitstream/2013/237402/3/tia4.pdf.

Svensén, B. 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. New

York: Cambridge University Press.

Temmerman, R. 2000. *Towards New Ways of Terminology Description: The Sociocognitive Approach*. Terminology and Lexicography Research and Practice. Amsterdam: John Benjamins Publishing Company. http://www.jbe-platform.com/content/books/9789027298638.

Van Campenhoudt, M. 2000. « De la lexicographie spécialisée à la terminographie : vers un "métadictionnaire" ? ». In Thoiron, Ph. & Béjoint, H. (éd.). *Le sens en terminologie.* Lyon: Presses universitaires de Lyon (Travaux du C.R.T.T.), 27-152. https://dipot.ulb.ac.be/dspace/bitstream/2013/237429/3/crtt.pdf.

Van Campenhoudt, M. 2001a. « Pour une approche sémantique du terme et de ses équivalents ». *International journal of lexicography* 14 (3) : 181-209.

Van Campenhoudt, M. 2001b. *Expérimentation de l'éditeur de catégories de données Salt Suite. Rapport final. Projet européen Standards-based Access to multilingual Lexicons and Terminologies (Projet SALT N° IST-1999-10951).* Bruxelles, Centre de recherche Termisti, Institut supérieur de traducteurs et interprètes.

Van Campenhoudt, M. 2002. « Lexicographie vs terminographie : quelques implications théoriques du projet DHYDRO ». *Travaux du LILLA* 4 : 91-103.

Van Campenhoudt, M. 2003. « L'évolution des dictionnaires de traduction du domaine maritime au XIX^{e} siècle : aux sources de De la quille à la pomme de mât ». *Chronique d'histoire maritime* 52 : 83-97.

Van Campenhoudt, M. 2004. « Réseau sémantique et approche componentielle des bases de données multilingues ». *International Journal of Lexicography* 17 (2) : 155-160.

Van Campenhoudt, M. 2016. « Terminologie et langues spécialisées dans les pays de langue romane ». In Albrecht, J. & Métrich, R. (éd.). *Manuel de traductologie.* Berlin : De Gruyter (Manuals of Romance Linguistics), 589-616. https://dipot.ulb.ac.be/dspace/bitstream/2013/235032/3/traductologie.pdf.

Van de Velde, W. 2014. « La base de données terminologique Babeliris : une modélisation sociale dans le respect des normes ISO ». *7^{e} Séminaire de terminologie de Bruxelles : Malentendus terminologiques et soins de santé.* 25 avril 2014. Vrije Universiteit Brussel. http://hdl.handle.net/

2013/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/219362.

Villeneuve, J-Fr. de. 2001. *Le lexique français-anglais, anglais-français de la bourse et des marchés financiers*, Paris: SOFICOM. http://catalogue.bnf.fr/ark:/12148/cb37646127z.

Vossen, P. 2004. "Eurowordnet: a Multilingual Database of Autonomous and Language-Specific Wordnets Connected via an Inter-Lingual-Index". *International Journal of Lexicography* 17 (2): 161-173. doi:10.1093/ijl/17.2.161.

Warburton, K. 2015. "Re-modelling the ISO TC37 Data Category Registry: Challenges and Theoretical Reflections". *TOTh workshop Terminology and Standards.* Luxembourg: European Parliament, Dec. 2015. http://www.termcoord.eu/wp-content/uploads/2015/12/3-Kara-ToTH-DCR-kul.pdf.

Wright, S.E., Windhouwer, M., Schuurman, I. & Broeder, D. 2014. "Segueing from a Data Category Registry to a Data Concept Registry". In *Terminology and Knowledge Engineering 2014. Berlin, Gesellschaft für Terminologie und Wissenstransfer & Deutsches Institut für Normung* 177-187. https://tke2014.sciencesconf.org/conference/tke2014/eda_en.pdf.

Wright, S.E., Windhouwer, M., Schuurman, I. & Kemps-Snijders, M. 2013. "Community Efforts Around the ISOcat Data Category Registry". In Gurevych, I. & Kim, J. (eds). *The People's Web Meets NLP*, 349-73. Berlin – Heidelberg: Springer. http://link.springer.com/10.1007/978-3-642-35085-6_13.

Wüster, E. 1968. *The Machine Tool: An Interlingual Dictionary of Basic Concepts; Comprising an Alphabetical Dictionary and a Classified Vocabulary with Definitions and Illustrations*. London: Technical Press.

Wüster, E. 1979. *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. 2 vol. New York: Springer.