

Categorisation frameworks in Termontography

Koen Kerremans

Erasmushogeschool Brussels

The content of a terminology base should be determined by the purpose(s) for which it will be used as well as the profile(s) of its potential users. In this view, each terminological project first requires defining what will be considered a term. In this article, we reflect on how – in the Termontography approach – requirements for multilingual terminology bases can be translated into frameworks of interrelated categories. These frameworks are templates for the extraction of terms and knowledge rich contexts from texts and will gradually evolve towards enriched and more fine-grained networks of semantic relations as the knowledge elicited from these texts is mapped to it. The implications of categorisation frameworks for methods in multilingual terminology description will be illustrated by frameworks set up in the FF POIROT and OMTOFIR research projects.

0. Introduction

The content of a terminology base should be determined by the purpose(s) for which it will be used as well as the profile(s) of its potential users. In this view, each terminological project first requires defining what will be considered a term. This can be done by consulting domain-experts, interviewing users and writing a requirements report and by analysing domain-specific texts in order to acquire insight for determining what knowledge is relevant.

In the **Termontography** approach (Kerremans et al. 2003), requirements for multilingual terminology bases are translated into language-independent frameworks of interrelated categories. Depending on their granularity level, these frameworks can provide detailed information with respect to the extraction of terms and knowledge rich contexts (Meyer 2001) from a multilingual corpus of texts. By adhering to a common base, terminographers will be able to decide more efficiently which terms are considered translation equivalents and thus need to be placed in the same terminological record. Moreover, when implemented in an application, a common framework with links to several terminologies can support the task of updating aligned or merged terminologies (Oliver et al. 1999; Steve and Gangemi 1996). In particular, this idea is present in projects for general language dictionaries such as Duden where common frameworks are used to support the automatic updating of lemmas occurring in electronic dictionary versions (Alexa et al. 2002).

There are several frameworks of interrelated categories available. These frameworks, such as WordNet (Miller 1995) and SIMPLE (Lenci et al. 2000), have primarily been developed for general language purposes only. The aim of this paper is to reflect on the categorisation framework in the Termontography approach. Apart from being a framework developed for special language purposes, we will discuss how this framework differs from models for harmonising multilingual, general language lexicons such as EuroWordNet (Vossen 1998) and MultiWordNet (Pianta et al. 2002), or even from generic frameworks for multilingual terminology bases (Vouros and Eumeridou 2002).

This article will be structured as follows: in the first section we examine some of the multilingual models for general languages. In the second section, we reflect on the (non-)applicability of these models in projects for special language purposes. Section three deals with the general design principles and features of the categorisation framework in Termontography. The implications of this categorisation framework for methods in multilingual terminology description will be illustrated by frameworks set up in the FF POIROT (section 3.1.1.) and OMTOFIR (section 3.1.2.) research projects.

1. Models for multilingual, general language lexicons

Four possible designs or models for multilingual, general language lexicons are presented in Vossen et al. (1997). In the first design principle, languages are linked by pairs. This model is especially useful in bilingual projects as it supports the establishment of precise equivalence relations across language-pairs. However, the model is not recommended when applied to more than three languages as it “multiplies the work by the number of languages involved” (Vossen et al. 1997: 2). This means for instance that in case of a multilingual database covering the 20 official languages of the European Union, one would need to establish 40 language-pairs.

Linking languages through an external intermediate language or Interlingua is in this sense an improvement over the first model as the number of language-pairs always equals the number of languages in the multilingual project. But an Interlingua should often be language-neutral as it needs to cover the entire vocabularies of all languages in a given project. Its artificial construction is therefore difficult to achieve and may even require constant revisions as a result of new words entering one or several languages.

This problem may be overcome in a third model in which all human languages are linked through one of the languages involved. It follows that in case of the 20 official European languages, only 19 pairs would need to be established for structuring the multilingual database. But since one human language is not able to cover all meanings encoun-

tered in the other languages as a result of the different settings (cultural, geographical, political, etc.) in which human languages are shaped, a multilingual database which is based on the design principles of this third model risks to be biased to the lexical and conceptual structure of the intermediate human language.

This drawback can be resolved in a fourth model where the intermediate layer is a non-structured list of categories. This list is a derivative from the supersets of categories in all the languages of the multilingual project that have been obtained through bottom-up analyses.

In the following subsections we focus on the models that have been applied for the development of the EuroWordNet (section 1.1.) and MultiWordNet (section 1.2.) lexical resources.

1.1. EuroWordNet

EuroWordNet is a multilingual lexical resource consisting of wordnets in several European languages. The wordnets are built along the same lines as the Princeton WordNet 1.5 (Fellbaum 1998), i.e. words relating to the same categories are grouped in synsets, which in their turn are related by means of basic semantic relations such as “hyponymy, meronymy, cause, roles (e.g. agent, patient, instrument, location)” (Vossen et al. 1997: 3). In an initial stage, EuroWordNet only covered Dutch, Italian, English and Spanish. After extension of the project, Czech, German, French and Estonian wordnets were added to it. EuroWordNet is considered to be an important resource for multilingual information retrieval.

In EuroWordNet, the fourth model has been adopted (section 1). In order to efficiently account for the number of languages involved and, at the same time, to guarantee the diversity of all languages, equivalent word meanings are linked through a non-structured list of categories, the so-called Inter-Lingual Index (ILI), while language-specific wordnets are stored independently in a central lexical database. According to Vossen et al., this design seemed “most beneficial with respect to the effort needed for the development, maintenance, future expansion and reusability of the multilingual database” (1997: 3).

Each category in the ILI functions as a record to which at least two language-specific wordnets are linked. All language-specific wordnets pointing to the same ILI record are considered equivalent. The equivalence relationship is indicated by the interlingual relation EQ_SYNONYM. But there are also markers to indicate complex interlingual relations, such as cases where lexical gaps occur or where meanings of words in several languages do not entirely overlap.

Most of the ILI-records have been derived from WordNet 1.5. In order to do this, a list of base-categories was defined for each language resource separately in an initial stage. Criteria to find those base-categories were: (a) the number of relations they share with other categories

and (b) their position in the hierarchy (Vossen et al. 1997). The 1059 base-categories which resulted from this analysis were then translated to the closest WordNet 1.5 synsets. Those synsets to which at least two language-specific base-categories were linked, were added to the ILI.

Difficulties in this approach are the identification of the right interlingual correspondence when a new language-specific wordnet is added in one language, the precise matching between a synset in the local wordnet and a synset in the ILI, or “how to control the balance between the languages” (Vossen et al. 1997: 3).

1.2. MultiWordnet

MultiWordNet is a multilingual lexical resource covering the English, Italian and Spanish languages. The model used for the development of MultiWordNet is based on the assumption that many conceptual relations defined for English in the Princeton WordNet (PWN), such as hyperonymy or hyponymy, can be shared across several other languages: if there are “two synsets in PWN and a relation holding between them, the same relation holds between the corresponding synsets in the new language” (Pianta et al. 2002: 293). Cross-language correspondence between synsets is defined by means of the relation ‘corresponds_to’.

In MultiWordNet the third model has been adopted (section 1). English synsets and relations serve as a framework for developing wordnets in the other languages. With this model, automatic procedures can be devised in order to speed up both the construction of corresponding synsets and the detection of divergences between PWN and the wordnet being constructed (Pianta et al. 2002).

A potential drawback of this model is that the analysis of the languages involved heavily depends on the lexical and conceptual structure of the English language. However, according to Pianta et al. (2002), this risk is considerably reduced by allowing the new wordnet to diverge, when necessary, from the PWN. In this way, MultiWordNet stresses the usefulness of a strict alignment between wordnets of different languages, while retaining the ability to represent true lexical idiosyncracies between languages (Pianta et al 2002; Bentivogli and Pianta 2003).

2. General frameworks for structuring multilingual terminologies

Lexical resources such as EuroWordNet and MultiWordNet can be used as supportive resources for a variety of NLP tasks (e.g. information retrieval and word sense disambiguation). They also constitute robust frameworks for supporting the development of structured terminologies. For instance, one of the major objectives of the Prometheus project was the development of a generic framework – based on EuroWordNet and

SIMPLE design principles – for organising multilingual terminological databases (Vouros and Eumeridou 2002). Another example is TermNet, a terminological database holding German terms on text technology and hypermedia, in which WordNet's design principles have been adopted by structuring the lexemes of the terminology in a network of related synsets. Moreover, in order to deal with the meronymy relation in the terminological resource, two types of meronymy were taken from EuroWordNet. As an extension to the WordNet lexical resource, several lexical relations were added to the TermNet model, such as the *ist_orthographische_Variante_zu* to indicate that at least two terms are exactly the same but have a different spelling (e.g. *Hyper-Link* and *Hyperlink*) and the relation *ist_Akronym_zu* which also indicates that two terms denote the same category but that one is an acronym of the other. An example of the latter is the term pair *HTML* and *Hypertext Markup Language* (Beißwenger et al. 2003).

Apart from generic relations, synsets in EuroWordNet and MultiWordNet can also be used to support the process of structuring multilingual terminology encountered in texts. However, these synsets are usually too general in order to cover the whole range of specific terms. They only provide the hyperonyms according to which terms may be further structured (e.g. the lexeme 'fraud' is a hyperonym of the term 'missing trader fraud') and these hyperonyms may even occur as terms themselves. This is for instance the case when they are used in a domain-specific text as a lexical variant of their hyponyms (e.g. 'scam' instead of its hyponym 'VAT scam').

Although the general language resources may provide (re)usable content (i.e. relations and synsets) for structuring multilingual terminology, they should not be used as categorisation frameworks for motivating all term selection processes, unless the purpose of the terminology project is to further enrich the content of these general language resources with domain-specific categories. An example of such a project is ArchiWordNet (Bentivogli et al. 2004).

3. The categorisation framework in Termonotography

The *Centre for Terminology and Communication (Centrum voor Vaktaal en Communicatie - CVC)* is working out a method, called Termonotography, for developing (multilingual) terminological databases in which theories and methods of the sociocognitive terminological analysis (Temmerman 2000) are combined with methods in ontology engineering (Sure and Studer 2003). The motivation for combining the two research areas derives from our view that existing methodologies in terminology compilation (Sager 1990; Cabré 1999; Temmerman 2000) and (text-based, application- and/or task-driven) ontology development have significant commonalities (Kerremans et al. 2003).

An important view in Termontography is that a knowledge analysis phase should ideally precede the methodological steps which are generally conceived as the starting-points in terminography: i.e. the compilation of a domain-specific corpus of texts (Moreno and Pérez 2001) and the understanding and analysis of the categories that occur in a certain domain (Meyer et al. 1997). This view results from the fact that terminological databases need to represent in natural language those items of knowledge or ‘units of understanding’ (Temmerman 2000) which are considered relevant to specific purposes, applications or groups of users (Aussenac-Gilles et al. 2002). In Termontography, the units of understanding as well as their intercategory relations are therefore structured in a common knowledge base or categorisation framework. On the one hand, this framework supports the information gathering phase during which a corpus is developed (Kerremans et al. 2003). On the other hand, it allows terminographers to establish specific extraction criteria as to what should be considered a ‘term’: i.e. the natural language representation of a unit of understanding, considered relevant to given purposes, applications or groups of users. Furthermore, the predefined knowledge also affects the terminographer’s working method as well as the software tools that will be used to support that working method (Aussenac-Gilles et al. 2002).

In section 3.1., we discuss the frameworks set up in two research projects. Next, we compare the design of the categorisation framework in Termontography with the general language models outlined in sections 1.1. and 1.2. (section 3.2.). In section 3.3., we reflect on some general issues pertaining to the development of the categorisation framework.

3.1. Examples of categorisation frameworks

In section 3.1.1., we will discuss a hierarchically structured categorisation framework partly used for the development of a quadrilingual terminological database in the FF POIROT project. In section 3.1.2., we will show that a categorisation framework can also become a complex network of categories and intercategory relations.

3.1.1. The FF POIROT project

Financial Fraud Prevention Oriented Information Resources using Ontology Technology (IST-2001-38248) is a European research project. The aim of the project is to explore the use of tools and methodologies to represent, mine and use an ontology of financial forensics in a wide variety of applications against value added tax (VAT) fraud and securities fraud. CVC’s main task in this project is to develop in four languages (English, Italian, French and Dutch) a terminological database. This mul-

tilingual terminology base will be used as a supportive resource throughout the developing stages of the ontology. It will assist ontology modellers in the formalisation process of domain-specific categories by providing terminological information in four languages about terms that refer to these categories. Moreover, the database will be integrated in several end-applications developed to identify for instance cases of securities fraud or VAT fraud in the languages involved.

In the context of VAT fraud detection, part of the categorisation framework that we will present in this section serves as information for the extraction of terms referring to categories that are required for identifying fraudulent transactions of goods between European member states. These categories are lexicalised in the national VAT legislations as well as the European directives on VAT. Field experts can point to these categories by visualising them together with their relations in a categorisation framework. One such important category is paraphrased in English as 'transactions for which no VAT is required'. This category is said to be culture-independent and human-language independent as all the European VAT legislations contain a section on particular transactions for which one does not have to pay VAT.

From the model visualised in figure 1 we can infer, through the relations 'is hyperonym of' and 'is hyponym of', that this category has four subcategories: 'transactions in which the supplier does not have the right to deduct VAT', 'transactions in which the supplier has the right to deduct VAT', 'transactions that occur outside the territory of the VAT legislation at stake' and 'transactions that are outside the scope of VAT'. Once these categories have been identified, the multilingual terminology that needs to be assigned to these categories, is searched for in the multilingual VAT law texts and mapped to the categorisation framework. For instance, in the Belgian VAT legislation, the first subcategory, 'transactions in which the supplier does not have the right to deduct VAT', is labelled in Dutch as *vrijstelling* and in French as *exemption*. These two terms are also used to denote the second subcategory 'transactions in which the supplier has the right to deduct VAT'. The relations between terms referring to the category 'transactions for which no VAT is required' and the terms that lexicalise the different subcategories are structured in the multilingual terminological database (Kerremans et al. 2003).

In this example, the meaning of the categories is paraphrased in English. However, the human language that is used in the categorisation framework merely serves as 'hub' to which the terminology in all the languages is mapped during the search phase (Kerremans et al. 2003).

The categorisation framework presented in this section merely consists of categories that have been structured hierarchically. In the next section, we will show that categorisation frameworks can also appear as complex networks in which categories are linked through conceptual and lexical relations.

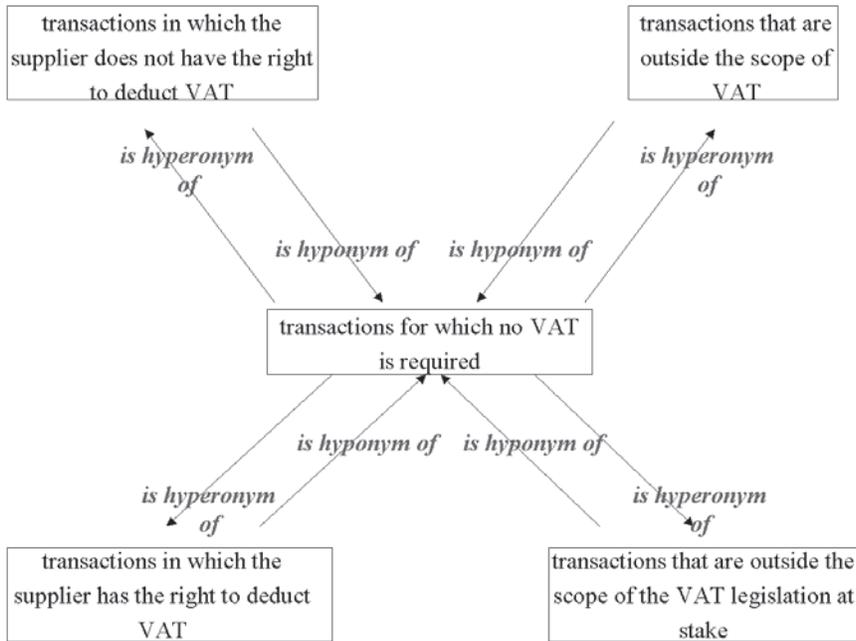


Figure 1: example of a categorisation framework in FF POIROT

3.1.2. The OMTOFIR project

One of the aims of the *Ontology-based, Multilingual Terminology on Functions in Retail* (OMTOFIR) project is to investigate to what extent the model presented in the ontologically-underpinned bilingual dictionary English-French of Dancette and Réthoré (2000) is reusable in the development process of a similar dictionary for the language pair ‘English-Dutch’. The advantage of this type of translation dictionary for translators is that they benefit from being subduced in a wealth of ontological information, i.e. information on how the term which has to be translated is related to other terms in the same lexical field or semantic network of related terms (Temmerman 2003). In the OMTOFIR project, Dancette and Réthoré’s terminological analyses of the terms denoting functions are taken as case-study (Vandervoort et al. forthcoming).

For instance, let us consider the terminological analysis of the term ‘dealer 1’. In French, this term is defined as follows: “Détaillant [...] à qui un fabricant ou fournisseur a accordé une concession 3 [...] pour la vente de ses produits [retailer who is offered, by a producer or supplier, a license for the selling of his products]” (Dancette and Réthoré 2000: 50, translation Koen Kerremans). As this French definition shows how the category ‘dealer 1’ is perceived in the English setting of retail, we claim that the terminological analysis in French can in fact serve as a

template for the development of a similar dictionary English-Dutch. In order to examine this, the terminological analysis in the dictionary on retailing was first converted into a network of categories and intercategory relations. This allows us to efficiently identify particular ‘knowledge chunks’ to which we map, after compilation of the Dutch texts on retailing, terms and relations found in the texts. In a knowledge chunk two categories are related to one another. The relation can be either conceptual (e.g. hyperonymy or hyponymy) or lexical (e.g. receive or grant). From the definition in Dancette and Réthoré (2000) we can derive the knowledge chunks listed below. Note that words or patterns referring to categories are placed between quotation marks and that words or patterns in italics refer to intercategory relations:

- a ‘dealer’ *is a hyponym of* a ‘retailer’
- a ‘retailer’ *is a hyperonym of* a ‘dealer’
- a ‘dealer’ *is given* a ‘license for selling products’
- a ‘dealer’ *sells* ‘products’
- ‘products’ *are sold by* a ‘dealer’
- a ‘license for selling products’ *is given to* a ‘dealer’
- a ‘producer’ *grants* a ‘license for selling products’
- a ‘license for selling products’ *is granted by* a ‘producer’
- a ‘supplier’ *grants* a ‘license for selling products’
- a ‘license for selling products’ *is granted by* a ‘supplier’
- a ‘license for selling products’ *is a hyponym of* a ‘license’
- a ‘license’ *is a hyperonym of* a ‘license for selling products’

As words can entail different meanings, categories and relations in the categorisation framework are presented to the user through glosses in one or several languages (see also section 3.1.1.). These glosses may either come from domain specialists or from textual resources. In order to translate the knowledge chunks into the categorisation framework, we found – for some of the terms denoting categories – the following English glosses in the CD-ROM version of the ‘Longman Dictionary of Contemporary English’:

- ‘dealer’ (someone who buys and sells a particular product, especially an expensive one)
- ‘retailer’ (a person or business that sells goods to customers in a shop)
- ‘license’ (an official document giving you permission to own or do something for a period of time)
- ‘product’ (something that is grown or made in a factory in large quantities, usually in order to be sold)
- ‘producer’ (a person, company, or country that makes or grows goods, foods, or materials)
- ‘supplier’ (a company or person that provides a particular product)

In the same dictionary, we found definitions for the following lexical relations:

- ‘sell’ ([the act of giving] something to someone in exchange for money)
- ‘give’ ([the act of letting] someone have something as a present, or to provide something for someone)
- ‘grant’ ([the act of giving] someone something or allow them to have something that they have asked for)

Figure 2 presents some of the knowledge chunks referring to the category paraphrased as ‘someone who buys and sells a particular product, especially an expensive one’ in a categorisation framework. The dotted arrows indicate how words in the lexicon – retrieved from texts – are mapped to their respective destinations in the categorisation framework.

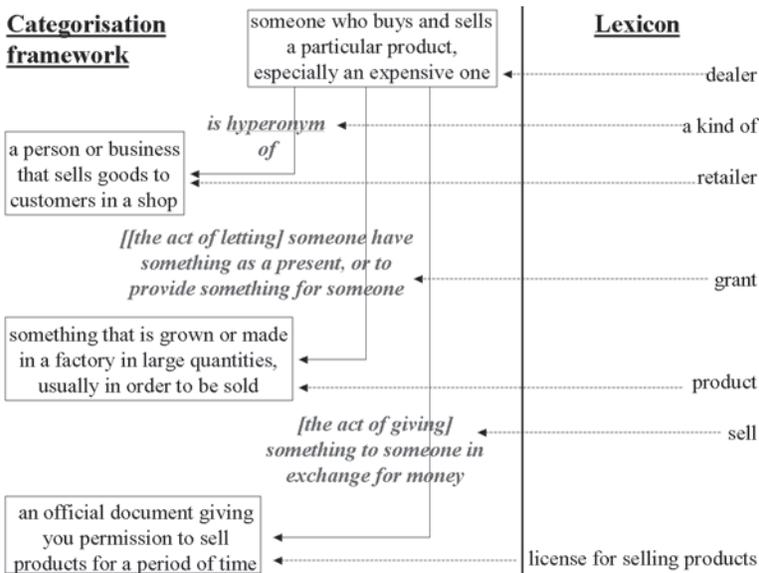


Figure 2: mapping to a categorisation framework in OMTOFIR

This figure shows how the categorisation framework is used as template for extracting terms and relations from a multilingual corpus of domain-specific texts. It allows us to scope the knowledge and restrict the selection of terms and relations only to those that lexicalise the knowledge chunks in the categorisation framework. This does not necessarily imply a static approach to terminology extraction. Depending on the requirements specified in each terminology project, the framework may also be

further enriched with culture-specific subcategories as a result of the cultural differences that may arise in the multilingual corpus. We will further explain this in the next section. We will also compare the design and content of the categorisation framework in Termonography to the EuroWordNet and MultiWordNet models presented earlier on (sections 1.1. and 1.2.).

3.2. Categorisation frameworks: features

This section discusses important features of the categorisation framework in Termonography. One important feature of the categorisation frameworks is that they do not necessarily have the same underlying model (section 1). For instance, the starting-point of the framework in the FF POIROT project was a list of culture-independent categories and subcategories, produced by domain-experts, that are considered important knowledge for the applications under construction. In the OMTOFIR project, the language-pair model (section 1) was chosen, taking English as the source language to which the Dutch terminology is compared.

The choice for different models in the categorisation framework results from one of the first methodological steps in Termonography: the analysis phase (Kerremans et al. 2003). This phase focuses on the analysis of the users as well as the possible applications and goals of the terminological database so that the choice for an underlying model (e.g. language-pair or Interlingua) depends on the type of terminological project (e.g. bilingual or multilingual project with one source language). It follows that compared to EuroWordNet and MultiWordNet, the categorisation framework in Termonography allows for different models to be represented.

Another important feature of the categorisation is the fact that all models are represented according to the same design principles: i.e. categories and, if necessary, intercategory relations are presented to terminographers in terms of human language phrases. This is mainly because paraphrases are easier to understand than an artificial Interlingua and because it is possible to provide all categories with paraphrases, even though some categories are culture-specific. For instance, the Italian term *esportatore abituale* is a VAT legislative term denoting a category that does not occur in the English UK VAT legislation and yet we are able to explain its meaning in English.

3.3. Categorisation frameworks: issues

In this section we discuss some important issues regarding categorisation frameworks in Termonography. One issue pertains to the reusability of the content of the categorisation framework. In order for the content of

the framework to be reusable, i.e. to be used more than once in several terminological projects, it is important that categories are presented in multiple dimensions instead of a univocal taxonomical hierarchy. The reason for doing this is because categories can be classified according to multiple dimensions:

[...] the concept “Optical Storage Media” can be classified according to “writability” as a “read-only media”, “write-once media” and “rewritable media”. It can also be classified according to “physical form” into any of the concepts “optical disc”, “optical tape”, “optical film”, “optical card” and “digital paper”. Vouros and Eumeridou (2002: 249)

Due to the multiple dimensions of categories, the location of a category in the framework should not be fixed. Terminographers should be able to reuse categories, including their specific relations to other categories, without being restricted by the structure of the framework from which categories are adopted.

The reusability issue does not imply that the taxonomical structure should be rejected completely. In order for the categorisation framework to be shareable, i.e. to be shared among several terminological databases, there should be a common upper-layer of categories according to which all other categories can be classified hierarchically. For instance, in Moreno and Pérez (2001) this upper-layer consists of the category ‘all’ further divided into the categories ‘event’, ‘object’ and ‘property’. The ‘property’ category has the two subcategories: ‘attribute’ and ‘relation’.

The third issue relates to the implementation of the categorisation framework in a computer system. When implemented, the categorisation framework becomes in fact a relational database from which a system must be able to deduce new facts, given the knowledge represented in and mapped to the framework. It follows that although the categorisation framework is presented to the terminographer in terms of human language paraphrases, there should be a formal knowledge representation language behind it which allows a system to ‘understand’ the meaning of domain-specific categories and to be able to derive new facts from given knowledge (which will facilitate to a large extent the terminographer’s task of structuring terminology in a database). In order to provide a computer system with adequate descriptions of e.g. domain-specific categories, a formal representation must support the explicit specification of semantic relations that exist among the categories (Kerremans et al. 2004). For instance, in the FF POIROT project, semantic relations between categories are formally represented in terms of lexons. Lexons are grouping elements further composed of a context identifier g (e.g. a European VAT directive), a starting term (e.g. taxable person) t_1 , a second term (e.g. tax) t_2 and two roles r_1 (e.g. pays) and r_2 (e.g. is paid by). Terms and roles appear in a semantic relationship which receives, through the use of the context identifier g , a particular meaning in a given context? (e.g. VAT). This ideational context is externalised by a set

of resources, such as documents, graphs and databases (Zhao et al. 2004).

A final important issue pertains to culture-specific information. Some terminological projects require the explication of possible semantic distinctions between equivalent terms. This may be the case in terminological projects in which legislative terms used in different countries are described and compared. For instance, although the English term *infanticide* and the Norwegian equivalent *barnedrap* both refer to the category paraphrased as ‘killing of a child by its mother’, in the English law the mother is accused of this offence when the child is under 12 months, whereas in the Norwegian law the term *barnedrap* only applies if the killing took place during or up to 24 hours after birth (Lind 2004).

4. Conclusion

In this paper we have emphasised that any terminological project should ideally start from an analysis of the users, applications and goals of the terminological database. In order to determine what (linguistic) knowledge is relevant given the specified requirements and, consequently, what should be considered a relevant term, we have proposed the development of a categorisation framework. This framework, which is an important component in the Termontography method, lists all the relevant categories and (if necessary) intercategory relations that are considered important within the scope of a terminological project. We have compared the categorisation framework in Termontography to general language models such as EuroWordNet and MultiWordNet and discussed important issues such as the representation of culture-specific information.

In further research, we intend to seek for tools that are able to implement and visualise categorisation frameworks. Such tools will be integrated in a workbench that will support the Termontography workflow.

Bibliography

- Alexa, M., B. Kreissig., M. Liepert, K. Reichenberger, L. Rostek., K. Rautmann, W. Scholze-Stubenrecht, & S. Stoye (2002). “The Duden Ontology: An Integrated Representation of Lexical and Ontological Information.” Paper presented at *OntoLex 2002* (University of Las Palmas de Gran Canaria, 27th May 2002).
- Aussenac-Gilles, N., A. Condamines & S. Szulman (2002). “Prise en compte de l’application dans la constitution de produits terminologiques.” Le Maitre, J. (ed) (2002). *Actes des 2^e Assises Nationales du GDR 13*. Nancy/Toulouse: Cepaduès Editions, 289-302.

- Beißwenger, M., A. Storrer & M. Runte (2003). "Modellierung eines Terminologienetzes für das automatische Linking auf der Grundlage von WordNet." *LDV-Forum* 19(1/2), 113-125.
- Bentivogli, L. & E. Pianta. (2003). "Beyond Lexical Units: Enriching Wordnets with Phrasets." Paper presented at the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL03, 12-17 April 2003).
- Bentivogli, L., A. Bocco & E. Pianta. (2004). "ArchiWordNet: Integrating WordNet with Domain-Specific Knowledge." Sojka, P., K. Pala, P. Smrz, C. Fellbaum and P. Vossen (eds) (2004). *Proceedings of the Second International WordNet Conference (GWC 2004)*. Brno: Masaryk University, 39-46.
- Dancette, J. & C. Réthoré (2000). *Dictionnaire analytique de la distribution; Analytical Dictionary of Retailing*, Montréal, Presses universitaires de Montréal.
- Cabré, M. (1999). *Terminology: Theory, methods and applications*. Amsterdam/Philadelphia: John Benjamins.
- Fellbaum, C. (1998). "A semantic network of English: the mother of all wordnets." *Computer and the Humanities* 32(2-3), 209-220.
- Kerremans, K., R. Temmerman & J. Tummers (2003). "Representing multilingual and culture-specific knowledge in a VAT regulatory ontology: support from the Termonography approach." Meersman, R. & Z. Tari (eds) (2003). *OTM 2003 Workshops*. Tübingen: Springer Verlag, 662-674.
- Kerremans, K., R. Temmerman & J. Tummers (2004). "Discussion on the Requirements for a Workbench supporting Termonography." Paper presented at the *Eleventh EURALEX International Congress*. (University of Bretagne Sud, July 6-10 2004).
- Lenci, A., N. Bel, F. Busa, N. Calzolari, E. Gola, M. Monachini, A. Ognonowski, I. Peters, W. Peters, N. Ruimy, M. Villegas & A. Zampolli (2000). "SIMPLE: a General Framework for the Development of Multilingual Lexicons." *International Journal of Lexicography* 13 (4), 249-263.
- Lind, A. (2004). "Navigating Through the Minefields. On the Making of an English-Norwegian Dictionary of Law." Paper presented at the *Eleventh EURALEX International Congress*. (University of Bretagne Sud, July 6-10 2004).
- Meyer, I. (2001). "Extracting knowledge-rich contexts for terminography: A conceptual and methodological framework." Bourigault, D., C. Jacquemin and M.-C. L'Homme (eds) (2001). *Recent Advances in Computational Terminology*. Amsterdam/Philadelphia: John Benjamins, 279-302.
- Meyer, I., D. Skuce, J. Kavanagh, L. Davidson (1997). "Integrating Linguistic and Conceptual Analysis in a WWW-based Tool for Terminography." Paper presented at the *Joint International Conference of the Association for Computers and the Humanities and the Association for Literary & Linguistic Computing* (Queen's University, June 3-7 1997).
- Miller, G.A. (1995). "WordNet: a lexical database for English." *Communications of the ACM* 38 (11), 39-41.
- Moreno, A. & C. Pérez (2001). "From Text to Ontology: Extraction and Representation of Conceptual Information." Paper presented at *Conférence TIA-2001* (INIST-CNRS, 3-4 May).
- Oliver, D. E., Y. Shahar, M. A. Musen, & E. H. Shortliffe (1999). "Representation of change in controlled medical terminologies." *Artificial Intelligence in Medicine* 15 (1), 53-76.

- Pianta, E., L. Bentivogli & C. Girardi (2002). "MultiWordNet. Developing an aligned multilingual database." Paper presented at *the 1st International Conference on Global WordNet* (Central Institute of Indian Languages, 21-25 January 2002).
- Sager, J. C. (1990). *A practical course in terminology processing*. Amsterdam: John Benjamins.
- Steve, G. & A. Gangemi (1996). "ONIONS Methodology and Ontological Commitment of Medical Ontology ON 8.5." Paper presented at *the 10th Knowledge Acquisition Workshop*. (Banff, November 1996)
- Sure, Y. & R. Studer (2003). "A methodology for Ontology-based Knowledge Management." Davies, J., D. Fensel, & F. Van Hamelen (eds) (2003). *Towards the Semantic Web. Ontology-Driven Knowledge Management*. New York: John Wiley and Sons, 33-46.
- Temmerman, R. (2000). *Towards New Ways of Terminology Description. The sociocognitive approach*. Amsterdam: John Benjamins.
- Vandervoort, V., U. Knops, H. Stengers, R. Temmerman, & K. Kerremans (forthcoming). "The Description of Function Terms in Retailing: Towards a Methodology for the Creation of Multilingual Terminological Resources."
- Vossen, P., P. Díez-Orzas & W. Peters (1997). "The Multilingual Design of the EuroWordNet Database." Paper presented at *the IJCAI-97 Workshop WP24 on Ontologies and Multilingual NLP* (Nagoya, 23 August 1997).
- Vossen, P. (1998). "EuroWordNet: building a multilingual database with wordnets for European languages." *The ELRA Newsletter* 3(1), 7-10.
- Vouros, G. & E. Eumeridou (2002). "Simple and EuroWordNet. Towards the Prometheus ontological framework." *Terminology* 8 (2), 245-281.
- Zhao, G., R. Coppens, K. Kerremans, J. Kingston, R. Meersman, R. Temmerman and F. Verlinden (2004) "Engineering an Ontology of Financial Securities Fraud." Paper presented at the *OTM 2004 Workshops* (Agia Napa, 25-29 October 2004).

