The translation of domain-specific languages and multilingual terminology management

Rita Temmerman & Uus Knops
Erasmushogeschool, Brussel

In 2003, in their editorial statement to the periodical *Terminology*, L'Homme, Heid and Sager pointed out that even though the disciplinary nature of terminology is still a subject of some debate, many people are now recognized as terminologists (2003: 151)[1]. Moreover, specialists who have a background in computer science, linguistics, translation, information science have also shown an interest in terminology theory and in methods for terminography. This special issue of LA, New Series, on the translation of domain specific languages and multilingual terminology management proves these authors' point that terminology is a multidisciplinary achievement, establishing cooperative links with the above-mentioned areas, to which we would like to add knowledge engineering and semantic Web application development. At the same time the contributions to this issue show the potential impact of terminology projects on the quality of translations in a rapidly changing multilingual environment. Terminology remains a research field in full development, striving to keep up with technological innovation. Professional translators will belong to the group of beneficiaries.

The need for terminology resources has increased due to the globalisation of scientific, technical and business exchanges and the development of international communication networks. Thanks to the WWW, easy access to existing multilingual terminology sources has materialised. Shared efficient multilingual terminology management can contribute to better quality translations and more effective communication.

At least three shifts have affected the discipline of terminology over the last decade: the shift towards computational terminology management, the linguistic shift in the theory of terminology and the ontology shift which combines the (semi-) formalised representation of semantic relationships with terminological management systems.

Taking the perspective of the translator, in this issue of LA we invited contributions that highlight research in domain specific languages, in terminology and in translation-oriented terminography (theme 1); contributions dealing with new developments in terminology theory and the relationship between the discipline of terminology and other cognitive disciplines like philosophy of science, epistemology and semiotics (theme 2); papers on multilingual terminography and computational terminology management (theme 3) and papers on terminological knowledge bases and ontologies (theme 4).

Theme 1: Domain-specific languages, terminology and translation, translation-oriented terminography

Translators always need state-of-the-art solutions when translating the terminology specific to e.g. legal, technical, scientific, medical or economic texts. This sub-section holds submissions describing research and experience in the field of domain-specific translations and papers exploring translator-oriented terminology. More specifically, contributors reflect on one or several of the following issues. What is the impact of the dictionary user's profile on the terminographer's working method? To what point and in what format should semantic relationships between terms and categories be made explicit? How can the semasiological approach in meaning description reinforce the onomasiological approach, and vice versa? Which problems in multilingual terminography are language-specific and which are not?

Ulrike Oster discusses how the understanding of 'term' in Traditional Terminology theory reduced the lexical problems of technical translation to a mere substitution of the source-text term by a target-text term. In translation studies, however, a number of issues have been highlighted which are not covered by traditional terminology theory, e.g. cultural specificity and the importance of textual and pragmatic considerations. In her article the author analyses how the new communication and cognition-oriented approaches to terminology account for these aspects of technical translation. Subsequently, she demonstrates in German-Spanish parallel texts on ceramics how these languages differ in word formation conventions (single words vs. multiwords, derivation vs. compounding), and in the adoption of different techniques to enhance text cohesion or to avoid text redundancy. She concludes her study by observing that translators need to have domain-specific knowledge, knowledge of domain-specific terms and linguistic knowledge of text conventions in order to be able not only to create adequate neologisms in the case of lexical gaps, but also to adequately assess translation proposals as found in conventional dictionaries.

Eva Wiesmann presents a new approach for the description of legal terms attempting to overcome the limitations posed by the traditional conceptual understanding of Wüster's theory of terminology. The central question deals with the role of definition, which is of importance in the delimitation of legal terms, but more particularly in the formation of legal concepts. Viewed from the perspective of a modern conceptual understanding, legal concepts are the result, not the object of definition. Often, however, they may be formed through definitions only in the core area, but not in the marginal area important for the differentiation of legal concepts, and this means that further debate on the objects of law is required. Wiesmann proposes a comprehensive approach where a legal term is not only defined, but also differentiated from related legal concepts. By defining legal terms, and not legal concepts, a basic precondition is created for taking into account not only the needs of the subject specialist but also the needs of the legal translator.

Taking legal translation as a case in point as well, **Marella Magris** emphasises that the needs of the potential users ought to be taken into account when the structure of a terminological database is defined. If the database is set up for translators, the terminologist should assess the specific problems typically found in legal translation as well as the use of terminological resources by legal translators. In her opinion users' needs and users' behaviour are extensively investigated in specialised lexicography, but seem to be a rather under-researched area in terminology and terminography. She aims at drawing up a list of requirements concerning terminological information for legal translators. A properly constructed and reliable language resource needs to meet these requirements. She draws on relevant lexicographic experience, on some empirical observations concerning the terminological needs of specialised translators and on translation-oriented terminography with particular reference to legal terminology.

In her paper on synonymy, polysemy and hyperonymy **María Carmen Acuyo Verdejo** aims at identifying and describing the main terminological problems arising from the continuing harmonisation of the different national legislations within the European Union. This has resulted in the creation of a new specialised language, that of European Law, which can cause synonymy, polysemy and hyperonymy to emerge when contacting with national legal languages. Both the comprehension and the quality of the translated text can be seriously affected by these phenomena. She illustrates this problem taking summons from two different legal systems, the Spanish and the Italian. Special attention is paid to terminology that is specific for summons, the person summoned and the type of text with which the person in question is summoned. She outlines some strategies to help out the translator who needs to solve terminological problems in this domain.

Judith Lavoie points out the lack of a specialised dictionary dealing with securities as a subject field in Canada. The concept of securities refers to a large, interdisciplinary sphere that includes preferred shares and partnerships, as well as fraud and broker dealers. In her methodological article she distinguishes four phases in the making of an English-French dictionary on securities: compiling literature on the subject, establishing the nomenclature, deciding on the microstructure and the description of the bilingual lexicon and its publication medium. The following objectives are set: first, to compile the Canadian terminology applying to securities; second, to offer to researchers, translators and professionals a lexicographical tool that is complete and useful in that it pays specific attention to the description of the conceptual relations among terms; and third, to verify the legal aspects of the terms defined.

Theme 2: New developments in terminology theory and adjacent disciplines

Terminology as a discipline has been enriching and strengthening its links with linguistic disciplines like cognitive semantics, computational linguistics, lexical functional linguistics as well as text linguistics, sociolinguistics and historical linguistics. This interdisciplinary influence has prompted quite a few scholars to question the principles of traditional terminology theory. The contributions to sub- theme two reflect on theoretical aspects of terminology, going into issues like: the concept versus the category approach, knowledge representation, guidelines for terminological definitions, how to deal with synonymy, polysemy, metaphor, etc. Contributions also highlight the differences and parallelisms between lexicography and terminography, terminology and knowledge management.

The purpose of **Tanja Collet**'s article is to define the term within the theoretical framework of text linguistics. As she sees it, in a text for specific purposes, a term has two fundamental functions: its naming function and its cohesion and coherence producing function. The definition of 'term' proposed in this article accounts for both functions of the term. In doing so the emphasis is on the term's original and dynamic contribution to text cohesion and to text coherence, two characteristics of a text that were not considered in Traditional Terminology.

As **Jeanne Dancette & Marie-Claude L'Homme** point out, it is now widely acknowledged that terms enter into a variety of semantic relationships with other lexical units and that classic taxonomies and meronymies represent only a small part of these relationships. This can be seen in recent specialized dictionaries that account for derivational relationships, co-occurrents, synonyms, antonyms, etc. It also has been underlined in several articles written by terminologists, linguists or computational scientists working with specialized corpora. They discuss the advantages and shortcomings of trying to account for semantic relations between terms using the formal lexico-semantic framework of Mel'čuk. Their research is based on a long-term project that aims at converting Dancette & Rhétoré's English-French paper dictionary on the subject of retailing into a relational database. They show that even if lexical functions have several advantages, a number of pragmatic decisions must be made to accommodate the description of specialized terms.

Philippe Selosse goes back in history and aims to bring to light the underlying pattern of the botanical nomenclature in the Renaissance. His analysis shows that the nomenclature was conceived within the frame of the Renaissance *episteme*, i.e. a logical and religious frame and is characterized by a threefold purpose: the terms have to give an essential definition of plants, to reflect the classification of the botanists, and be easy to remember. The nomenclature is structured by philosophical (logical division), linguistic (anaphora, polysemy) and cognitive (typicality) principles. The author's analysis leads to a definition of the concept of "nomenclature" in the Renaissance.

Tommaso Pellin studies the lexicon of political economy in nineteenth century China and highlights the influence of the contemporary social background on the coinage of technical lexicons. During the second half of the nineteenth century, Chinese scholars translated many Western works in various scientific disciplines, which involved the coinage of new terms. Pellin describes the impact of the expectations of contemporary readers on the process of coinage of the language of political economy. He analyses the first two Chinese translations of English politico-economic essays which appeared on either side of 1895. He shows the link between the lexicological choices of the translators and the expectations of their public.

Theme 3: Computational multilingual terminology management

The means to construct terminological resources have evolved considerably. Domain-specific texts are available and easily accessible (e.g. on the Internet) and techniques borrowed from natural language processing, information retrieval, corpus linguistics, AI, etc., facilitate the extraction of specialised knowledge. Still, terminological knowledge can be used to further improve major applications, such as computer-aided translation, document indexing and filtering, and corporate memory management. Various types of information in electronic format (corpora, databases, and internet) have an impact on methods and principles for terminography and give rise to computer-based terminology management systems and applications of computer-based terminology management; like the exploitation of corpus-based term extraction and translation.

Anja Rütten reflects on why and how conference interpreters need special software. Her article shows the differences between the software requirements of interpreters and those of translators or terminologists. On the basis of these considerations, a model for a special software solution for conference interpreters is presented.

Pilar Sánchez-Gijón considers how electronic communication accelerates the exchange of knowledge and information in areas of specialized knowledge. This state of affairs forces anyone involved in such communication (e.g. technical writers, technical translators) to remain up to date with new developments. Not only do professionals belonging to this group have to master the standard terminology of each specialized domain, they must also assimilate and understand the subject matter within which they are working. This article proposes a method for assembling and using specific corpora via the Web, with a view to extracting from them systematic and bilingual knowledge relating to terminology, the conceptual relations between terms, and the knowledge that they represent. Special attention is devoted to the strategies that will enable professionals to use such corpora in English and in Spanish.

Oliver Streiter, Natascia Ralli, Isabella Ties and Leonhard Voltmer introduce the reader to BISTRO, the online platform for termino-

logy management which facilitates structuring terminology without entry structures. BISTRO supports the translation process in various phases. It hyperlinks a terminological database with bilingual and trilingual corpora and provides tools for term extraction (TE), term recognition (TR) and keyword-in-context (KWIC). BISTRO's architecture is open for new tools and contents, providing at the same time the interface for the management of the underlying data structure and the constant update of the terminological data.

Carles Tebé and Teresa Cabré point out that computer-aided translation systems (CAT) based on Translation Memories (TM) are a widely diffused technology that uses database and code-protection features to improve the quality, efficiency and consistency of the human translation process. These systems basically consist of a textual database in which each source sentence of a translation is stored together with the target sentence. This textual database is combined with a terminological database (TDB), which is used by translators to independently store terminological equivalences or translation units of particular value. In their paper the authors outline a first draft of a methodology that describes the preparation of a bilingual terminology from – and within – TM applications. The bilingual corpus produced is called the 'terminological memory' of the translator.

Theme 4: Terminological knowledge bases and ontologies

There are as many terminology resources as there are users, but these resources have at least one common feature: they offer a structured description of the vocabulary used in a specific domain of knowledge. The descriptions may vary according to user specifications (machine or human) and the domains described. In recent years, however, there has been a shift from terminological databases to terminological knowledge bases (TKB). These TKBs are often referred to as 'ontologies', i.e. knowledge repositories in which not only categories (terms) are defined, but also (the) relationships between these categories. In other words, the discipline of terminology is widening its scope towards knowledge representation and knowledge management and toward the semantic web. On the one hand, reflection is therefore needed on how insights in terminology can contribute to the requirements of new semantic web users in a multilingual setting. On the other hand, we need to investigate how terminology management can benefit from new developments in IT and NL processing, and more specifically, how terminographers and terminology management systems can make use of ontologies when creating resources for a variety of user groups.

Pierre Lerat shows that the Web is a very good tool to validate technical translations provided philological precautions are taken. A linguistic method for using the Web consists in making requests on strings of characters as long as possible, between quotation marks, then confronting the obtained results with other sources. It is common knowledge that the Web is

transitory and varies greatly in value; nonetheless using this method will be efficient. Useful requests are made through classes of objects and sets of predicates. Comparing frequencies is a good means of evaluating the respective vitality of concurrent terminological phrases and their social uses (official, technical, professional or journalistic texts). So neologisms and archaisms become traceable.

For **María Teresa Cabré** linguistic engineering is frequently based on a descriptive analysis of terms and their combinations. Tools are created to detect units and structures that can be relevant for knowledge extraction. The integration of all the tools needed by terminologists in a platform enables the transformation of specialised texts and the construction and enrichment of ontologies. In her article Cabré outlines the design of TERMINTEGRAL, an open multimodular platform for terminology work. This platform has been constructed by reusing tools previously developed by IULATERM, a research group of IULA at the Pompeu Fabra University (Barcelona). TERMINTEGRAL enables, on the basis of automatic text processing, the construction of terminological glossaries and ontologies.

Koen Kerremans believes that 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. He reflects on how – in the Termontography approach developed at CVC Erasmushogeschool Brussels – 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 them. The implications of categorisation frameworks for methods in multilingual terminology description are illustrated by frameworks set up in several research projects.

Peter Spyns and Jan De Bo's article provides an introductory overview of the different research domains (computational linguistics, terminography, artificial intelligence (AI), philosophy and database semantics) for which ontologies and the emerging field of the Semantic Web have become a main point of interest. They point out that each of these domains uses a different definition of 'ontology'. The DOGMA ontology engineering methodology is presented with special emphasis on the specific role and contribution of (multilingual) terminography in this ontology. In addition, the authors explain what ontologies can offer to advance linguistics and terminography.

Proper names often constitute a problem in translation. The contribution by **Thierry Grass, Denis Maurel and Mickaël Tran** proposes an ontology which represents the basis for a multilingual database of proper names, Prolexbase. It is being set up for treatment of proper names in the framework of the Prolex project, a research programme supported by the French Ministry of Industry in collaboration with two firms working on the

market of language technologies: Systran and Exalead. The aim of this collaboration is to create a multilingual database of proper names containing information applicable to machine translation, computer aided translation, data research as well as spelling dictionaries. Beside a set of language-dependent and language-independent relations associated with a logical model, the database is founded on a four level ontology: the level of instances, the linguistic level, the conceptual level and the metaconceptual level. The authors describe the different levels of the ontology and their implementation in the database using French and German examples.

The topics covered by the articles in this issue of *Linguistica Antverpiensia* constitute an overview of research carried out in the intertwining subject fields of domain-specific languages, terminology and translation. Not only do we find contributions on new developments in terminology theory and adjacent disciplines, also the impact of computational terminology management and the potentialities of terminological knowledge bases and ontologies for professional translators are demonstrated. The editors are confident that this issue of *Linguistica Antverpiensia* will encourage further research and developments in terminology theory and in methods for terminography.

---

[1] L'Homme, M.-C., U. Heid & J.C. Sager (2003) "Terminology during the last decade (1994-2004)" *Terminology* 9 (2), 151-161