

Terminology and Lexicography Research and Practice (TLRP)

Terminology and Lexicography Research and Practice aims to provide in-depth studies and background information pertaining to Lexicography and Terminology. General works include philosophical, historical, theoretical, computational and cognitive approaches. Other works focus on structures for purpose- and domain-specific compilation (LSP), dictionary design, and training. The series includes monographs, state-of-the-art volumes and course books in the English language.

Editors

Marie-Claude L'Homme
University of Montreal

Kyo Kageura
University of Tokyo

Consulting Editor

Juan C. Sager

Volume 13

Terminology in Everyday Life
Edited by Marcel Thelen and Frieda Steurs

Terminology in Everyday Life

Edited by

Marcel Thelen

Zuyd University, Maastricht

Frieda Steurs

Lessius University College

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Terminology in everyday life / edited by Marcel Thelen, Frieda Steurs.

p. cm. (Terminology and Lexicography Research and Practice, ISSN 1388-8455 ; v. 13)

Includes bibliographical references and index.

1. Language and languages--Terminology. 2. Terminology. I. Thelen, Marcel. II. Steurs, F. (Frieda)

P305.T4437 2010

401.4--dc22

2009045277

ISBN 978 90 272 2337 1 (Hb ; alk. paper)

ISBN 978 90 272 8859 2 (Eb)

© 2010 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

Introduction	1
<i>Marcel Thelen and Frieda Steurs</i>	
SECTION I. Terminology and smaller language	
Synonymy and variation in the domain of digital terrestrial television: Is Italian at risk?	11
<i>Franco Bertaccini, Monica Massari and Sara Castagnoli</i>	
Language (policy), translation and terminology in the European Union	21
<i>Márta Fischer</i>	
The situation and problems of Hungarian terminology	35
<i>Ágota Fóris</i>	
Translation-oriented terminology work in Hungary	47
<i>Judith Muráth</i>	
Towards a national terminology infrastructure –The Swedish experience	61
<i>Henrik Nilsson</i>	
SECTION II. Best practices in terminology management	
Terminology on demand: Maintaining a terminological query service	81
<i>Claudia Dobrina</i>	
Frames, contextual information and images in Terminology: A proposal	97
<i>Mercedes García de Quesada and Arianne Reimerink</i>	
How much terminological theory do we need for practice? An old pedagogical dilemma in a new field	123
<i>Vassilis Korkas and Margaret Rogers</i>	

Ontological support for multilingual domain-specific translation dictionaries	137
<i>Rita Temmerman and Sancho Geentjens</i>	

SECTION III. Possibilities of terminological databases for different applications

In praise of effective export terminology	149
<i>Danielle Dubroca Galin, Ángela Flores Garcia, Valérie Collin Meunier and Marc Delborge</i>	

Computer aided term bank creation and standardization — Building standardized term banks through automated term extraction and advanced editing tools	163
<i>Jody Foo and Magnus Merkel</i>	

Competency-based job descriptions and termontography. The case of terminological variation	181
<i>Koen Kerremans, Peter De Baer and Rita Temmerman</i>	

Proposals to standardize remote sensing terminology in Spanish	195
<i>Lara Sanz Vicente and Joaquín García Palacios</i>	

SECTION IV. Terminology in a medical setting

The PERTOMed project: Exploiting and validating terminological resources of comparable Russian-French-English corpora within pharmacovigilance	213
<i>Cedric Bousquet and Maria Zimina</i>	

Instrumentality in cognitive concept modelling	233
<i>Paul Sambre and Cornelia Wermuth</i>	

Biographical notes	255
--------------------	-----

Author index	263
--------------	-----

Subject index	267
---------------	-----

Introduction

Marcel Thelen and Frieda Steurs

1. General

This volume contains a selection of papers that were presented at the international conference on terminology that was organised by NL-Term (the Dutch-Flemish Association for Dutch Terminology) in close cooperation with the Department of Translation and Interpreting¹ of Lessius University College Antwerp, and under the auspices of the Nederlandse Taalunie (the Dutch Language Union, a cooperation between the Dutch and Flemish Ministries of Education and Culture to promote the Dutch language).

As an association whose aim it is to bring together individuals and institutions working with, doing research on or just being interested in terminology, NL-Term regularly organises symposia on topics of relevance to terminology. This conference is a new initiative of NL-Term and was the first of a series of international conferences on a tri-annual basis.

The conference took place on 16 and 17 November 2006 at Lessius University College in Antwerp (Belgium). The local organisation was in the hands of the Department of Translation and Interpreting of Lessius University College. The title of the conference was “Terminology and Society. The impact of terminology on society”. The conference was preceded on 13 November by a workshop organised by TermNet on terminology planning and policies.

The main aim of the conference was to contribute to the efforts of the European Association for Terminology (EAFI) and other national and international organisations to create platforms for the exchange of information on advances in terminology science and its applications.

The conference had an impressive international audience of participants from professional terminology bodies and companies, to translators and interpreters, and universities teaching terminology. It brought together theory and practice, terminology and translation & interpreting, larger and smaller languages and

1. As of 2006, the Department's name is Department of Applied Language Studies.

various fields of application. The conference together with the pre-conference workshop was preceded by a two-day international conference of EAFT in Brussels. Together, these formed the “semaine belge de la terminologie”, a unique one-week international platform on terminology, initiated by the Department of Translation and Interpreting of Lessius University College.

The fact that two terminology associations, i.e. a bi-national one – NL-Term – and a European one – EAFT – both had an international conference on terminology in one week and that this was made possible for a great part by a university teaching translation and interpreting, viz. the Department of Translation and Interpreting of Lessius University College, together with the large number of international participants at the “Terminology and Society” conference, but also the fact that many participants from the first conference took part in the second, shows that terminology is alive and that terminology needs to be taken seriously as it has a great impact on everyday life, which was precisely the main theme of this conference.

Conceptual systems differ from country to country, from company to company, which may result in confusion of ideas or communication glitches. It is not uncommon for users in different places to coin entirely new terms, derive terms from existing ones whose meanings may or may not overlap, or to attach new meanings to existing terms.

Words are important carriers of meaning and form an essential part of communication. Ambiguous terminology has been known to engender miscommunication, which in its turn may have any number of potential, damaging consequences.

In our increasingly global, international world, effective and unambiguous communication between a motley crew of potential partners has become crucial. Our growing mobility and migratory lives have made intercultural communication a very hot issue indeed, and given it a strategic, political function.

Terminology plays a part in a huge array of communication situations and is used to discuss a vast number of subjects by myriad communication partners.

The international information society has created a huge demand for multilingual technical, scientific and legal documents, and the consequences for translation can hardly be denied.

Translating job specific texts massively stimulates both the language industry and language technology. Even with maximally deployed human resources, the demand cannot be met. A huge amount of translating, localizing and publishing work is waiting to be done. We are set increasingly tight deadlines, while the number of languages into which documents are to be made available keeps on growing.

The world of multilingual communication is a perfect barometer of this. From an *economic* point of view, the market for translation, interpreting, multilingual document management and so-called language services is worth billions. Inside

today's globalization context, it grows exponentially each year. With a global annual growth potential of 30%, the translation and interpreting industry is one of the fastest growing industries in the world. In the United States, the translation industry generated an income of 8.8 billion dollars in 2005. Worldwide, the translation market yields a 30 billion turnover every year. The European Union alone spends €1.1 billion a year on translation and interpreting services. Those costs are minimal, compared to the millions of losses companies and institutions would suffer if they were to opt for monolingualism.

Terminology affects more than just translation. Modern countries keep adjusting their language and translation policies to the new demands of our multilingual, multicultural society. Worldwide, translators and interpreters are deployed in politically and socially delicate areas: in health care, asylum procedures (community interpreting), legal procedures (court interpreting), etc. Opting for a specific translation policy – translation as a fundamental right versus translation as a special favour – is intrinsically linked with whether or not a community grants equal social, economic, political, linguistic opportunities and rights to minorities in today's multicultural society.

In each of those activities – translating the European legislation, starting a file on bio-energy, preparing an international congress on diabetes, or interpreting for the International Court of Justice in The Hague – terminology plays a pivotal part. Without specialized terminology and correctly understood concepts inside a specific field, correct communication becomes impossible.

Interesting in this respect, is the question of how so-called 'small languages' deal with this phenomenon. They often feel pressurized by the large international markets, while globalization often – wrongfully – gives the impression that we are moving towards a highly monolingual (read: English-language) market.

Nothing could be further away from the truth: increased globalization and internationalization actually boost 'localization'. The more global a company goes, the more individually its local markets must be approached. Companies that fail to respect their customers' language and culture score badly on the marketing policy board and will feel the consequences.

Of old, the European Commission has focused firmly on multilingualism, from the core conviction that all member state languages are equal – a very important stand indeed, and crucial to the European policy. If all citizens of the European Union's member states are to have equal rights, those same citizens also have a right to communicate in their own language. Linguistic identity is essential, but there is more: all Europeans must be able to read what laws have been enacted, what procedures followed by the European policy makers.

That approach of multilingualism has yielded a terminology infrastructure in Europe in the form that revolves around one European Association for Terminology

(EAFI) and national terminology associations such as NL-Term in the Netherlands and Flanders. Important, on a policy level, is the fact that those infrastructures receive the support of their national governments.

2. Contributions in this book

The contributions in this book all deal with terminology and in particular its central role in society, which was the theme of the international conference “Terminology and Society. The impact of terminology on everyday life”.

Some papers were written as research papers (e.g. Bertacinni and Massari), while others present research projects in their initial stages (e.g. Garcia and Reimerink). Again others are of a more general nature (e.g. Fóris). What all these papers have in common, however, is that they demonstrate that terminology is important for everyday life in many respects. The book consists of 4 main sections, each dealing with a different aspect of the central theme.

Section I, terminology and smaller languages, is the largest section and contains five papers. The first paper by Bertacinni and Massari (*Synonymy and variation in digital terrestrial television: is Italian at risk?*) deals with a new system of broadcasting by television, Digital Video Broadcasting over Terrestrial (DVB-T), that, because it is interactive, allows the user to become an active user instead of a passive consumer. The authors describe an investigation into the terminology of this new technology in Italian and French, not typically smaller languages, but nevertheless smaller in comparison to English. They found that Italian and French still lack a sufficient terminology for this field, and that English is used as the lingua franca, making it possible that the technical knowledge about the field is circulated internationally, but at the same time creating a barrier for those people who are not proficient enough in English. Fischer (*Language (policy), translation and terminology in the European Union*) discusses the translation and terminology work at the European Union, in particular the EU policy on translation and terminology and their impact on the work done. She distinguishes between “multilingual term creation within one conceptual system” (viz. the EU conceptual system) and “term transfer between different systems” (viz. the conceptual systems of member states). EU-level terms are first created in a limited number of languages (primary term creation), and only then through translation in other languages on the basis of these primary terms (secondary term creation). In her view, translation plays a role in “multilingual term creation” only, and not in “term transfer” because here only existing equivalents have to be found. Because of this way of working, EU-level translations may have an impact on the official languages of the member states. Fóris (*The situation and problems of Hungarian terminology*) argues

for the need for Hungarian terminology to develop quick and well-founded terminological classifications, since the rapid scientific and technological developments alongside with economic and political changes going on in Hungary and the broadening of conceptual systems of a number of areas (such as services, administration and education) make this necessary for lack of Hungarian counterparts. In this paper she describes the current state of Hungarian terminology and the problems that have to be solved. **Muráth** (*Translation-oriented terminology work in Hungary*) deals with the same topic, i.e. terminology in Hungary, though from a different perspective. She describes how in particular bilingual and multilingual terminology in Hungary was neglected as a discipline of linguistics for the last 20 years, and that now translation-oriented terminology is taking over. In her paper she outlines the history of this type of terminology in Hungary including the parties involved in its development and then discusses a number of examples and problems faced on the basis of examples from the social sciences. In doing this, she compares Hungarian and German. The last paper in this section is by **Nilsson** (*Towards a national terminology infrastructure – the Swedish experience*). Nilsson gives an outline of the Swedish Centre for Terminology (TNC) and describes its programme, TISS – Terminology Infrastructure for Sweden. Two of the main components of TISS are a terminology portal containing a national term bank – the “Rikstermbanken”, and a terminology coordination programme. In this respect, TISS resembles the Dutch-Flemish terminology policy of the Dutch Language Union (Nederlandse Taalunie). TISS is an adequate answer to problems that smaller languages face as regards their language and their national terminology.

The theme of **Section II** is best practices in terminology management from different points of view. It consists of four papers. The first one, by **Dobrina** (*Terminology on demand: maintaining a terminological query service*), describes the terminological query service of the Swedish Centre for Terminology (TNC). One of its most interesting features is that this service includes “terminology on demand”, a query system where tailor-made terminology is called for. The system also handles terminology-related questions. Dobrina gives a detailed description of the procedures of this service, but also of the demands posed on the staff. The TNC seems, thus far, to be unique for this query system. The second paper in this section, *Frames, contextual information and images in terminology: A proposal* by **Garcia de Quesada** and **Reimerink**, analyses and evaluates a number of possible applications of the Berkeley FrameNet project – originally put forward by Fillmore et al. (2003) – and the Spanish FrameNet, to the area of coastal engineering. In their proposal, they group together images and texts on coastal engineering in accordance with these “Fillmorian” frames. In this way, they are able to draw conclusions about the syntactic contexts of the various terms, the semantic characteristics of the contextual items of these terms, and the place of the terms in a network

of related terms. On the basis of these findings, they give a number of suggestions for the actual presentation of data to the end user of a term base on coastal engineering. The third paper in this section, by **Korkas and Rogers** (*How much terminological theory do we need for practice? An old pedagogical dilemma in a new field*), deals with what terminology training as part of a (postgraduate) translation programme should/could entail. At the heart of the problem is the question what should/could be the ideal division between theory and practice in terminology training. Before answering this question, the authors dwell on the questions of what can be understood by theory, and what elements of this theory should be taught. These are the questions that every terminology lecturer has to find an answer to. They propose to include at least the term-word distinction, and the format of definitions and contexts. Then they discuss how theory could be linked to practice. The core of this link, they suggest, is problem solving. The fourth and last paper in this session, (*Ontological support for multilingual domain-specific translation dictionaries*) by **Temmerman and Geentjens**, proposes a multilingual terminological translation dictionary where the terminological content makes use of knowledge representation for the description of terms. As a good example of such a dictionary the authors outline the dictionary produced by Dancette and Réthoré (2000), *Dictionnaire Analytique de la Distribution. Analytical Dictionary of Retailing*. For such a dictionary that includes knowledge representation, they use the termontography method developed by Temmerman, which uses ontological models for the representation of terminological information. The field analysed is the automotive field and the languages studied are English, French, German, Dutch and Italian.

The theme of **Section III** is exploring possibilities of terminological databases for different applications. It contains four papers. The first one, by **Dubroca Galin, Flores Garcia, Collin Meunier, and Delbarge** (*In praise of effective export terminology*), presents a project on export terminology. The project aims at describing the terminology used to market local products in their countries of origin and the terminology transfers in translated international trade documents. The authors describe such examples as the Spanish Armuña lentils and the French foie gras and pralines. The project outlines the transfers from Spain to French-speaking European countries and vice versa. The project shows that promotional terminology is an extremely important but difficult type of terminology since it changes constantly, unlike other types of terminology. The next paper in this section, by **Foo and Merkel** (*Computer aided term bank creation and standardization – Building standardized term banks through automated term extraction and advanced editing tools*), argues that for authoring and translating, standardised term banks are indispensable in terms of terminological consistency and avoidance of confusion and frustration on the part of the audience. To fill such standardised term banks

requires a lot of manual labour. Term extraction may reduce this, but even then a lot of manual post-extraction work is needed. To overcome this and reduce the amount of manual labour even further they propose a method that combines efficient editing tools (alignment and alignment-related tools) and the criterion of quality or relevance of extracted candidate terms. In this way, they will be able to detect terminological inconsistencies and to build a standardised term bank. Such a term bank can be a good instrument for quality assurance. **Kerremans, De Baer, and Temmerman** (*Competency-based job descriptions and termontography. The case of terminological variation*) continue this section and deal with competence management as part of HRM in large companies. Competences play an important role in job descriptions, job planning and staff evaluation. Consistency and standardisation of competences and their descriptions are essential especially since more and more competences and competence-based job profiles are communicated by automatic exchange through a semantic web environment. Competences typically include knowledge, skills and attitude. They discuss a project, PoCeHRM, that tries to reduce terminological confusion over competences by standardisation. For this they make use of termontography, which is a combination of socio-cognitive terminology and ontology engineering. The last paper in this section is *Proposals to standardize remote sensing terminology in Spanish* by **Sanz Vicente, and Garcia Palacios**. The language of investigation in this paper is Spanish. In the area of remote sensing, Spanish has not been able to catch up with English as regards the terminology of remote sensing, with the result that Spanish is full of English terms. In this paper, the authors propose a methodology for the creation and standardisation of Spanish terms for the remote sensing domain. This methodology is the result of close cooperation between terminologists, linguists, and domain experts. It is based on communicative linguistics and corpus-based terminology and aims at (1) facilitating communication, and (2) help translators and interpreters with text comprehension and text production in the domain.

Section IV deals with terminology in a medical setting, and contains two papers only. The first of these, *The PERTOMed Project: exploiting and validating terminological resources of comparable Russian-French-English corpora within Pharmacovigilance* by **Bousquet and Zimina**, discusses a multidisciplinary project by a number of institutions in France, PERTOMed, designed to study a number of areas including pharmacovigilance, i.e. “the collecting, monitoring, researching, assessing and evaluating information from healthcare providers and patients on the adverse effects of medications”. Terminology plays a crucial role here. One of the objectives of the project was to build a specialised internet corpus (in Russian) on pharmacovigilance. In doing this, the authors tested a number of term extraction tools that were applied to comparable multilingual texts. The result of the project is a trilingual Russian-French-English database freely available on the

PERTOMed server. This section is concluded by a paper by **Sambre and Wermuth** (*Instrumentality in cognitive concept modelling*). The central theme of this paper is the associative relation of instrumentality in multidimensional terminological definitions. The field of investigation for this paper was the medical subdomains of microsurgery and cardiosurgery, and the data were assembled in a German-French multidisciplinary corpus. Instrumentality is an important concept in medicine, in particular in the two domains of investigation. As an associative relation, instrumentality is defined only rather superficially in linguistics and terminology, especially its syntactic, grammatical and lexical heterogeneity. The authors argue that an approach in line with cognitive grammar yields more detailed descriptions capable of handling instrumentality more satisfactorily. In their paper they outline the investigation done and summarise its results. They conclude by a number of recommendations both for terminology and terminological software tools.

SECTION I

Terminology and smaller language

Synonymy and variation in the domain of digital terrestrial television

Is Italian at risk?

Franco Bertaccini, Monica Massari and Sara Castagnoli

This article presents the results of a bilingual (French-Italian) corpus-based terminological study on the language of digital terrestrial television (DVB-T). The research sets out to describe the main features of this specialized language from a socio-terminological perspective by assessing the role played by English and by investigating synonymy and variation within the domain-specific discourse in the two countries. The analysis revealed that both Italian and French DVB-T terminologies include a large number of English borrowings and are marked by a proliferation of synonyms and variants, probably for lack of standardization in this new developing domain. The existence of equivalences among synonymic forms in the two languages is used to suggest that a new generation of terminological databases should be developed which allow for cross-linguistic correspondences to be established beyond the main-term level.

Keywords: specialized language, socio-terminological approach, lexical borrowings, synonymy, variation, cross-linguistic equivalence, digital terrestrial television

1. Introduction: Motivations and methodology

In recent years, Digital Video Broadcasting -Terrestrial (DVB-T) has often been presented as a new communication medium which will change the traditional notion of television as a simple tool for broadcasting contents. DVB-T is an innovative system for the digital transmission of television signals which paves the way for a new television era, characterized by multimedia contents and interactivity and bound to turn passive TV viewers into active users.

The evolution of DVB-T technology is the main motivation for the present terminological study, which aims at investigating which linguistic phenomena characterize a still-developing domain like digital terrestrial television. Our hypothesis is that the DVB-T domain does not yet have a clear and fixed terminology,

and that in the general ‘chaos’ that is the language of telecommunications, English plays the role of *lingua franca*.

The research, ultimately aimed at producing a bilingual Italian-French term-base on the domain of digital terrestrial television, is based on comparable specialized corpora of Italian (about 500,000 words) and French (532,000 words). In order for the corpora to be representative of the language used in different communicative settings, i.e. by different types of users and in different contexts, we have included various text types, from technical manuals to informative texts like user guides. A third, smaller corpus of English texts (about 300,000 words) was also assembled as a reference resource.

In the following sections we set out to analyse some of the main linguistic features of the domain of digital terrestrial television following a socio-terminological approach, focusing first on the permeability of the Italian language to English words and on the role played by the latter in communication among different types of users, then on the extent to which this specialized language is characterized by synonymy and variation.

2. The language of DVB-T: Main features and the notion of ‘state of necessity’

The language of DVB-T is not an independent jargon in its own right but forms part of the wider and better-known language of Information & Communication Technology (ICT) and, in particular, of the language of the Internet.

In spite of this linguistic dependency, however, the specialized language of ICT cannot fully satisfy all the communication needs generated by the birth of new devices and technologies in the digital terrestrial television domain. From a linguistic point of view, the introduction of a new technology implies the creation of new expressions which help define the new concepts. In other words: the appearance of a new technology determines the rise of a ‘state of necessity’, i.e. the need to create terms, or borrow them from another language, in order to enable communication in the new domain. Language possesses several means and resources to name new concepts, objects and activities. In the domain under examination, the research has highlighted three main trends for Italian: re-semanticization (or terminologization),¹ calque and lexical borrowing.

In the following sections we will concentrate on the latter phenomenon, which results in the widespread presence of English words in this specialized language and which may eventually lead to the gradual disappearance of Italian from this domain.

1. The term refers to the process by which a general-language word or expression is transformed into a term in a special language (ISO 1087-1: 2000).

2.1 Lexical borrowing

The use of English words is a widespread phenomenon in this field. Out of 311 Italian terms extracted during the research,² we identified 93 borrowings (i.e. 29% of the total): according to Haugen's (1950) best-known taxonomy of borrowed items, 76 of these are loanwords (i.e. form and meaning are copied completely), and 17 are loanblends (i.e. terms consisting of a copied part and a native part).

This state of affairs is probably the result of DVB-T technology's origins in English-speaking countries. As pointed out by Sobrero,

[l]a dinamica tradizionale prevede che il paese più innovatore, e all'avanguardia in un determinato settore, esporti negli altri paesi, insieme al know how, anche la parte più importante del lessico specialistico [...] relativo a tale settore. (2002: 268)

The high presence of loans may therefore be due, in the first place, to the important role played by English in the diffusion of the technology. However, it might also be connected to what we previously defined as a 'state of necessity': it can be hypothesized that, this new technology having been exported so fast, 'target' languages such as Italian did not have enough time to elaborate their own terminologies. In effect, borrowings normally merely form the first step in the process of acquisition of a new technology, since at a later stage the recipient language activates its own linguistic resources to create an ad hoc terminology, often inspired by the donor model. Interesting examples of synonymic pairs formed by an English loanword and an Italian term extracted from our Italian corpus are *transport stream* – *flusso di trasporto*, *payload* – *carico pagante*, *zapper* – *ricevitore di base* and *smart card* – *carta intelligente*.³

As we have just observed, the general tendency is for languages to borrow words from English as the need arises, and then supplement them with native terms. Terminologies often continue to coexist, even though with differences in the relative frequencies of use, and thus give rise to a wide range of synonyms and variants. Some forms may eventually prevail or fall into disuse, but this is not about to happen in the domain under examination.

Our research has revealed that the presence of English words is significantly higher within the terminology related to the new digital medium than in the terminology related to the underlying mathematics and engineering. Some subdomains, in other words, turn out to be more influenced by English than others. For instance, English is often used in the subdomain of the new services and applications offered by DVB-T technology, especially with respect to the MHP platform and the set-top

2. Lists of candidate terms were obtained by making word- and cluster lists with *WordSmith Tools*, and then presented to domain experts for validation.

3. Here and in the rest of the article italics are used to mark domain-specific terms.

box; the language used to describe the processes of TV signal transport and reception, on the other hand, appears less contaminated. Much depends, in other words, on the relative recentness of the different subdomains, i.e. the terminology belonging to the first subdomain may still be unstable whereas the second, in use in other subdomains for years, may be more deep-rooted and fixed.

Let us close this section with a look at the borrowing behaviour of French in this domain. French seems to resort to borrowings to a lesser extent than Italian: out of 339 extracted terms, we have identified 57 borrowings, namely 33 loanwords and 24 loanblends. With respect to the DVB-T domain, French tends to create new expressions rather than adopt foreign words, possibly in line with the country's linguistic policy of preservation of the national language.

3. Synonymy and terminological variation

One of the cornerstones of traditional terminology is the so-called 'univocity principle', according to which only one term should be assigned to a concept and vice versa. The principle is thought to ensure effective and efficient communication, whereas its violation is perceived as a source of ambiguity.

In the last decade, however, the univocity principle has repeatedly been questioned (see e.g. Temmerman, 1997). Cabré, for instance, commented that:

[l]a théorie veut qu'en terminologie chaque concept soit exprimé au moyen d'une seule dénomination, mais, une fois de plus, la réalité nous oblige à admettre l'existence de dénominations concurrentes pour une seule notion. On peut ainsi dire que deux unités sont synonymes quand elles désignent le même concept. (1998: 188)

A number of scholars have started advocating the need to acknowledge that synonymy and variation do not belong exclusively to general language but also characterize specialized terminology; this would appear particularly true for domains which are subject to profound changes, where harmonization only occurs in ideal cases and concurrent terms continue to coexist (Mayer, 2002: 118).

This idea is indeed confirmed by the analysis of the specialized language of digital terrestrial television in Italy, where the introduction of this new technology is yet to be matched by the elaboration of an adequate specialized language, and the lack of terminological standardization has led to a proliferation of synonymic forms and variants. Based on the analysis of our Italian corpus, out of 311 Italian terms we detected 76 synonyms and 59 variants. This has led to the creation of

135 terminological records dedicated to non-main terms, i.e. there are about as many synonymic forms and variants as main terms in the database.⁴

The high number of synonyms is due, above all, to the simultaneous presence of foreign and Italian terms designating the same concept. In addition, these are often paralleled by acronyms and abbreviations which further increase the range of available synonymic expressions for the same concept. Some significant examples are given in Table 1 below.

As can be observed from the examples in Table 1, the proliferation of synonyms may be linked to the appearance of multiple translations for the same English term, which then coexist with the original. For example, *intestazione* and *testa* are both possible translations of the English term *header*, and the three forms are used in the corpus.

A similar tendency can be observed in the French corpus: out of 339 French entries, we counted 71 synonyms and 92 variants. French thus shows more variants (92) compared to Italian (59): this is arguably due to the tendency for French to coin native acronyms or abbreviations, such as *TEB* for *BER* (*Taux d'Erreurs sur les Bits* – *Bit Error Rate*), *IES* for *ISI* (*Interférence Entre Symboles* – *Intersymbol Interference*), *TNT* for *DTT* (*Télévision Numérique Terrestre* – *Digital Terrestrial Television*), etc., which are used alongside the English ones in technical communication.

Table 1. Co-existence of Italian terms, foreign terms and acronyms/abbreviations

Main term	Synonym(s)	Variant(s)
<i>Guida elettronica ai programmi</i>	Electronic programme guide	EPG
<i>Interfaccia di programmazione delle applicazioni</i>	Application programming interface	API
<i>Intestazione</i>	Testa Header	
<i>Macchina virtuale Java</i>	Java virtual machine	JVM
<i>Piattaforma multimediale domestica</i>	Multimedia home platform	MHP Piattaforma MHP

4. Two main criteria were used to determine the relative status of terms. First, we decided that acronyms and abbreviations were always to be treated as variants, even when more frequent in the corpus than full lexical forms (in which case the information was recorded within terminological records). On the other hand, frequency was used as the main criterion to decide whether the status of main term should be assigned to native Italian terms or to the corresponding English loanword.

3.1 'Physiological' and 'pathological' synonymy

A deeper analysis of the above phenomenon has revealed the existence of two types of synonymy, linked to different causes. More precisely, by adopting a socio-terminological point of view, i.e. by focusing on the actual uses of the language by different users, it may be argued that the DVB-T terminology is affected simultaneously by 'physiological' and 'pathological' synonymy (see Bertaccini et al., 2006 on these notions). The former consists of a functional kind of synonymy, originating in diastatic differences, whereas the latter, as its name suggests, is arbitrary and may cause confusion.

The two types of synonymy do not present the same distribution within the analysed corpus. Unjustified 'pathological' synonymy is more frequent than functional synonymy, given that the recent development of the domain has not yet allowed the standardization of terminology both within and across different user groups.

Nevertheless, the research has highlighted some cases of physiological synonymy which deserve to be analysed. An interesting example is the series of terms denoting the device which connects the television to an external signal source, i.e. the so-called set-top box. Our terminological research extracted twelve different terms from our Italian reference corpus, either synonyms or variants, used to indicate the device in question, namely *ricevitore DVB-T*, *decoder*, *adattatore digitale*, *ricevitore DTT*, *ricevitore digitale terrestre*, *decodificatore per DTT*, *decodificatore digitale*, *set-top-box*, *STB*, *STB IRD*, *set top box IRD*, *STB integrated receiver decoder*. The long list of terms may initially point to an unjustified proliferation of synonyms, but some of them did at least have a functional role. *STB IRD*, for example, is the term preferred by organizations which carry out research in the legal, technological and marketing domains, such as the Ugo Bordoni Foundation in Rome. Conversely it is not used in business settings, where the terms *ricevitore*, *decoder* and *set-top-box* are favoured; these are also the most common among aerial fitters, retailers and shopkeepers.

On the other hand, as mentioned in previous sections, pathological synonymy is often connected to the coexistence of foreign and native terms, which may be caused by different translation choices and gives rise to a wide range of equivalent expressions and acronyms/abbreviations. Let us take the following list as an example: *televisione interattiva*, *profilo televisione interattiva*, *profilo della trasmissione interattiva*, *profilo interactive broadcasting*, *radiodiffusione interattiva*, *interactive broadcasting*. The list comprises one loan word (*interactive broadcasting*), a loanblend (*profilo interactive broadcasting*) and a series of structural calques from English (i.e. *profilo televisione interattiva*, *televisione interattiva* and *radiodiffusione interattiva*). The use of this type of synonyms being arbitrary, it may represent an obstacle during the creation of terminological databases, and puzzle non-expert users.

4. Cross-linguistic equivalences

As previously mentioned, our analysis has revealed that a comparable proliferation of synonyms and variants occurs in the Italian and French corpora. In addition, the analysis has highlighted the existence of a network of relationships between the terminologies in the two languages, not only at the level of main terms, but also with respect to synonymous forms: in other words, when comparing the Italian and French glossaries, cross-linguistic equivalences are also detected between synonyms and variants.

Cross-linguistic equivalences are a key issue for terminological research, and particular attention is being given to the possible creation of new databases where users can perform cross-searches even between synonyms and variants in different languages. Such an instrument would not only meet different user needs by offering them the possibility to consult the database multidirectionally, it would also allow translators to detect and get information on the correct equivalent for any given term presented as a synonym or variant in the database. Whether real equivalence can be found between synonyms and variants in different linguistic repertoires remains a matter of debate: there is hardly any equivalence between main terms, let alone at the level of variants. 100% correspondences may be rare between synonymous forms across languages, but some examples of full equivalence do exist, e.g. in the case of loanwords from English and of acronyms/abbreviations. In those cases, cross-linguistic correspondence is borne out by the terms' morphological identity.

In what follows, we will try to provide evidence of the existence of cross-linguistic equivalences between Italian and French by using the formal criterion of isomorphism, i.e. by analysing in detail the morphological structure of the terms listed in Table 2.

First of all, as mentioned above, in many cases the English abbreviation exists in both Italian and French, e.g. *DVB-T*, *MHP*, *PID*. In these cases, there is full morphological correspondence. Then there are these cases in which an Italian acronym or abbreviation (borrowed from English) corresponds to two French acronyms/abbreviations, one of which is a loan-word from English while the other is derived from the native French main term: for instance, *BER* – *BER/TEB*, *JVM* – *JVM/MVJ*.

Other cases of morphological equivalence are the phrasal forms following the 'noun + abbreviation' pattern, e.g. *piattaforma MHP* – *plateforme MHP*, *rete SFN* – *réseau SFN*. Isomorphism also occurs between phrasal forms which present the same structure in both languages, as in the following examples: *controllo di accesso* – *contrôle d'accès*, *sistema di accesso condizionato* – *système d'accès conditionnel*,

Table 2. Cross-linguistic multi-level equivalences

	Main term	Synonym(s)	Variant(s)
IT	<i>Accesso condizionato</i>	<i>Controllo di accesso</i>	<i>CA, Sistema di accesso condizionato, Accesso condizionale</i>
FR	<i>Accès conditionnel</i>	<i>Contrôle d'accès, Système de contrôle d'accès</i>	<i>AC, Système d'accès conditionnel</i>
IT	<i>Tasso d'errore sul bit</i>		<i>BER, Tasso di errore</i>
FR	<i>Taux d'erreurs sur les bits</i>	<i>Taux d'erreur binaire</i>	<i>TEB, BER, Taux d'erreur, Taux d'erreurs</i>
IT	<i>Televisione digitale terrestre</i>		<i>DTT, DVB-T, Digitale terrestre, TV digitale terrestre</i>
FR	<i>Télévision numérique terrestre</i>	<i>Télévision numérique de Terre, Télévision hertzienne terrestre, Télévision numérique hertzienne</i>	<i>TNT, TVNT, DVB-T, TV numérique terrestre</i>
IT	<i>Smart card</i>	<i>Carta intelligente</i>	
FR	<i>Carte à puce</i>	<i>Carte intelligente</i>	
IT	<i>Piattaforma multimediale domestica</i>	<i>Multimedia home platform</i>	<i>MHP, Piattaforma MHP</i>
FR	<i>Plate-forme multimédia domestique</i>		<i>MHP, Plateforme MHP, Plateforme multimédia domestique</i>
IT	<i>Macchina virtuale Java</i>	<i>Java virtual machine</i>	<i>JVM</i>
FR	<i>Machine virtuelle Java</i>		<i>MVJ, JVM</i>
IT	<i>Rete a frequenza singola</i>	<i>Rete isofrequenziale</i>	<i>Rete SFN</i>
FR	<i>Réseau monofréquence</i>	<i>Réseau à fréquence unique</i>	<i>Réseau SFN</i>
IT	<i>Identificatore di pacchetto</i>		<i>PID</i>
FR	<i>Identificateur de paquet</i>		<i>PID</i>

tasso di errore – taux d'erreur, TV digitale terrestre – TV numérique terrestre, carta intelligente – carte intelligente.

The possibility to establish cross-linguistic equivalences at any level within termbases would allow users to identify the equivalent for a given term irrespective of its form.

5. Conclusion

The results of our research suggest that the linguistic panorama related to the new DVB-T technology is still evolving. The delay in the elaboration of ad-hoc terminologies has been shown to favour the use of English terms in both scrutinized recipient languages. This may enable the circulation of knowledge at an international level, by electing English as lingua franca. The use of foreign terms, however, may constitute a barrier for people who fail to reach a certain level of competence. In this case, English may create a great disparity between users, increasing the so-called 'digital divide'. On a sadder note, if the frequent use of foreign words increases the passiveness of a language with respect to the creation of new terminologies, thus leading to its impoverishment, as some authors suggest, this may result in the gradual disappearance of the Italian language from this specialized domain.

References

- Bertaccini, F., Prandi, M., Sintuzzi, S. and Togni, S. (2006). Tra lessico naturale e lessici di specialità: la sinonimia. In R. Bombi, G. Cifoletti, F. Fusco, L. Innocente and V. Orioles (eds.), *Studi linguistici in onore di Roberto Gusmani (171–192)*. Alessandria: Edizioni dell'Orso.
- Cabré, M.T. (1998). *La terminologie. Théorie, méthode et applications*. Ottawa: Les Presses de l'Université d'Ottawa.
- Haugen, E. (1950). The analysis of linguistic borrowing. *Language*, 26, 210–231.
- ISO 1087–1 (2000). *Terminology work – Vocabulary – Part 1: Theory and application*. Geneva: International Organization for Standardization.
- Mayer, F. (2002). Sinonimia ed equivalenza. In M. Magris, M.T. Musacchio, L. Rega and F. Scarpa (eds.), *Manuale di terminologia. Aspetti teorici, metodologici e applicativi (115–133)*. Milano: Hoepli.
- Sobrero, A.A. (2002). Lingue speciali. In A.A. Sobrero (ed.), *Introduzione all'italiano contemporaneo. La variazione e gli usi (237–276)*. Laterza: Roma.
- Temmerman, R. (1997). Questioning the univocity ideal. The difference between socio-cognitive Terminology and traditional Terminology. *Hermes, Journal of Linguistics*, 18, 51–91.

Language (policy), translation and terminology in the European Union

Márta Fischer

The aim of this paper is to analyse the impact of an EU language and translation policy and the peculiarities of translation and term-creation in an EU context. I first present the legal basis for a language and translation policy, focusing on national-level involvement in corpus planning, and, secondly, I analyse how institutional translation policy and practice affect the linguistic (language, text) and extra-linguistic (translator, receiver) aspects of translation. Thirdly, by revealing the complexity of conceptual systems at an EU-level and extending the model of Sager (1990), I analyse the stages of multilingual primary and secondary term-creation in an EU context. In the creation of secondary terms I draw a distinction between *intra*-conceptual and *inter*-conceptual term-transfer. I finally outline the linguistic and conceptual variability of EU terms.

Keywords: language policy, translation policy, status planning, corpus planning, primary term formation, secondary term formation, multilingual primary term formation, *intra*-conceptual term-transfer, *inter*-conceptual term-transfer, EU terminology

1. Introduction

In discussions about language policy issues, increasing importance is being attributed to the far-reaching impact of translation in an international context and to the notion of lesser translated languages (Branchadell & Lovell, 2005). The European Union, itself a mixture of both intergovernmental and supranational elements, forms a peculiar case in point. Due to the direct effect of Community legislation and the principle of democracy, the language regime is multilingual and translation into all official languages mandatory. The constant obligation for all official languages to create a special language and terminology for the EU domain is of outstanding importance in the light of serious concerns about the threat of domain loss in some languages. In an EU context, translation into and from the official languages ensures by regulation the ‘EU function’ of those languages,

producing many thousands of pages every year. Apart from those quantity aspects, however, it seems worth taking a look beyond the legal basis and to analyse both this and the impact of translations into those languages.

2. Language policy and translation

The brackets used in the title indicate the lack of any explicit legal reference to a coherent policy area for language issues in either primary or secondary legislation. However, if, in general, we understand by language policy top-down, deliberate efforts to influence language use, whether explicitly codified or not, an implicit policy may be deduced, led by the principles of equality and diversity. This is further underlined by the European Commission's very first communication on a framework strategy for multilingualism in 2005 and its second communication on multilingualism published in 2008¹. In both communications, multilingualism is referred to as a policy area, in line with the principle of subsidiarity. Similarly, in spite of the absence of any explicit reference to a translation policy, the management of translations, costs, logistics and translator training constitutes the elements of a translation policy at institutional level. Since translation activity at EU-level, unprecedented in any other international organisation, is a direct result of the principle of equality, it seems necessary to analyse the legal basis of this principle and the competences of the European Union.

In order to trace the competences of the EU in the field of language, three communication levels need to be differentiated at EU-level – and competences accordingly: firstly, communication *within* EU institutions; secondly, communication between EU institutions and European citizens; and, finally, communication *among* the citizens of the EU. The first and second levels cover the institutional level of the EU whilst the third refers to the European citizens' level. Starting with the latter, no legislative competence is conferred upon the EU to define what languages European citizens should use or learn, since education (including language teaching) is not a harmonized policy. Measures may be complementary to Member States' actions in order to promote their citizens' language skills in the form of Community programmes, actions and initiatives or recommendations made by the European

1. Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions: A new framework strategy on multilingualism. COM (2005) 596; Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee and the Committee of the Regions. Multilingualism: an asset for Europe and a shared commitment. COM (2008) 566

Council². Conversely, competence was conferred upon the Council of the European Union by the Treaty to regulate language use at an institutional level.³ Accordingly, the Council set down the linguistic regime in its very first Regulation in 1958, defining the official languages and working languages on an equal footing. Below we shall discuss the implications of the Regulation from the point of view of language planning at EU and national levels in more detail. (Fischer, 2007)

Status and corpus planning (Kloss, 1969) gain relevance if we analyse national involvement in language planning at an EU-level. Status planning refers to the selection and spread of a language, by assigning it a social and legal status, whilst corpus planning refers to the codification and elaboration of the language (including terminological modernisation) in order to make the language suitable for every function. In the EU context, some elements of language planning are transferred to the EU-level, whilst others remain within the national competence. To start with, status planning is carried out by the Council of the European Union. The decision here is not about selecting a vernacular to be the standard language, but to give an additional EU status to the language(s) officially recognized at a national level⁴. As a result, Member States define the EU status of their official languages, since they act unanimously in the Council in this matter. The question then becomes how corpus planning is linked to status planning at EU-level.

At first it may seem strange to speak about corpus planning for languages which have long been recognized and standardized at a national level. However, we must bear in mind that, not least due to the supra-national nature of the Union, a new domain has emerged where official languages have to fulfil new functions. As I shall argue below, especially from a legal point of view, a specific language, an 'EU language' is called for and to be developed in order to allow for a clear delimitation from national regulations. In addition, new terms are constantly being introduced, for

2. In 2002 the heads of states and governments called for further action to improve the mastery of basic skills, in particular by teaching at least two languages from a very early age and establishing a linguistic competence indicator. Presidency Conclusions of the Barcelona European Council. 15 and 16 March 2002.

3. Article 290 of the Treaty establishing the European Community

4. Although the framework of the present paper does not allow us to discuss this decision in more detail, it is important to highlight the shortcomings of the equality principle. From the perspective of all languages spoken on the EU territory, the equality principle is reduced to the languages enjoying official status at national level. Hence, the (mostly monolingual) language policies of the MEMBER STATES were expanded to the EU-level. This *de jure* inequality of the languages in the EU is counterbalanced by actions promoting regional and minority languages, by funding programmes available for all languages and by the Conclusions of the Council (on 13 June 2005) on the use of additional languages at institutional level with costs to be borne by the requesting Member State.

which equivalents in all official languages have to be provided. Hence, corpus planning plays a crucial role in ensuring the fulfilment of this new function.

This seems all the more important since, contrary to status planning, national involvement in corpus planning is rather limited at EU-level. Before accession, the *acquis communautaire* was translated in the Member States, but after joining the Union, the translation of documents (including the development of EU terminology) was transferred to the EU institutions. As a result, after EU accession, corpus planning is carried out in the EU institutions and not by or in national authorities or translation centres. The national level may contribute indirectly by delivering highly skilled translators to the EU or, more directly, by formalizing co-operation between EU translators and experts at home. In any case, the limits of national involvement in corpus planning for the official language(s) make co-ordinated terminology work both at EU and national levels especially important. This is increasingly felt inside the EU and in some Member States in view of the quality of translations and terminology.

3. Translation policy and translation

At an institutional level, the practical implementation of the equality principle is ensured by the translation services of the EU institutions, each establishing their own practice and policy. Before turning to the process of translation in the EU institutions, it is important to note that, from a legal and status point of view, there are, theoretically, no translations – only different language versions, all being equally authentic. From a dynamic perspective, however, the equality principle would presuppose parallel, multilingual drafting. With twenty-three languages this is not feasible in practice, so in most cases documents are drafted in one, two or three languages and then translated into others. As a result, the *de facto* inequality of the official EU languages makes some languages act as source languages whilst other languages only serve as target languages in translation. Source languages are generally the so-called procedural languages of the institutions used for internal communication.⁵

5. In general discourse, those internal languages are referred to as ‘working languages’. However, in Regulation 1958 no distinction was made between the terms ‘official languages’ and ‘working languages’, but considerable liberty is left to the institutions to set their own language use in their rules of procedures. Therefore, those languages, generally referred to as ‘working languages’ are in fact the procedural languages of the institutions. In this paper, adhering to the Regulation’s terminology, we use the term ‘official languages’ for the official languages and working languages as laid down in Regulation 1958, while with the term ‘procedural languages’ we refer to the languages used in the institutions for internal communication.

Dominating languages, institutional practice and the peculiarity of the EU decision-making process affect both the linguistic (language, text) and extra-linguistic (translator, receiver) aspects (Klaudy, 2004) of translations⁶. First of all, the source languages in which documents are drafted are mostly English, French and to a lesser extent German; that is, the procedural languages of the European Commission. This is all the more understandable, as the Commission is the initiator of legislative acts, and so most legislative documents are 'born' here. However, it is rather difficult to trace the original language of texts since they may be drafted in more than one language. Alternatively, the language may be affected by interference from other languages, since texts are drafted by officials whose mother tongue is not always the language of drafting. Similarly, source texts are modified at several steps in the decision-making process and likewise translated by different translators. It is important to note that those aspects also make research in parallel corpora or terminology difficult, since the notion of source language and source text is rather vague. In view of the impact of the EU framework on the target language and texts, those aspects were analysed in several studies and works (see for example Koskinen, 2000; Piehl & Vihonen, 2006; Pym, 2000; Wagner et al., 2002). From the perspective of the receiver, the clarity and readability of texts are crucial since a language difficult to understand may run counter to the Union's aim to bring citizens closer to Europe. The need for good quality Community legislation has been emphasized in several documents and events in the EU.⁷

The last aspect, the role of translators, is also influenced by the conditions which EU institutions have established for translation. Focusing on the European Commission, Koskinen (Koskinen, 2000: 86) differentiates among three categories: *intracultural* translation (within an *anational* EU culture), *intercultural* translations (when communicating between the EU and national cultures) and, finally, legal translations which are treated as a separate case. Being part of the EU administration, the translator is bound to be affected by the culture and domestic discourse which the institutions have developed. As Koskinen argues, this is the reason why even *intra*- and *intercultural* translations tend to follow the logic and pattern of legal translations, overlooking the need for cultural transfer in this type of translation. We will follow a similar line of argument in the next section of the paper by analysing the translation of terminology *between* and *within* conceptual systems in the EU context.

6. I was able to discuss those factors within the framework of informal meetings with translators and terminologists of the EU institutions in Brussels and Luxemburg on 23 and 28 March 2007.

7. See Declaration No 39 attached to the final Act of the Treaty of Amsterdam, Council Resolution of 8 June 1993, on the quality of the drafting of Community legislation, as well as the seminar series on the quality of legislation organized by the Commission's Legal Revisers Group.

4. Translation and terminology

4.1 The notion of translation in terminology

Sager (1990: 80) differentiates between primary and secondary term formation. Primary term formation is a monolingual activity, the process of designating a new concept, whilst secondary term formation starts from an already existing term and is either mono- or multilingual. Accordingly, it may have two aims: the revision of an existing term within one linguistic community (a monolingual activity) or the transfer of an existing term into another linguistic community (an inter-lingual activity). This latter process, the transfer of terms into other linguistic communities is generally referred to as 'translation'. In the EU context, however, it seems necessary to refine the notion of translation in those term-formation processes. In Sager's model, the three activities (the monolingual creation of a new term, the monolingual revision of an existing term and the inter-lingual transfer of an existing term) are distinguished on the basis of new/existing terms or mono-/multilingual activities. In the following, an attempt will be made to widen the scope of the model to include EU specific aspects and involve an additional dimension, namely, the conceptual level.

Starting with the primary creation of terms, it is important to note that it may not only be mono- but multilingual as well. In that case, multilingual terminology is created in a parallel way, excluding the need for translation. In other words, *multilingual primary term-creation* is not about translation, but a simultaneous, multilingual activity aiming at the designation of one concept in several languages. In countries and institutions with more official languages, multilingual term-creation would be the ideal way of creating terms. In this case, from a conceptual point of view, primary term-creation is carried out *within* one conceptual system, but in *more* languages. However, for most languages, multilingual primary term-creation generally makes up a two-step process in practice: (1) primary term-creation in some languages followed by (2) secondary term-creation (translation) into the other languages.

Turning to secondary term-creation, i.e. the translation of an existing term, an additional dimension is suggested for consideration, namely, the existence of one or more conceptual systems. From this conceptual point of view, we may differentiate two cases in the translation of terms: term-transfer into another language *within* the same conceptual system (*intra-conceptual term-transfer*) and term-transfer *between* different conceptual systems (*inter-conceptual term-transfer*). In the first case, although terms have to be translated, the process is carried out within the same conceptual system. This is characteristic of step (2) in multilingual term-creation in practice, as discussed in the previous paragraph. In the second

case, the difference between conceptual systems plays a significant role. Here, the process of translation ideally takes an onomasiological approach, comparing first the two conceptual systems and then finding or creating the equivalent target term for a source term. The difference between conceptual systems and the problem of equivalence is especially relevant in social sciences (different legal, economic systems) but plays a role in natural sciences as well (Muráth, 2002; Arntz, 1994; Schmitt, 1994).

In an EU context, the conceptual approach also plays a significant role since both the EU and several national conceptual systems interact, with twenty-three languages describing them. Below I shall first present the conceptual systems and languages interacting at EU-level and then analyse the peculiarities of multilingual term-creation in the EU conceptual system.

4.2 The interaction of languages and conceptual systems at EU-level

To analyse the conceptual systems at EU-level it is necessary to show the potential combinations of languages which may relate to conceptual systems. To make a clear delimitation possible, 'conceptual system' is understood as the legal system, and 'language' as the official language within a state. Relying on Sandrini's model (2004) on trans-national inter-lingual communication, we differentiate three categories. The first can be described as a one-to-one relationship when one language enjoys official status in one state and the language is linked to a single conceptual system (Hungarian-Hungary). The second is a more-to-one relationship typical in states where more languages have official status and, hence, more languages may belong to one and the same conceptual system (Swedish and Finnish – Finland). Finally, a one-to-more relationship is when one language belongs to more conceptual systems, that is, the language is officially recognized in several states (French – Canada and France). Due to its (partly) supra-national nature, the EU developed its own conceptual system, which we refer to below as 'the *EU conceptual system*'. This system may be compared to the second category: one conceptual system having at present twenty-three official languages.

However, the twenty-three official languages not only describe the EU conceptual but also the national conceptual system(s). Take the example of German, which serves several national conceptual systems (Austria, Germany, Belgium) *and* the EU conceptual system. As a result, the complexity of translations and terminology work at EU-level lies in the fact that all the above language-conceptual system combinations may be present whilst translating EU-texts. In other words, both terms belonging to the EU and terms belonging to the national conceptual system(s) may occur in the same EU text. The terms belonging to the EU conceptual system will henceforth be called 'EU-terms' and the terms of the national

conceptual system ‘national terms’, with a further explanation to follow in Section 4. We can, therefore, differentiate two different processes: on the one hand the designation of new EU-concepts in twenty-three languages (*multilingual primary term-creation* of EU terms) and, on the other, the translation of national terms (*inter-conceptual term-transfer*). In other words, it may involve both *multilingual primary term-creation within* the EU conceptual system and translation (*term-transfer*) *between* national conceptual systems.

There is, moreover, constant interaction between the national and the EU conceptual system also. The EU conceptual system cannot be considered separately, or completely isolated from the national conceptual systems since, at the same time, its official languages belong to one or more national conceptual systems. In other words, there is originally no ‘EU-language’ which exclusively belongs to the EU conceptual system. If such a language is needed, it must inevitably be based on the current twenty-three languages. As Sandrini (2004) points out, in neutral translation an EU specific legal language is called for in order to make a clear distinction between national and EU-level regulations possible. Within each language, a specific part (whether terminology or textual features) relates to the EU conceptual system. This calls for a constant check with national conceptual systems in order to see whether or not the proposed designation at EU-level is already linked to a national concept. In other words, a continuous check is needed in order to avoid the conceptual variability of one and the same term – as will be discussed later. In the next section our analysis focuses on the first process at EU-level, that is, on *multilingual term-creation within* the EU conceptual system.

4.3 Multilingual term-creation within the EU conceptual system

As discussed above, primary term-creation may be both monolingual and multilingual. Within the EU conceptual system the equality of official languages would, in theory, presuppose that primary terms are created in all official languages in parallel. In other words, for one EU concept, twenty-three designations should be given simultaneously. In reality, similarly to the drafting of texts, the creation of terms is carried out in two steps: terms are first created in the dominating languages (mainly in the procedural languages of English, French and German) and then *translated* into all other languages. This means that, in most languages, target terms (secondary terms) are created on the basis of a source term (primary term), by translation. Turning to the conceptual dimension of terminology translation, it is important to emphasise that this process, whether by or without translation, is still an *intra-conceptual* activity, since the terminology has to be created within one (the EU) conceptual system. As a result, the creation of EU terminology can be described as a two-step process: (1) *multilingual primary term-creation* for the

dominant languages followed by (2) a secondary activity, an *intra-conceptual term-transfer* for most other languages.

The importance of translation in term-creation at EU-level is also underlined by the participants taking part in the process. Whilst, for primary terms, the process of conceptual thinking and designation is carried out by politicians, experts and civil servants (depending on the stage of decision-making), secondary terms are created by the translators/terminologists in the EU institutions. As a result, although the approach still has to be onomasiological (starting from the concept), the choice is already influenced by the presence of a primary term and by the fact that the term has to be *translated*. The fact that, for most languages, terms are created by translation makes a mere onomasiological approach questionable. Translators depart from a primary term which inevitably influences their choice of the target language designation of the same concept.

In addition, developments and needs at national level may also influence the translators' choice. Eurocrats may not have the same preferences as translators or linguists. They tend to accept a 'foreign-sounding' term in the national language, on the basis of which the original term may easily be deduced. This aspect is all the more important since, as mentioned before, citizens rather need ordinary, non-technical terms that are easy to understand. The translator has, therefore, to balance the different needs. Another challenge the translator may face is when the professional community at national level starts using a term long before a document is to be translated. It becomes even more difficult if there is no consensus on a specific term in the target language. Secondary term formation at national level usually takes place within the professional community among experts who have the time to discuss the integration of a foreign term into the national conceptual and linguistic system. If this has not yet been achieved, the translator has to produce a term – and usually within a very short period of time. As a result, EU translations may accelerate (or even put an end to) consensus seeking at home. Nevertheless, the creation of a term at EU-level does not mean its automatic acceptance in the professional community, which in turn may result in parallel designations of the same concept; an example of the linguistic variability of terms. If, at a national level, no authority has the power to decide which term to adopt, the translator is left alone with the choice.

Another difficulty is caused by the vagueness of the EU concepts themselves. Both the given policy area and the field of science may influence the degree of exactness of a specific subject field. In a policy which is not harmonized (education, for example) it may actually be in the political interest to avoid clear terms (more examples to be found in Wagner, 2000). Similarly, as Felber and Schaefer point out in their model (Muráth, 2002), the scientific field may influence the proportion of 'terminologized' and 'non-terminologized' special vocabulary in a

subject field. Generally, the special vocabulary of natural sciences tends to be more exact, having more terminologized vocabulary than social sciences.

4.4 Linguistic and conceptual variability of terms at EU-level

Two questions will be dealt with in this section: the different perceptions of the notion of 'EU term' and the linguistic and conceptual variability of terms in the EU context. Starting with the question of what may be considered as an EU term, several difficulties arise. The first one may come from the different approaches of translators, terminologists, linguists or experts. Translators tend to consider terms in the broadest sense, wishing to include everything which makes their work easier into a terminological database. Similarly, the definition of an EU term also depends on differing perceptions. In the broadest sense, all terms appearing in an EU text may be called EU terms, which would mean that even terms belonging to national conceptual systems could be referred to as EU terms. Limiting this approach, we consider an EU term to be one which belongs to the EU conceptual system, that is, a term which designates an EU concept. However, in a second step, this definition may be further reduced to terms which directly relate to the peculiar nature of the Union; terms which make the EU what it is. It is beyond the limits of this paper to elaborate this question in more detail or to argue in favour of one approach or the other. Instead, we would refer to a practical approach, the IATE Best Practice (Interactive Terminology for Europe), which lays down general input criteria for terms to be included.⁸ The database takes a rather translator-oriented approach. Included are not only terminology in the narrowest sense, but also special expressions (not necessarily terms) – that is to say, all items considered to be useful for translation. The IATE terminology database is an important element of corpus planning at EU-level and of outstanding importance both for translators and for the general public.

The interaction of conceptual systems and languages at EU-level may lead to the formal and conceptual variability of terms; whereby variability means that either one concept belongs to more designations (formal) or one designation refers to more concepts (conceptual). The formal variability of terms may result from term-creation for different target groups. For example, one designation may be given for Eurocrats and another for European citizens who need more understandable, clear terms. Another example of the formal variability of terms may come from the translation of EU directives (Somssich, 2003). EU directives are first translated at EU-level and, then, for a second time, at national level by incorporating the

8. IATE. Best Practice for Terminologists. 2007. BP revised, draft 1. Inter-institutional Committee for Translation and Interpretation. IATE Terminology Coordination Team.

directive into national law. Since lawyers operating at the national level are not bound by the wording of the directive, a different national designation may be found for the same concept. As a result, one term may occur in the directive and another in national law, both referring to the same concept. Turning to conceptual variability, the variation of terms may result from the interaction of the EU and the national conceptual system within the same language. As was previously underlined, in order to see whether the proposed designation at EU-level is already linked to a national concept, continuous checks impose themselves. However, if a national term is used to describe an EU concept, the term will relate both to an EU and to a national concept. This may be more typical of the dominating languages, which are used as primary languages in the process of conceptual thinking.

5. Conclusion

In order to trace inter-relationships among language policy, translation policy, translation and terminology work in an EU context, this paper has followed a similar pattern. In the first section, the legal basis of a language and translation policy and the competences of the EU were analysed. It was shown that, although not codified as a policy area, an implicit language and translation policy may be deduced. The EU has the formal competence to regulate language use at an institutional level. From the point of view of language planning, the EU-status of national languages is defined by the Member States, acting unanimously in the Council. National involvement in corpus planning, however, is rather limited, since, after a country's accession, translation and terminology work are transferred to EU institutions. Since a specific EU-language and terminology is developed in all the official languages, corpus planning plays a huge role in the quality aspect of terminology work. Those activities are directly carried out at EU-level (e.g. inter-institutional terminology work, IATE) whilst, at national level, contribution may only be indirect, in the form of providing highly skilled translators and formalizing co-operation between translators in the EU and experts at home.

In the second section, the impact of translation policy on translation was examined. Each EU institution, having developed its own practice and culture, constitutes a unique framework for translation which influences both the linguistic and extra-linguistic factors of the translation process. The dominance of some source languages, the vague notion of the source text, interference of other languages, all influence the quality of translations.

In order to refine the notion of translation in terminology, a conceptual dimension and EU specific aspects were included in Sager's (1990) model on primary and secondary term-creation. Starting with the conceptual aspect, it was

argued that, in an ideal case, primary term-creation may also be multilingual (multilingual primary term-creation). In secondary term-creation, a distinction was made between term-transfer (translation) within one conceptual system (*intra-conceptual term-transfer*) and term-transfer into another conceptual system (*inter-conceptual term-transfer*). In terms of EU-specific aspects, three different processes were distinguished. Whilst for some languages (mainly English, French and German) terms are created simultaneously (*multilingual primary term-creation*), for most languages terms are created via a secondary activity, that is, by translation (*intra-conceptual term-transfer*). In this case translators most often start their translation based on a primary term which inevitably influences their choice for target language designation. However, term-creation in the EU not only concerns creating terms for new EU concepts, but also translating national terms. (*inter-conceptual term-transfer*). As a result, the complexity of translation and terminology work at EU-level lies in the fact that it may involve both *primary term-creation* or *intra-conceptual term-transfer* within the EU-conceptual system and *term-transfer* between national conceptual systems. Those factors further underline the impacts of EU-level translations on the official languages, on the quality of texts and the role of translators, calling for further research in this field.

References

- Arntz R. (1994). Terminologievergleich und internationale Terminologieangleichung. In Snell-Hornby M. (Hrsg.), *Übersetzungswissenschaft. Eine Neuorientierung* (283–311). Tübingen & Basel: Francke Verlag.
- Branchadell, A. and Lovell, M. W. (ed.). (2005). *Less Translated Languages*. Amsterdam & Philadelphia: John Benjamins.
- Fischer, M. (2007). Fordítás(politika) és terminológia az Európai Unióban. In Heltai, P. (szerk), *A XVI. Magyar Alkalmazott Nyelvészeti Kongresszus előadásai* (806–811). Pécs & Gödöllő: MANYE, Szent István Egyetem.
- Haugen, E. (1983). The implementation of corpus planning: Theory and practice. In J. Cobarrubias and J.A. Fishman (eds.). *Progress in language planning* (269–289). The Hague: Mouton.
- Klaudy, K. (2004). *Bevezetés a fordítás elméletébe*. Budapest: Scholastica.
- Kloss, H. (1969). *Research Possibilities on Group Bilingualism: a Report*. Quebec: International Center for Research on Bilingualism.
- Koskinen, K. (2000). *Beyond Ambivalence. Postmodernity and the Ethics of Translation*. Academic Dissertation. Tampere: University of Tampere.
- Muráth, J. (2002). *Zweisprachige Fachlexikographie. Pécsér Beiträge zur Sprachwissenschaft*. Budapest: Nemzeti Tankönyvkiadó.
- Piehl, A. and Vihonen I. (toim.) (2006). *Vuosikymmen EU-suomea*. Helsinki: Kotimaisten kielten tutkimuskeskus.
- Pym, A. (2000). The European Union and its Future Languages. Questions for Language Policies and Translation Theories. *Across Languages and Cultures* 1 (1) 1–17.

- Sandrini, P. (2004). Transnationale interlinguale Rechtskommunikation: Translation als Wissenstransfer. In Burr, I. und Müller, F. (Hg.), *Rechtssprache Europas. Reflexion der Praxis von Sprache und Mehrsprachigkeit im supranationalen Recht* (139–156). Berlin: Duncker & Humblot.
- Schmitt A., P. (1994). Die ‘Eindeutigkeit’ von Fachtexten: Bemerkungen zu einer Fiktion.’ In Snell-Hornby M. (Hrsg.) *Übersetzungswissenschaft. Eine Neuorientierung* (252–283). Tübingen & Basel: Francke Verlag.
- Somssich, R. (2003). A jogfogalmi megfeleltetés problémái a közösségi jogban az irányelvek átültetésének szintjén – a jogi ‘fordítás’ sajátos formája. *Magyar Jog* 50 (12) 746–753.
- Sager, J. C. (1990). *A Practical Course in Terminology Processing*. Amsterdam & Philadelphia: John Benjamins.
- Wagner, E. et al. (2002). *Translating for the European Union Institutions*. Manchester: St Jerome.

The situation and problems of Hungarian terminology

Ágota Fóris¹

The aim of this article is to offer a general view of Hungarian terminology, using concrete examples to analyse the experiences gained in the Hungarian-speaking territories within Hungary and outside its borders.

In Hungary, the ever-increasing number of new products and scientific achievements, and the broadening of conceptual systems of services, administration, education, etc. made quick and scientifically well-founded terminological classification necessary. The Hungarian economy and society have been transformed these past two decades. The two most important steps in the substantial changes of terminology were the change of political system in 1989, and accession to the European Union in 2004. In the paper I will focus on the terminological problems caused by such variable circumstances.

Keywords: small languages, Hungarian terminology, economy, society, language planning

1. FÓRIS ÁGOTA, (born: 1970, Hungary), habil. (2006), PhD (2002), linguist. Her field of research includes LSP lexicography, researching technical and scientific vocabulary, and terminology. She is the editor and author of the Hungarian-Italian and Italian-Hungarian Technical and Scientific Dictionary, and the author of the book titled *Hat terminológia lecke* [Six Lectures on Terminology], and many other publications. She is a member of the Board of the EAFT (European Association For Terminology); she is an elected member of the Dictionary Work-committee and the Applied Linguistics Work-committee of the Hungarian Academy of Sciences, an elected member of the Board of the Society of Hungarian Association of Applied Linguists and Language Teachers, and a founding member of the MaTT (Council of Hungarian Terminology).

She works at the Károli Gáspár University (Budapest, Hungary), Faculty of Arts, Department of Hungarian Linguistics, Associate Professor; she is the Head of the Terminology Research Group (TERMIK). (E-mail: aforis@t-online.hu)

1. Introduction

The achievements of Hungarian linguistics are barely known in Western Europe, hence this short overview of the territories and countries where Hungarian is spoken, as well as the historical background of Hungarian terminology. Publication of such texts goes back several centuries, but most of them have been published in scientific literature in Hungarian.

In the second half of the paper I will focus more in detail on the situation of Hungarian terminology research at the turn of the 20th and 21st centuries, and evaluate those findings. This paper does not aim to present and assess the situation of terminology teaching; instead it aims to provide a general picture of Hungarian as a small and less widely known language. Finally, after assessing the situation, I will outline further tasks for the future, primarily with regard to the organization of research.

1.1 Linguistic environment –

The Hungarian language and speakers of Hungarian

98.5% of the ten million people living in Hungary speak Hungarian as their mother tongue. Hungarian is also the country's official language. A significant number of native Hungarian speakers also live outside Hungary's borders. Rough estimates suggest that 5 million such people exist: 3 million of them in neighbouring countries, with the biggest group in Romania. One third of all Hungarian speakers live outside Hungary.

The minority languages spoken inside Hungary are, in decreasing order of importance (number of speakers): German, Slovakian, Croatian, Gypsy (Lovari and Beas), Romanian, Serbian, and Slovenian. The languages spoken by people who immigrated to Hungary are: Polish, Greek, Armenian, Bulgarian, Ruthenian, Ukrainian, and one language that has become more widespread in recent years: Chinese.

The most widely learnt foreign languages in the country are English and German. Most young people speak English and German, while German and Russian are more widespread amongst older people. Other languages taught in Hungary include: French, Italian, Spanish, Latin and (less frequently) Finnish, Japanese, Chinese, Korean and Swedish.

1.2 A short history of Hungarian terminology

Since the 16th century, the transformation of social, economic and technical life in Europe has given rise to millions of new concepts, which have created a new terminology. Formal acknowledgement of the fact that increasing knowledge and the

resulting linguistic-lexical changes indeed affect the whole population, led to language planning. It is widely known that in the 17th–19th centuries, one of the priorities of intellectual life in Europe was to develop national languages that met the challenges of science, industry and economic development.

In the Hungarian-speaking language area, Latin used to be the language of religion, science, education, and developing technology (the official language was Latin until 1844). Growing industrial activity made German more and more widely used, which manifested itself in the large number of German terms in languages for specific purposes (LSPs, i.e. in the language of crafts). Following the example set in Western Europe, where efforts were already being made to create independent Italian, French, German national languages, a national movement took shape to reform the Hungarian language. In Hungary, the first organized assignments to form new terms were carried out during the period of the ‘neologist movement’ (*nyelvújítás*, 1772–1872) – the theoretical and linguistic issues of those efforts are discussed in several Hungarian books and articles (e.g. Tolnai, 1929; Pais, 1955; Kovalovszky, 1955; Fábíán, 1984). New terms were recorded in LSP dictionaries. The modernization of the languages of economy and industry followed the modernization of scientific language. One period of the Hungarian neologist movement coincided with the development of capitalism in Hungary: growing industrialization, increasingly mechanized agriculture, and the construction of the railway network not only brought new tools, machines, and work processes with them, but also their terminology. The establishment of an independent Hungarian industry and agriculture also required new Hungarian terminology. In addition, the languages of literature, conversation and poetry were fundamentally renewed during this period. The basic questions of terminology, a discipline that developed later on, were implicitly handled in the discussions held during the neologist movement and published by those engaged in the debate. Already in 1834, József Bajza recognized that the formation of terms and the classification of terminology are concept-based (Bajza, 1834).

Terminological work carried out during the time of the neologist movement focused on the formation of terms. Within this deliberately defined limit, the questions concerning the organization of terms into a system and the development of a terminology system remained in the background. The system of terms developed according to the logical systems of the different professions. Deliberate terminological classification took place only in the 1900s, through the needs of specialized scientific branches.

The large-scale post-war industrialization had a positive influence on Hungarian terminology research. One of the most significant achievements of the interdisciplinary co-operation between linguistics and the specialized scientific

branches was the progress made in terminology research throughout the middle of the 20th century, and the appearance of several LSP dictionaries.

Half a century ago, János Klár and Miklós Kovalovszky published a volume of studies entitled '*Műszaki tudományos terminológiánk alakulása és fejlesztésének főbb kérdései*' [The Major Questions of the Development and Improvement of our Technical Scientific Terminology] (Klár & Kovalovszky, 1955). The book groups the results, collected by the authors, of terminological changes, analyzes existing deficiencies and outlines further tasks for the future. Despite the topicality of the issue, no similar works followed this volume in the 20th century. It took until 2005 for a summarising monograph on terminology to be published (Fóris, 2005). Systematization of technical terminology, in other words, began in the 1950s, when its principles and methods were established (based mostly on the achievements of the neologist movement and the Soviet School of Terminology) – and, as an achievement, the publication of the technical explanatory dictionary series began (cf. Pusztai, 1988).

In the 1970s specialized translation groups were founded at Hungarian universities. So-called 'training in specialized translation' was launched in 1974, and at the same time the teaching of language for specific purposes began (Klaudy, 1993). All this had a positive effect on the theoretical foundation of terminology and on the articulation and solution of practical problems alike. The collection and recording of the terminology of mining and heavy industry, for example, brought about significant achievements.

An elaborate set of Hungarian terms and systems of terminology were developed during the 17th–20th centuries in relation to the historic crafts, branches of science, sports, technical sciences and related industries, conventional agriculture, etc. This system called for further growth to adjust to the accelerating pace of development (for details see Fóris, 2007a).

2. Present-day situation and problems of Hungarian terminology

2.1 Two important steps regarding terminology: The changed political system (1989), and accession to the European Union (2004)

As in other languages, most of the current Hungarian terminological problems appeared in the last two decades of the 20th century, as a result of the effects of economic and social globalization in our language (cf. Fóris, 2007b). The ever-increasing number of new products and scientific achievements, and the broadening of conceptual systems of services, administration, education, etc. made quick and scientifically well-founded terminological classification necessary. This was also

the time when the Hungarian economy and society were transformed. However, the decreased economic output could not provide the financial requirements for terminological development.

Under such circumstances, a comprehensive research network of terminology could not develop, and it was pared down to the isolated examination of fractions of a problem area. The old and new terminological paradigms coexist, creating confusion as to the proper treatment of both practical and theoretical issues. According to a lot of linguists, for example, terminology spans a linguistic problem area that is independent of the conceptual system; that is, they focus on whether the lexeme is well-formed according to Hungarian norms, but ignore how it is related to concepts.

Until the turn of the millennium, no independent degree courses in terminology was offered by any of the Hungarian universities. No co-ordinated efforts to compile reference works that would serve terminological development have been made, even though the drawbacks of terminological disorder have been voiced in the fields of economy, law, public administration and others. For instance, the Departments of Linguistics, Economics and Law of the Hungarian Academy of Sciences organized a joint debate on the topic in May 2006, where a lot of problems were brought up.

Let us examine an example from economics. The transformation of the Hungarian economy began in the 1980s (with, amongst others, the introduction of a new system of taxation in harmony with the market; the formation of a dual-level banking system, that is, a central bank and commercial banks; conversion to a fixed [currency] price regime, etc). Still, only at the end of that decade were the legal and political conditions crucial to privatization and the multi-party system created, which made a peaceful transition possible. The two most important steps in the substantial changes in the terminology of economics were the change of political system in 1989, and accession to the European Union in 2004.

The Soviet model of economic organization collapsed in 1989, taking the *conceptual system of socialist economy* down with it. The ideological basis of socialist terminology was a unified approach to economy, according to which the conceptual systems of *economics* and *business and finance* merged. By the 1990s a new terminological system had to be created to correspond to the Western-European conceptual system (this has been more or less achieved, but several problems remain). As a result, the terminology of economics, business and finance had to be disentangled, and in order to do so, post-war Hungarian terminology needed to be consulted (Chikán, 2006).

Before joining the EU in 2004, we had to gradually adjust to the conceptual system of the European Union, and harmonize our legal, administrative, economic and business practice (e.g. laws, standards). The so-called language of the

European Union caused a lot of difficulties, primarily for translators translating legal texts, regulations and the like into Hungarian. More often than not, the European conceptual system does or did not match the Hungarian one (some concepts are missing entirely, and therefore do not have a name in Hungarian). Consequently, the need to describe the concept and to find a name for it usually appear at the same time. The adaptation of EU terminology to Hungarian calls for considerable effort, since appropriate Hungarian lexemes must be found for existing (and defined) conceptual EU systems, without having them overrule the Hungarian terms already in use. When a Hungarian lexeme is missing and a new one has to be created, it must be done according to the norms and rules of Hungarian terminology and word formation. The main problem lies in the absence of a Hungarian protocol, that would co-ordinate the development of new terminology, as well as evaluate its results. Separate efforts are being made simultaneously, as a result of which several synonyms are put into use at the same time (Várnai, 2005).

2.2 Terminology work and research: Huge room for improvement

Let us examine some specific areas where a higher intensity of terminology work is essential, e.g. reference works, terminology databases, standards, language planning.

Hungary is lagging behind in the creation of synchronic terminology reference works. The Hungarian National Corpus is an openly accessible corpus of present-day Hungarian, but it does not contain any LSP texts. Other terminology databases, or reference lexicons are not publicly accessible. Several specialized dictionaries have been compiled with commercial aims, and a number of Hungarian institutions have databases for terminological purposes and for internal use, but none of them is accessible freely and openly. Terminological harmonization has not yet taken place within the Hungarian language area.

EU membership obliges Hungary to keep introducing European standards and to eliminate clashing national standards. A further difficulty is caused by the fact that the majority of new EU standards (more than 70%) are not available in Hungarian, only in English (not enough manpower or money to fund their translation). Also, the terminology section of standards is available in English only and access to it expensive. For the internal European market to work, technical obstacles to commerce must be eliminated. A means to that end is the development of common European standards which member states are forced to introduce as national standards and impose over their existing, national standards. Currently there are nearly 20 000 European standards, and the number increases by 1 500 every year. Due to the shortage in funding, a mere 23% of them are available in Hungarian, and the rest (including the terminology standards) are Hungarian national standards in English (personal communication by *József Haba*, the chief

advisor of the Hungarian Standardization Board, 2005, cf. Fóris 2005). Given that the leaders of Hungarian SMEs have a poor command of foreign languages, this puts the Hungarian economy and industries at a serious disadvantage compared to businesses in those EU countries where standards are available in the mother tongue. Furthermore, research, translation, interpretation and education, etc. are all facing serious difficulties.

Hungary lacks centralized and scientifically organized language planning that would extend to related fields and territories. The Hungarian research conducted inside and outside the country's borders is not co-ordinated. No common norms are created or mutually accepted either. As regards the Hungarian speaking areas outside Hungary, Hungarian researchers have to cope with several problems concerning language use, terminology and translation. The fact that LSP dictionaries, unified terminological registers and databases are also lacking in Hungary, makes the task of Hungarian authors and/or translators living outside Hungary even more difficult. They face an especially demanding task concerning the translation of geographical names, place-names, names of institutions, as well as terminology within the newer branches of medical science, biology and economics (Lanstyák & Szabó Mihály, 2002, 2005; Péntek, 2004). Professional fields that developed during the past – sports originating from the Far East, the wellness and service industries, or ICT – lack tradition and thus have several problems. These past few years, instruction in Hungarian has become possible in vocational schools located in territories outside the borders of Hungary but populated by Hungarians. Unfortunately, Hungarian professionals who completed their studies at Romanian, Slovakian universities, etc. – who by the way speak excellent everyday Hungarian – are not willing to teach in Hungarian because they are not familiar with the Hungarian terms and the Hungarian terminological system of their profession. 'Now that the possibilities for vocational instruction in Hungarian have vastly improved, it is the Hungarian teachers who refuse to teach in Hungarian, claiming they are not familiar with Hungarian technical language' (Péntek, 2004: 241). Hungarians living outside the country's borders live in a bilingual environment. The biggest advantage they could have in order to succeed is a high command of both their mother tongue and the majority language. They must be able to use everyday Hungarian and technical Hungarian, as well as be familiar with the standard and technical variety of the majority language. To this end, they need fast and easy access to textbooks, dictionaries, terminological databases, and other reference works; and the simplest way to make this possible would be to set up freely accessible databases on the internet.

J. Péntek drew our attention to the fact that Hungarians living within the country's borders have easy access to ICT terminology, but for Hungarians living abroad, accessing new terminology may be problematic. Even international

companies accept the fact that some languages have an official status; for example, computer programmes officially sold in Romania are either in English or Romanian (Péntek, 2003).

Several European countries currently co-ordinate and plan terminology work on a national level (e.g. Cabré, 1999; Hartmann, 1999; Pusztay, 2002; Gaivenis, 2002). Detailed information on international terminological research, events and publications can also be found in several books and articles, and on the websites of various terminology associations.

In the field of Hungarian linguistics, terminological questions are dealt with in many studies based on translation studies (e.g. Albert, 2005). In relation to lexicography, the question of recording terms in LSP dictionaries is the most frequently discussed topic (e.g. Heltai, 1988; Fóris, 2002; Dróth, 2003). Many papers have also been written on teaching foreign professional languages and intercultural communication (e.g. Borgulya, 1999). In the domain of terminology some improvements have recently been made.

2.3 The situation of Hungarian terminology

Personal research, carried out between 2004 and 2006, has resulted in the following conclusions about the situation of Hungarian terminology (cf. Fóris, 2005, 2007a):

New concepts and matching terms come into being on a much larger scale and within shorter periods of time. The new discoveries and their related terminological elements are put into use very quickly. The classification of new concepts and the formation of new terms must be done within a short period of time.

Before the introduction of certain tools, processes and concepts, no precise description or conceptual and terminological definition tends to be formulated; as a result, a term is often accidentally introduced for the occasion. That causes problems in language use as well as the conceptual system. As terms quickly integrate and spread, it is very difficult to correct the false terms arising from this quick terminological development of language for special purposes (LSP) and language for general purposes (LGP).

As the role of scientific knowledge is vital in terms of the development of knowledge itself, scientific analysis is also required in the field of terminological development. The formation of conceptual and terminological systems is only possible if more specialized and linguistic knowledge come together.

The clear use of the terminology of a given language can only be guaranteed by defining concepts in a manner that corresponds to present-day scientific knowledge and can be used at different linguistic levels without any contradictions.

The frequent professional insufficiencies and mistakes in the entries of (LSP and LGP) dictionaries describing technical issues may be attributed to the

fact that lexicographic and terminological work is carried out without consulting representatives from the related professional fields. No such people take part in the revision of dictionaries prior to their publication. The results of my research into the definitions of terms in Hungarian dictionaries match the results of Zajankauskas obtained in Lithuanian dictionaries (Fóris, 2006; Zajankauskas, 2006).

Dictionary publishers often discard established traditions in order to gain quicker and easier access to information. This primarily manifests itself in missing or inaccurate definitions of concepts. Dictionary users look up the meaning of words unknown or partially known in vain and in many cases they are simply unable to find an appropriate answer.

Few new dictionaries exist. The same applies to glossaries, encyclopaedias and electronic databases containing the terms and definitions of concepts in certain technical fields.

No real information system shows the results of Hungarian terminological research and practical work.

To solve the above problems, lengthy co-operation of several researchers and experts is needed. Recently made terminology efforts are promising – such as:

2.4 Initiatives and solutions in Hungarian terminology

Terminological work may have several aims. The following five areas play an important role in Hungarian terminology.

1. The creation of new terms to replace missing or incorrect ones. – These days, the task of coining new terms is becoming more complex. On the one hand, more and more new concepts come into being that lack a Hungarian name. Recently, quite a number of new crafts, scientific branches, sports, etc. appeared whose Hungarian terms are in the making (although in the case of sports from the Far East, for example, English terms are being used). On the other hand, their rapid development does yield many incorrect terms.
2. The systematization and maintenance of the terms belonging to a certain domain. – The key steps in this process are: collecting terms, classifying them, based on the concepts and the terminological system, putting them into a system, and making corrective suggestions. The number of new tasks escalates due to the stagnation of terminological classification for several decades and the fast expansion of the term set in every field.
3. Teaching terminology. – It is our task to present the terminological system of a given profession in a nuanced fashion. (i) According to one approach, the teaching of terms and the terminological system is an inherent part of professional training, and is usually done in the mother tongue. (ii) According to another approach, the professional language of the given field is also taught in

a foreign language (in Hungary this is done in the training of specialized translators). (iii) Another type of educational objective could be the presentation of a given profession's terminology along with the general academic questions of terminology, either as part of an independent professional training programme or in further education. I am referring to the training of terminologists, which is already a fact in some European countries (no such training currently exists in Hungary).

4. The creation of terminological databases. – Terminology bases form the most effective means of ensuring free access to terms and terminological systems.
5. Research into terminology. – International, European and national terminological systems change so fast that the effects of those processes must be checked at various levels. The above-mentioned results of my research support this statement and indicate that these are just the first steps in fathoming the changes in terminology.

Two essential organizations of Hungarian terminology were founded in Szombathely in 2005 and 2006:

The MaTT (*A Magyar Nyelv Terminológiai Tanácsa* / Council of Hungarian Terminology) was formed in 2005, as a subsidiary of UNESCO. It aims to support the co-operation of professionals and researchers, and provide the theoretical co-ordination of Hungarian terminological works.

The TermIK (*Terminológiai Innovációs Központ* / Terminology Innovation Centre) was founded in September 2006 at the Institute of Intercultural Studies (Berzsenyi Dániel College, Szombathely). In 2009 it was relocated to the Károli Gáspár University (Budapest) and continues its work under the name Terminology Research Group. It aims to play a determining role in the creation and maintenance of the national and international network of relations in favour of theoretical research into terminology, and also to offer improvements and services for applications. The Termik carries out its activity both on a national and international level. Within the widespread and multi-layered terminology tasks, it wishes to play a key role in the creation of theoretical foundations. We would like TermIK to play a catalytic role in the co-ordination of the extensive tasks related to terminological problems.

3. Conclusion

In my paper I overviewed the situation, present and past, and problems of Hungarian terminology before and after our accession to the European Union. As not all speakers of Hungarian live inside Hungary, this study spans the

terminological situation in- and outside Hungary. After evaluating the present situation, I outlined the main tasks that must be tackled.

A final note: this paper was written and delivered at the 'Terminology and Society' conference. Since then, Hungarian terminology efforts have become more intense. Let me mention one important achievement that helps disseminate information and scientific findings: the international journal entitled 'Magyar Terminológia' (the Journal of Hungarian Terminology) was launched, and its first volume published in June 2008. I hope we will be able to report on more achievements and fewer problems.

References

- Albert, S. (2005). Un type spécial de *contresens*: apparition d'une tautologie dans le texte-cible. In K. Károly and Á. Fóris (eds.), *New Trends in Translation Studies. In Honour of Kinga Klaudy* (157–175). Budapest: Akadémiai Kiadó.
- Bajza, J. (1843). *Nyelvünk műveléséről*. Budapest.
- Borgulya, Lné (1999). *Bankkenntnisse II. Kunden- und Mitarbeitergespräche*. Budapest & Pécs: Dialóg-Campus.
- Cabré, M. T. (1999). *Terminology. Theory, Methods and Applications*. Amsterdam & Philadelphia: John Benjamins.
- Chikán, A. (2005). *A gazdasági szaknyelv (át)alakulása a rendszerváltás után*. [The (trans)formation of language for economics after the change of the political system] Talk at the Conference 'A társadalomtudományok szaknyelve', 12 May 2005, MTA (Hungarian Academy of Sciences).
- Dróth, J. (2003). Egy korszerű szakszótár elkészítésének alapjai [The principles of an up-to-date technical thesaurus]. *Magyar Nyelvőr*, 127 (2), 159–167.
- Fábián, P. (1984). *Nyelvművelésünk évszázadai*. Budapest: Gondolat.
- Fóris, Á. (2002). *Szótár és oktatás* [Dictionaries and Education]. Pécs: Iskolakultúra.
- Fóris, Á. (2005). *Hat terminológia lecke* [Six Lectures on Terminology]. Pécs: Lexikográfia Kiadó.
- Fóris, Á. (2006). Lexicographical Definition of Terms in Hungarian Dictionaries. In E. Corino, C. Marelllo and C. Onesti (eds.), *Atti del XII Congresso Internazionale di Lessicografia, Torino, 6–9 settembre 2006. Proceedings. XII EURALEX International Congress* (776–772). Alessandria: Edizioni dell'Orso.
- Fóris, Á. (2007A). Hungarian terminology today. In J. Pusztay (ed.), *Terminology and Lexicology in Middle-Europe* (15–24). Szombathely: BDF.
- Fóris, Á. (2007B). Terminology and social-economic globalization. *Terminologija*, 14, 49–60.
- Gaivenis, K. (2002). *Lietuvių terminologija: teorijos ir tvarkybos metmenys* [Lithuanian Terminology: an Outline of Theory and Regulation]. Vilnius: Lietuvių Kalbos Institutas.
- Hartmann, R.R.K. (ed.) (1999). *Dictionaries in language learning. Recommendations, national reports and thematic reports from the TNP sub-project 9: Dictionaries*. Berlin: Thematic Network Project in the Area of Languages.
- Heltai, P. (1988). Contrastive Analysis of Terminological Systems and Bilingual Technical Dictionaries. *International Journal of Lexicography*, 1 (1), 32–40.

- Klár, J. and Kovalovszky, M. (1955). *Műszaki tudományos terminológiánk alakulása és fejlesztésének főbb kérdései* [The Major Questions of the Development and Improvement of our Technical Scientific Terminology]. Budapest: MTESZ.
- Klaudy, K. (1993). Fordítástudomány a világban – fordításoktatás Magyarországon. In K. Klaudy (ed.), *In Harmadik Magyar Alkalmazott Nyelvészeti Konferencia* (39–45). Miskolc: Miskolci Egyetem.
- Kovalovszky M. (1955). Tudományos nyelvünk alakulása. [The Formation of our Scientific Language]. In D. Pais (ed.), *Nyelvünk a reformkorban* (227–312). Budapest: Akadémiai Kiadó.
- Lanstyák, I. and Szabó Mihály G. (eds.) (2002). *Nyelvi érintkezések a Kárpát-medencében*. [Languages in Contact in the Carpathian Basin]. Pozsony (Bratislava): Kalligram Könyvkiadó – A Magyar Köztársaság Kulturális Intézete.
- Lanstyák, I. and Szabó Mihály, G. (2005). *Hungarian in Slovakia*. In A. Fenyvesi (ed.), *Hungarian Language Contact Outside Hungary. Studies on Hungarian as a minority language* (47–88). Amsterdam & Philadelphia: John Benjamins.
- Magyar Terminológia [Journal of Hungarian Terminology] (2008). <http://www.akademiai.com/content/1789-9486>.
- Pais, D. (ed.) (1955). *Nyelvünk a reformkorban*. Budapest: Akadémiai Kiadó.
- Péntek, J. (2003). *Anyanyelv és oktatás*. [Mother Tongue and Education] Csíkszereda: Pallas Akadémia.
- Péntek, J. (2004). Magyar nyelvű tudományosság – kezdet és vég? [Science in Hungarian – the beginning and the end?]. In J. Péntek (ed.), *Magyarul megszólaló tudomány* (233–242). Budapest: Lucidus.
- Pusztai, I. (1988). A szaknyelvi kutatások kérdései. [Questions of research into LSP]. In J. Kiss and L. Szűts (eds.), *A magyar nyelv rétegződése* (120–130). Budapest: Akadémiai Kiadó.
- Pusztay, J. (2002). Nyelvi tervezés a kis finnugor (uráli) népeknél. [Language planning at small Finno-Ugrian (Uralic) peoples] In K. Gadányi and J. Pusztay (eds.), *Közép-Európa: egység és sokszínűség. A Nyelvek Európai Éve 2001 zárókonferenciájának előadásai* (246–251). Szombathely: BDF.
- Tolnai, V. (1929). *A nyelvújítás*. [The neologist movement.] Budapest: MTA.
- Várnai, J. Sz. (2005). Európai uniós terminológia és fordítás – múlt és jelen. [Terminology and translation of the European Union – part and present] *Fordítástudomány*, 7 (2), 5–15.
- Zajankauskas, S. (2006). *Apie klystkelius reiškiant poliškumą ir susijusias sąvokas*. [On wrong paths in defining poliškumas and related concepts.] Talk at: Tautinių kalbų terminologija ir globalizacija – Terminology of National Languages and Globalization. Vilnius, Lithuania, Lietuvių Kalbos Institutas Terminologijos Centras, 11–13 October 2006.

Translation-oriented terminology work in Hungary

Judith Muráth

In Hungary the translation of specialized texts poses considerable problems, even for professional translators, and, as a result, translation-oriented terminology work has become more and more prominent. The need for translation features in a large part of the business life spectrum, and translation-oriented terminology work in one or more foreign languages also requires the co-operation of linguists, translators and experts in the given fields. Briefly, the aim of this paper is to introduce those institutions which can be considered as having paved the way for current work in translation-oriented terminology. At the same time, terminology research and other activities carried out before and after EU accession will be addressed. Problems arising in the field of terminology will then be discussed, using concrete examples from the social sciences and comparing German with Hungarian.

Keywords: translation-oriented terminology work, specialized dictionaries, term bases, practice and theory, unambiguousness, one-to-one correspondence, contextual autonomy, text level and system level, social science, social pension

1. The problem defined

The translation of specialized texts causes considerable difficulties, even for professional translators. In Hungary, as a new EU member, work in the field of translation-oriented terminology is therefore becoming more and more prominent. The vocabulary of domain-specific language is constantly being amended. Unconstrained developments in science and technology have also contributed to the dynamic of the Hungarian language. On top of that, the democratic upheavals of 1989 aggravated the changes in both the economy and politics. The transition from a centrally-planned to a market-economy not only had individuals facing considerable difficulties; it also manifested itself in the field of LSP and the related terminology. Since then, Hungary has joined the European Union, and the issue of specialized translation has acquired a new dimension. Translation embraces

wide-ranging professional interests and bi-, perhaps even multi-lingual terminology work requires the collaborative efforts of translators, linguists, terminologists and experts in the specific fields. Even though, fortunately, more and more authorities and researchers now take on terminology work, there is still no sign of co-ordinated collaboration – apart from a few admirable initiatives. First and foremost, co-operation is explicitly formulated and initiated by specialized translators and by the compilers of specialized dictionaries, including term bases. The current situation being so very closely related to that preceding the regime change, activities since 1980 will be looked at first.

2. Recent terminology activity in Hungary

In the '70s terminology research formed a relevant part of LSP research at Hungarian universities. However, with the development of 'Term to Text', such aspects of specialized communication as 'text analytical', 'stylistic-pragmatic', as well as 'intercultural' came to the forefront, and so terminology work as an activity of linguistics was forced more and more into the background. Terminology was in the hands of state authorities and specialist organizations: the Hungarian Academy of Sciences, the Hungarian Standards Institute as well as banks' and companies' in-house departments. Terminology was not an independent research area but closely linked to specialized lexicography, and, so to speak, formed the initial phase in preparing specialized dictionaries (cf. Muráth 2000a: 39; Gárdus et al. 1980).

At the same time, a counter-tendency could be detected within university education. In 1973 a Specialized Translator and Interpreter programme was introduced at the Faculty of Philology of ELTE (the Eötvös-Loránd-University of Budapest) as an FE (Further Education) course. Shortly afterwards, from 1974 to 1986, initiated by the then Ministry of Education, programmes in Specialized Translation were introduced in eight so-called Specialist Universities and combined with the main curriculum. They were: the Technical University of Miskolc (1974), the Natural Science Faculty of the Kossuth Lajos University of Debrecen (1976), the Agrarian University of Debrecen (1978), the Agrarian University of Gödöllő (1979), the Natural Sciences Faculty of the Eötvös Loránd University of Budapest (1979), the Faculty of Economics at the University of Pécs (1979), the Horticultural University of Debrecen (1980), and, finally, the Szentgyörgyi Albert Medical University of Szeged (1986). Only later, in 1990, did the Foreign Trade College of Budapest and the Budapest University of Technology join the group, providing Further Education for Specialized Translators and Interpreters (cf. also Klaudy, 1997: 177). As a result, the need for work on translation-oriented terminology developed, thus stimulating research in said institutions. Those independent

attempts, however, did not reach beyond the circle of individual institutions and only became more generally known when the democratic changes of 1989 and accession to the European Union in May 2004 gave new impetus to the strengthening of monolingual and, above all, translation-oriented terminology work.

To facilitate and co-ordinate work in terminology, several measures were agreed upon in Hungary. To undertake the translation work which Hungary had had to cope with pre-accession, a team was established within the Hungarian Ministry of Justice whose task lay in co-ordinating translation and terminology work. In 2003, HUTERM, an online terminology discussion forum for translators, interpreters and language revisers was set up. In May 2005, at the Berzsenyi Dániel College in Szombathely, under the umbrella of the Hungarian UNESCO Commission, the Council on Hungarian Terminology (MATT) was founded – chaired by Vilmos Voigt of ELTE, Budapest. Conference sections as well as round-table discussions were devoted to terminology. Examples of such gatherings are: the Congress on Applied Linguistics (Miskolc 2005; Gödöllő 2006), the International Conference on Interdisciplinary Aspects of Translation and Interpreting in June 2005 in Pécs (cf. Muráth et al. 2007), as well as the Terminology Conferences of 2005 and 2006 in Szombathely. In 2006, again at the Berzsenyi College in Szombathely, a Terminology Innovation Centre was established, while a Terminology Documentation Centre was opened at Pécs University's Faculty of Business and Economics (in collaboration with the University Library).

3. Practice and theory

In Hungary, for twenty-odd years now, translation-oriented terminology work has primarily been a *hands-on activity* in training centres for specialized translators, organized within the normal curriculum: starting with terminology-related textual analysis, via the supervision of degree theses, to the compilation of bi-lingual terminology collections. Not only linguistics experts were involved, but also engineers and economists – in other words, experts in the fields concerned. Gradually, dictionaries were being revised, which led to the compilation of specialized dictionaries and term bases. This yielded centres for terminology work, with the aim of contributing to specialized translation within a particular area by various means – specialized dictionaries or computer-aided term bases.

The first of such initiatives was connected with the University of Miskolc (cf. Gárdus et al. 1980). The LSP research work undertaken there went hand-in-hand with terminology work and the production of specialized dictionaries. The Agrarian University of Debrecen is well-known for its compilation of multi-lingual specialized dictionaries on agricultural affairs. The six-language specialized

dictionary of plants published in 2006 involved a long period of activity by many individuals (cf. Gallyas et al. 2006; Andrásy 2006: 91). The Faculty of Economics at the University of Pécs can also look back on a long, active and corpus-based activity in specialized dictionary production – above all in the fields of statistics, economics, social policy and the European Union. The bi-lingual specialized dictionaries in German and Hungarian were based on international projects. The terminological dictionary of statistics (cf. Bösel et al. 1997) and the 2-volume Context Dictionary ‘Economics and Social Policy’ – one part dealing with Economics, the other with Social Policy – were intended for a broad circle of interested parties. Users include students in specialized translation programmes, students of Economics studying German as an LSP, as well as economists and sociologists, translators and interpreters (cf. Muráth et al. 1998; Muráth 2007). Apart from the printed dictionaries, term bases were also envisaged for specialized translators in Hungary. The ‘Nature and Environmental Protection’ project is run by the St. István University of Gödöllő (English-Hungarian) (cf. Dróth 2003) whilst ‘General Statistics’ and ‘Dynamic Terminology of Economics and Social Policy’ (German-Hungarian), ‘Biomass’ (German-English-Hungarian), ‘the European Union’ (German-Hungarian) resort under the Faculty of Business and Economics of the University of Pécs. Some are in their completion phase and will soon be offered to interested parties via the internet.

In several institutions, the work involved in compiling specialized dictionaries and term bases, as well as the efforts devoted to developing the up-to-date teaching of Translation and Terminology to students, have led to both individual and joint research. Practice (i.e., praxis) needs a *theory*, or at least theoretical consideration, which paves the way towards practice. Research contributing to translation-oriented terminology work is inevitably interdisciplinary – in that it approaches the problem from different starting-points. In Hungary the issues belonging to the relevant directions include contrastive lexicological terminology research (cf., inter alios, Heltai 1979, 1981, 1985, 1988; Muráth 2000a, 2002, 2007; Hubainé Oláh 2005), research into the processing of terminology via computer-based methods and the development of suitable technology (cf. Muráth 1992, 2000b; Dróth 2003; Erdős 2004), socio-linguistic research (cf. Muráth 2000a, 2002, 2005; Kovács 2004), corpus-linguistic analyses (cf. Muráth inter alios. 2000a: 165; Dróth et al. 2006; Heltai 2006a, 2006b; Károly 2006), terminology, terminology policy and translating in the European Union (Várnai 2004; Fischer in this volume) together with the term in the translation process (cf. Muráth 2000a: 156, 2002, 2007; Heltai 2004; Rádai-Kovács 2004).

4. The term and the translation process

For a long time, and for whatever reason, translation-oriented terminology work long remained the exclusive domain of practitioners trying to develop a theory. That is why the General Theory for Terminology (GTT) seemed the logical source to fall back on. It was introduced by the Soviet school (primarily Reformatski) into the Hungarian specialized literature. Klár and Kovalovszky, who in 1955 published a book on the origin and development of Hungarian scientific terminology for technology, based their argumentation on the Soviet School of Terminology. The work (83 pages in total) was the sole monograph in this specialized field in Hungary for 40 years. All more recent publications (cf., inter alios, Szépe 1982; Kurtán 2003; Fóris 2005) are based on this theory. The work, although published by the Language Preservation and Translation Department of the Federation of Associations of Technology and Natural Sciences, devoted no more than a paragraph (cf. Klár et al. 1955: 75) to translation. In the '80s, following the publication of the Wüster Theory in *Magyar Nyelvőr* (cf. Pusztai, 1980), the already well-known terminology theory was further endorsed and the development of a translation-oriented theory hardly seemed rational.

Only in the '90s, in the course of several international dictionary and database projects – carried out under the leadership of Muráth of the Faculty of Economics at the University of Pécs in collaboration with, firstly, colleagues from Jena and Pécs¹ and, later, from Graz² (cf., inter alios Muráth 2000a, 2000b, 2007) – was it demonstrated that, from a translation process point of view, the classical Theory of Terminology (regarded at the start of the project, in 1994, as the basis for further work) did, in fact, require revision. Most of all, the practical work involved in a bilingual German-Hungarian corpus dealing with the relevant issues of Economics and Social Policy in Germany, Austria and in Hungary from 1993 to 1997, showed that what was needed was a dynamic specialized vocabulary of Economics and Social Policy. Reality simply did not correspond with the related literature.

1. The first database and dictionary project was undertaken with the cooperation of two statisticians, Martin Bösel of the Friedrich-Schiller-University of Jena and Katalin Rédey of the Faculty of Economics of the University of Pécs (cf. Bösel et al. 1991, 1997).

2. Research and editing of Economics Terminology was carried out in three international projects of 1994 – 1998 supported by the Foundation Action Austria-Hungary, the Economics Faculty of the Janus-Pannonius-University of Pécs and the Institute for Translator and Interpreter Training of the Karl-Franzens University of Graz. Project partners were Erich Prunč, Edina Dragaschnig, Wolfgang Eisenhut, Irene H. Pogány, Wolfgang Titus Tockner and also Marianne Zserdin. From the second project onwards, in parallel to Economics, Social Policy was also included in the research, and in this way the research field was broadened in terms of terminology in an area much less researched – as the economy had already demonstrated (cf. Muráth et al. 1998).

In Germanistics-related LSP research, modelled on Wüster, the priority of the concept to designation, one-to-one correspondence ('Eineindeutigkeit'), system relevance and context autonomy were regarded as attributes of the term (cf. Hoffmann 1984: 163; Fluck 1985: 33, Wüster 1970: 94). Also Klár and Kovalovszky, who researched the scientific stratum of Hungarian technical language, insist on 'unambiguousness' ('Eindeutigkeit') in the sense of 'one-to-one correspondence' ('Eineindeutigkeit')³. Context Autonomy was, admittedly, not explicitly formulated by the Hungarian academics, but this follows from the System Character of the term as declared in Point 2 (cf. Klár et al. 1955: 41). In contrast to this theory, those attributes of the term could rarely be seen in the corpus already addressed. As a rule, hierarchical concept relationships (as is customary in Natural Sciences and in Technology) could not be established in this case. The specialized vocabulary in the areas researched was quite heterogeneous in the extraordinary political and economic situation at the time of the regime change in Central and Eastern Europe, and only part of it could be classified as terms. For most terms it is a matter of relative unambiguousness – which only follows from the context. Further, the completed analyses – also in the form of case studies – drew attention to the fact that the specialized terminology of Economics (including Social Policy) is much influenced at the concept and designation level by specific characteristics, by the different levels of development of individual countries, and, simply, the differences between countries in both economic and management systems (cf. Muráth 2000a: 93, 2002: 85). In several cases, concept contamination could be detected in the text – even in known terms. Those cognitions, following Gerzymisch-Arbogast, led to the definition of a theory which, apart from the system level of the language, also took into consideration the text level of the language (cf. Muráth 2000a: 156, 2002: 138; Gerzymisch-Arbogast 1996). In contrast to the classical terminology

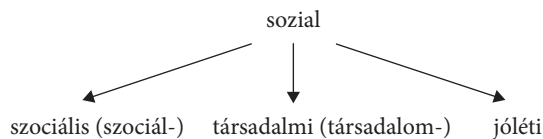
3. In the 1950s Klár and Kovalovszky assembled the requirements relating to the Hungarian terms in technical language under 8 items. The first requirement was 'exactness' and 'unambiguousness' (in German 'Exaktheit' and 'Eindeutigkeit', in Hungarian 'egzakt' and 'egyértelmű'): '... a synthetic word or expression should express only one concept' (Klár and Kovalovszky 1955: 41, translated by myself -- JM). In this context, the word 'unambiguousness' (in German 'Eindeutigkeit', in Hungarian 'egyértelműség') is to be understood as 'one-to-one correspondence' ('Eineindeutigkeit'), deriving from the above definition. Wüster's term to describe 'one-to-one correspondence' ('Eineindeutigkeit'), is in Hungarian ('egyegyértelműség') simply not possible. One possible solution would be 'total unambiguousness' ('teljes egyértelműség'), but Klár and Kovalovszky do not use this formula. Consequently, the term 'egyértelműség' (unambiguousness) is a polyseme in Hungarian relating to both 'one-to-one correspondence' and 'unambiguousness', since Klár and Kovalovszky make no distinction between 'Eineindeutigkeit' and 'Eindeutigkeit', i.e. mean by 'unambiguousness' 'Eineindeutigkeit' ('one-to-one correspondence'). As a result, the distinction between the two Wüster-terms (Eineindeutigkeit, Eindeutigkeit) is not taken into consideration.

theory, which, by stressing the system of concepts, gives precedence to the concept level, the translator, should, in my opinion, move between text level and system level during the translation process. In translation, the text level (the original text) should be the starting point. Terms should be conceptually examined within the text, and terminological system information, in contrast to the conceptual development of the terms, should also be relied on (cf. Muráth 2000a: 158, 2002: 139; Gerzymisch-Arbogast 1996: 245p). Meanwhile, this theory appears to be supported by more recent publications in the field of Term and Translation (cf. Heltai, 2004; Rádai-Kovács, 2004).

5. Terminology work in the social sciences

The Social Sciences present an especially interesting (although, at the same time, problematic) field, and drastic changes in the Social Security system have aggravated the situation. Not only in Hungary, but also in Germany and Austria, is the vocabulary of the Social Sciences marked by a particular dynamic. In addition, a number of country-specific features within the German language must be considered. In certain areas – in Hungarian as well as in German – neologisms occur (amongst others, so-called EU-neologisms). In the following, I would give a few examples of the difficulties translators face when translating from German into Hungarian and vice versa. For this analysis, German and Hungarian legal texts, periodicals and newspaper articles, as well as dictionaries, were collected and evaluated.

5.1 Collocations and quasi-synonyms in the Hungarian language



The word *sozial* (social), having to be translated into Hungarian, is itself a delicate matter. It appears in German in a variety of LSP expressions, either as an attributive (*soziale Sicherheit*), or as a determiner, that is, as the first element of a compositum (*Sozialversicherung*). Possible equivalents in Hungarian are either *szociális* (*szociál-*), *társadalmi* (*társadalom-*) (*gesellschaftlich*, relating to society) or, more rarely: *jóléti* (*gemein-*, welfare). Both will be treated as synonyms – which they by no means are. Only from the micro-context does it become clear which is to be used as the Hungarian equivalent. On the one hand it is a question of collocation, whilst, on the other hand, the final choice depends upon the understood content of the two adjectives together: *társadalmi* (rarely, *társadalom-*) rather more a noun, a

substantive, which, first of all, relates to society as a system, or, secondly, means the ‘togetherness’ of men/mankind (cf. Pusztai, 2003: 1311). *Szociális* (*szociál-*), on the other hand, has as its first meaning, *társadalmi*, but it also means the social situation of a person and is used in compound terms addressing this issue (cf. Pusztai 2003: 1273). Hardly ever, in translation, can both adjectives appear as equivalents. One exception is created, for example, by *Sozialpolitik*, whose translation is context-dependent. Since the above considerations are not exactly widely known, even dictionaries give false equivalents. A few examples pertaining to German: Hungarian (cf. Muráth et al. 1998 vol. 2; Hessky 2000, Várnai 2004; Fata 2005):

<i>soziale Entwicklung</i>	–	<i>társadalmi fejlődés</i>
<i>soziale Ordnung</i>	–	<i>társadalmi rend</i>
<i>soziale Errungenschaften</i>	–	<i>társadalmi vívmányok</i>
<i>Sozialversicherung</i>	–	<i>társadalombiztosítás</i>
<i>Sozialabgaben</i>	–	<i>társadalombiztosítási járulék(ok)</i>
<i>Sozialkritik</i>	–	<i>társadalomkritika</i>
<i>Sozialkunde</i>	–	<i>társadalomismeret</i>
<i>Sozialstruktur</i>	–	<i>társadalmi szerkezet</i>
<i>Sozialpolitik</i>	–	<i>társadalompolitika</i>
As opposed to		
<i>soziale Ausgaben</i>	–	<i>szociális kiadások</i>
<i>soziale Einrichtungen</i>	–	<i>szociális intézmények</i>
<i>soziale Leistungen</i>	–	<i>szociális juttatások</i>
<i>soziales Netz</i>	–	<i>szociális védőháló</i>
<i>soziale Sicherheit</i>	–	<i>szociális biztonság</i>
<i>Sozialhilfe</i>	–	<i>szociális segély</i>
<i>Sozialwohnung</i>	–	<i>szociális lakás</i>
<i>Sozialpolitik</i>	–	<i>szociálpolitika</i>
As opposed to		
<i>Sozialstaat</i>	–	<i>jóléti állam</i>

5.2 Differences in designation levels in language comparison

German	Hungarian
<i>Krankenkasse</i>	<i>egészségpénztár</i>

The Hungarian social security law prescribes what contributions insured individuals must pay to the *Krankenkasse* (literally Sick Service). However, in Hungarian this is not referred to as ‘Kranken-’ (sick), but ‘egészség-’ (health), and so, in a mirror

translation, this would give ‘Health Fund’ (in German, ‘Gesundheitskasse’). It is perfectly evident, of course, that, even though in the German Social Security System the term ‘sickness’ (Krankheit) and in the Hungarian System the term ‘health’ (Gesundheit) are stressed, the responsibilities of both funds are identical. The one question being: how to translate both terms? Should one opt for an ‘adaptive’ translation, such as ‘die ungarische Krankenkasse’ (the Hungarian Sick Fund) or, rather, an alienating version such as ‘die ungarische Gesundheitskasse’ (the Hungarian Health Fund)? Also: should the ‘adaptive’ translation (‘a német egészségpénztár’ = German Health-Fund) or the alienating translation (‘a német betegpénztár’ = German Sick Fund) be used in Hungarian? Before WWII, Hungary actually did have a so-called *betegsegélyező pénztár* (Krankenkasse). This designation is codified in current dictionaries, even though it sounds archaic, and no health care service in Hungary bears that name today.

5.3 National terminology and EU terminology

Finally, one further term will be mentioned, used not only in German but at both the national language and the EU-level.

German	Hungarian
<i>Sozialrente</i>	<i>szociális nyugdíj</i>

5.3.1 *Sozialrente in Germany*

The term *Sozialrente* (Social Pension) is used colloquially in Germany. It goes back to Chancellor Otto von Bismarck’s Social Legislation or Social Law. 1891 indeed saw the introduction of the statutory Pension Insurance, a pension also known as the Social Pension (cf. also Fata 2005: 108).

Example 1

Die abgespeckte gesetzliche Rentenversicherung würde ihre Aufgabe als einkommens- und beitragsbezogenes Alters-Grundsicherungs-System also auch künftig erfüllen. Allerdings darf nicht vergessen werden, daß Rentnerhaushalte heute schon neben der Sozialrente weitere Einkünfte haben. Dazu gehören insbesondere Zusatzrenten aus einer betrieblichen Altersversorgung und Vermögenseinkünfte.

Mit Reformen unter 20 Prozent. iwd – Nr. 9 vom 27. Februar 1997

[The trimmed-down statutory Pension Insurance was designed to fulfil the task of providing an income and a contribution-related, old age, basic pension system for the future, but it should not be forgotten that today’s

pensioner households have an income over and above the social pension. In this category lie supplementary pensions from a company pension scheme and income from assets.]

Example 2

Die Dreißig- bis Vierzigjährigen sind fast einhellig der Überzeugung, sie müssen im Alter verarmen, wenn das System der umlagefinanzierten Sozialrente nicht endlich auf Kapitalfonds umgestellt werde. Und es wächst der Unmut, daß ihnen heute gesetzliche Rentenbeiträge vom Bruttogehalt abgezogen werden, die zusammen mit dem, was der Arbeitgeber für sie zu zahlen hat, bis zu 1600 Mark monatlich ausmachen können für sie anscheinend verlorenes Geld. Denn die staatlich organisierten Rentenkassen verteilen diese Beträge doch gleich an die heutigen Rentner, das Geld wird sofort aufgebraucht.

Otto Mayer (Ossietzky): Rente aus Kapitalfonds kontra Sozialrente www.linksnet.de/artikel (25.01.2001)

[The thirty-to-forty year old age group are almost unanimously convinced that they will become impoverished in their old age, if, that is, the current shared cost (pay-as-you-go) pension system is not finally replaced by capital funding, and there is a growing displeasure that today's statutory pension contributions are deducted from their gross salary, which, together with what the employer has to pay for them can amount to DM 1,600 per month – which they perceive as lost money. Since the state-organized pension funds redistribute these contributions immediately to today's pensioners, the money is immediately used up.]

5.3.2 Sozialrente in the EU context

Scrutiny of EU information pages, at least where German can be selected as a language of communication, yields the following information. In Denmark the *Sozialrente* is a pension for disabled people. It is financed by taxation and is independent of any earlier occupational or previous contribution-based payment. In Portugal, the *Sozialrente* is a non-contributory pension, but not a Disability Pension.

Research in Denmark, Germany and Portugal confirms that *Sozialrente* in all three countries means payment provided by the Social Security system. Conditions, however, are totally different in the individual countries, so at no time can they be regarded as identical, even though they may have the same designation. Despite this, a correct interpretation is possible, since what we have are two different levels, one national (Germany) and one EU, where German is used as a tool for understanding. A greater problem arises when translating into Hungarian. The Hungarian term *szociális nyugdíj* (cf. Várnai 2004: 1073) is a neologism, on which little can be built in Hungarian. Since the term is undefined and codified with no information, one can only guess at the concept behind it. One thinks of Social

Assistance, since in Hungary all pensions are based on contributions and years of service, and, at most, Social Assistance (cf. however also Fata, 2005: 108, where the information concerning the German context is clear). Based on this research, one can be quite clear as to what *Sozialrente* means in Germany, in Denmark, and in Portugal, although there still remains the unsolved question of how those different contents can be differentiated in Hungarian by means of suitable designations.

6. Conclusion

As shown in this paper, there is a great variety of terminological activity at Hungarian universities, with a long tradition of activity. Cases in point are the centres for the training of specialized translators, which are at the same time centres for translation-oriented terminology work. The practical work is increasingly supported by theoretical considerations. Since the above activities are still generally unrecognized in Hungary, those activities must be followed up – and this even more intensely – in the immediate future. First of all, however, they must be made known to a wider public. Co-operation by the relevant institutions would contribute significantly to an upsurge in terminology work in Hungary. By means of the example of translation-oriented terminology work in the field of the Social Sciences, and by means of a number of examples, the difficulties involved in this work and from which the problems in different relationships emerge, can be pinpointed. The dynamic and terminological uncertainties in the Hungarian language, the differing social systems in the German-speaking countries and in Hungary, as well as the Social Science terms used at EU level, imply different levels of analysis. To be able to establish the actual situation, more in-depth textual analyses must be undertaken, in which not only the positions of terms should be interpreted in the conceptual and designation system, but also the use of terms in texts. Analyses should be carried out both contrastively and at all three levels referred to, by interdisciplinary and, at the same time, international teams. Finally, the codification of terms and methods of terminology work should be unified at a European level.

References

- Andrássy, G. (2006). Hatnyelvű növénytermesztési szótár. In M. Fekete-Silye (ed.). *Porta Lingua*, 2006, 91–97. Debrecen: ATC.
- Böselt, M., Muráth, J. and Rédey, K. (1991). *Magyar – német, német – magyar statisztikai kisszótár*. Pécs: JPTE KTK.
- Böselt, M., Muráth, J. and Rédey, K. (1997). *Statisztikai kisszótár. Magyar – német, német – magyar. / Statistisches Wörterbuch Ungarisch – Deutsch, Deutsch – Ungarisch*. Budapest: KSH.

- Dróth, J. (2003). Egy korszerű szakszótár elkészítésének alapelvei. *Magyar Nyelvőr*, 2003, 2, 159–167.
- Dróth, J. and Turcsányi, G. (2006). Az üregi nyúl mint egzotikus európai állat. Gondolatok a természet- és környezetvédelmi terminológia fordításáról. In J. Dróth (ed.). *Szaknyelv és szakfordítás* (54–61). Gödöllő: SZIE GTK.
- Erdős, J. (2004). Terminológia és fordítástechnológia a BME fordítóképzésében. In J. Dróth (ed.). *Szaknyelv és szakfordítás* (106–114). Gödöllő: SZIE GTK.
- Fata, I. (2005). *Ungarisch-deutsches, deutsch-ungarisches Fachwörterbuch zur Rentenversicherung*. Szeged: Grimm.
- Fischer, M. (in this volume). Language(policy) Translation and Terminology in the European Union.
- Fluck, H-R. (1985). *Fachsprachen*. Tübingen: Francke.
- Fóris, Á. (2005). *Hat terminológiai lecke*. Pécs: Lexikográfia.
- Gallyas, Cs. and Petrikás, Á. (eds.) (2006). *Növénytermesztési szótár*. Budapest: Mezőgazda.
- Gárdus, J., Sipos, G. and Sipőczy, Gy. (eds.) (1980). *Szaknyelvoktatás – szaknyelv kutatás. Tanulmányok a felsőoktatás köréből*. Budapest: FPK.
- Gerzymisch-Arbogast, H. (1996). *Termini im Kontext. Verfahren zur Erschließung und Übersetzung der textspezifischen Bedeutung von fachlichen Ausdrücken*. Tübingen: Narr.
- Heltai, P. (1979). *A Contrastive Study of Hungarian and English Agricultural Terminology*. [dissertation, University of ELTE, Budapest].
- Heltai, P. (1981). Poliszmia a terminológiában. *Magyar Nyelvőr*, 105, 4, 451–462.
- Heltai, P. (1985). The Relationship Between General and Scientific Vocabulary. *Euralex Bulletin* 2, 2, 1–3.
- Heltai, P. (1988). Contrastive Analysis of Terminological Systems and Bilingual Technical Dictionaries. *International Journal of Lexicography* 1, 1, 32–40.
- Heltai, P. (2004). Terminus és köznyelvi szó. In J. Dróth (ed.). *Szaknyelv és szakfordítás* (25–43). Gödöllő: SZIE GTK.
- Heltai, P. (2006a). Párhuzamos szaknyelvi korpusz munkálatai. 6. *Szaknyelvi szimpózium*. Szeged.
- Heltai, P. (2006b). Szakmai kommunikáció és szaknyelv. In M. Fekete-Silye (ed.), *Porta Lingua – 2004* (37–42). Debrecen: DE ATC.
- Hessky, R. (ed.) (2000). *Deutsch-ungarisches Handwörterbuch*. Budapest & Szeged: Nemzeti Tankönyvkiadó/Grimm.
- Hoffmann, L. (1984). *Kommunikationsmittel Fachsprache*. Berlin: Akademie.
- Hubainé Oláh, Á. (2005). A terminológia oktatásának aktuális problémái. In J. Muráth and Á. Hubainé Oláh (eds.). *A XXI. század kihívásai a szakfordítóképzésben* (51–55). Pécs: PTE KTK.
- Károly, K. (2006). Szövegkutatás és fordítástudomány. In J. Dróth (ed.), *Szaknyelv és szakfordítás* (33–40). Gödöllő: SZIE GTK.
- Klár, J. and Kovalovszky, M. (1955). *Műszaki tudományos terminológiánk alakulása és fejlesztésének főbb kérdései*. Budapest: MTESZ.
- Klaudy, K. (1997). *Fordítás I. – Bevezetés a fordítás elméletébe*. Budapest: Scholastica.
- Kovács, I. J. (2004). *Gazdasági terminusok szociolingvisztikai aspektusban*. [PhD dissertation, University of ELTE, Budapest].
- Kurtán, Zs. (2003). *Szakmai kommunikáció*. Budapest: Nemzeti Tankönyvkiadó.
- Muráth, J. (1992). Computer und Sprache. Überlegungen beim Aufbau einer Terminologie-Datenbank an der Universität Pécs. In N. Bradean-Ebinger (ed.). *LINGUA DEUTSCH 5., Fremdsprachige Hefte*. (27–35). Budapest: Lehrstuhl für Germanistik im Institut für Fremdsprachen an der Universität für Wirtschaftswissenschaften.

- Muráth, J., Dragaschnig, E., H. Pogány, I. and Zserdin, M. (1998). *Wirtschaft & Sozialpolitik – aktuell. Wörterbuch Deutsch – Ungarisch, Ungarisch – Deutsch*. Pécs & Graz: JPU/KFU. Bd. 1: Wirtschaft. Bd. 2: Sozialpolitik.
- Muráth, J. (2000a). *Zweisprachige Fachlexikographie in Theorie und Praxis – dargestellt am Beispiel der ungarisch-deutschen und deutsch-ungarischen Wirtschaftskommunikation*. [PhD dissertation, University of ELTE, Budapest].
- Muráth, J. (2000b). Datenbank zur Wirtschaftsterminologie unter dem besonderen Aspekt des demokratischen Wandels. In A. Schnaider (ed.). *Projektberichte 1995–1999*. Wien-Budapest: Aktion Österreich-Ungarn Wissenschafts- und Erziehungskooperation.
- Muráth, J. (2002 [2003]). *Zweisprachige Fachlexikographie*. Budapest: Nemzeti Tankönyvkiadó.
- Muráth, J. (2005). Wörterbuchbenutzung und Fachübersetzerstudenten. Ihre Erwartungen an ein Fachwörterbuch. In H. Gottlieb, J.E. Mogensen and A. Zettersen (eds.). *Symposium on Lexicography XI* (401–415). [LEXICOGRAPHICA Series Maior 115] Tübingen: Max Niemeyer.
- Muráth, J. (2007). Terminologische und fachlexikographische Forschungen an der Wirtschaftswissenschaftlichen Fakultät der Universität Pécs. In J. Muráth and Á. Oláh-Hubai (eds.). *Interdisziplinäre Aspekte des Übersetzens und Dolmetschens/ Interdisciplinary Aspects of Translation and Interpreting* (465–482). Wien: Praesens.
- Pusztai, I. (1980). A bécsi terminológiai iskola elmélete és módszertana. *Magyar Nyelvőr*, 104, 1, 3–16.
- Pusztai, F. (ed.) (2003). *Magyar Értelmező Kéziszótár*. Budapest: Akadémia.
- Rádai-Kovács, É. (2004). Terminológiakurzus a fordítóképzésben. In M. Fekete-Silye (ed.). *Porta Lingua 2004* (263–276). Debrecen: DE ATC.
- Szépe, Gy. (1982). A szaknyelv és a mindennapi nyelv kapcsolata. In *A technika tanítása* 5, 129–139.
- Várnai, J. Sz. (ed.). (2004). *Official Terminology of the European Union*. English-Hungarian-French-German. Budapest/Bicske: Morphologic/ Szak.
- Várnai, J. Sz. (2004). Az európai uniós terminológia magyar nyelvű egységesítése. In J. Dróth (ed.). *Szaknyelv és szakfordítás* (64–78). Gödöllő: SZIE GTK.
- Wüster, E. (1970). *Internationale Sprachnormung in der Technik. Besonders in der Elektrotechnik*. (Die nationale Sprachnormung und ihre Verallgemeinerung) Dissertation von 1931. Berlin & Bonn: Bouvier.

Towards a national terminology infrastructure

The Swedish experience

Henrik Nilsson

In October 2002, the Swedish Centre for Terminology, TNC, launched a programme (TISS, Terminology Infrastructure for Sweden) aimed at proposing an enlarged terminology infrastructure for Sweden. The programme, financed by the Swedish Ministry of Enterprise, Energy and Communications (Näringsdepartementet), described the basic prerequisites and needs for such an infrastructure. The suggested main components of the terminology infrastructure are: (1) a terminology portal containing a national term bank and (2) a terminology co-ordination programme. During 2005, TNC's ideas from the TISS programme were taken into consideration in two Swedish Government bills. This paper presents TNC's work aimed at achieving a terminology infrastructure in Sweden, including preparations for the construction of a national term bank and the testing of the software of IATE (= Inter-Active Terminology for Europe) for this purpose, and the lobbying work connected to those activities.

Keywords: national term bank, *Rikstermbanken*, terminology co-ordination, terminology infrastructure, terminology training, terminology work, the Swedish Centre for Terminology, TNC, *Terminologicentrum*, National Board of Health and Welfare

1. Introduction

An adequate and quality-assured term bank will be an efficient tool for both companies and authorities, as well as for the development of new ICT-services.

When in early 2006, Ulrica Messing, the previous Minister for Communications, made this comment upon the Government's decision to grant TNC 1.5 million SEK¹ for the establishment of a national term bank, that not only constituted an important step towards the realization of a national terminology infrastructure in

1. About 140 000 euros.

Sweden, but also an indication of the growing terminological awareness in Swedish society.

This paper will retrace the steps taken towards realizing a national terminology infrastructure in Sweden, by first presenting the programme named TISS (Terminology Infrastructure for Sweden) completed in 2004, then by discussing the on-going project aimed at creating a national term bank (*Rikstermbanken*), with special focus on its contents, technology (including an evaluation of the IATE software), and usage.

2. The concept of 'terminology infrastructure'

The idea of a terminology infrastructure is not new. It was underlined in the Pointer project, in 1996, where the following definition of 'terminology infrastructure' was put forward:² framework of

- institutions, companies (LEs and SMEs), associations, self-employed professionals, etc.;
- their terminological activities; and
- co-operation and communication networks (on both the physical and logical levels) they operate in for a given application area

In Budin & Wright (1997), it was later described as 'all arrangements and configurations of people working together, of institutions dedicated to or responsible for terminology-related activities, producing and using different kinds of information resources, reference materials, archives, databases, etc.'³ Galinski (1998) presented a horizontal terminology infrastructure as being composed of five main structural elements or aspects (of which two or more, often can or will be combined, and are or should be institutionalized in order to be effective): (1) terminology (planning) policy; (2) (systematic) terminology creation; (3) information and documentation in the field of terminology; (4) terminology associations (primarily for individuals); (5). purpose-oriented co-operation groupings in private industry or between private industry and public institutions (for the sake of creating and/or sharing terminological data). Similar thoughts were also voiced in 1999 in Greece, within the framework of EPOS (National Programme for Terminological Co-ordination)⁴, where a terminology business plan, human networks for terminology, and a

2. Proposals for an Operational Infrastructure for Terminology in Europe, <<http://www.computing.surrey.ac.uk/ai/pointer/>>.

3. Budin & Wright, 1997, p. 892

4. Valeontis, 2006

national terminology database were part of the programme, and in the EAFT 'Brussels Declaration – for international cooperation on terminology'.⁵ All this and more has served as input for TISS (see 5. below).

3. Preconditions

In contrast to translation and interpreting, however, the importance of terminology is beginning to be recognized. However, there is still a long way to go before industry and society as a whole realize the value of consistent terminology efforts.⁶

In Sweden, several favourable preconditions have influenced the development leading up to the financial state grants for the creation of a national term bank. One of them is a raised terminological awareness, of which there are several indications in Sweden today: Even though they might not always be referred to as such, discussions in the media often circle around the definition of concepts. More and more companies provide glossaries on their websites and use terminology as a marketing means. Lately, several job offers for terminologists have also been published, indirectly indicating a raised awareness. That awareness has boosted the prospects of enlarging the existing terminology infrastructure and making it a truly national one. Another important precondition not to be forgotten is the raised awareness of language in general: articles, prime time TV-shows and advertisements all, and increasingly, refer to language and even terminology. Discussions about the so-called 'domain loss' (the inability to use Swedish within certain domains) and the influence of English on the Swedish language, have been frequent. Attempts at creating a language law stating that Swedish is the official language of Sweden (which is currently not the case) besides the five official minority

5. "The representatives of national and international terminology associations, networks and documentation centres, [...] recognizing the need for co-operation among all actors and stakeholders at the global level and, in particular, to share terminological resources in a co-ordinated way; concerned to strengthen terminology development and dissemination infrastructures, call upon States and governments, intergovernmental bodies and international organizations, and bodies involved in language policies to [...] promote the dissemination and accessibility free of charge of terminologies, above all those contained in official documents of governments and international institutions; support the creation of terminology infrastructures in major economic groupings", <http://www.eaft-aet.net/orb.aw/class=file/action=preview/id=2844/en.pdf>

6. John Graham, Deuterm

languages⁷ were abandoned in 2006, but have now (March 2007) been revived.⁸ More important, however, is the existence of an established national centre for terminology and its activities, as well as other existing terminology networks.

4. A national centre for terminology – the TNC

The Swedish national centre for terminology, TNC, is the hub of Swedish terminology work, and also one of the oldest terminology centres in the world. The Swedish Centre for Technical Terminology (*Tekniska nomenklaturcentralen*, TNC) was founded as early as in 1941 on the initiative of the Academy of Engineering Sciences (IVA) and other interested parties, such as engineers and inventors. In 2001 the ‘old’ TNC became the *Terminologisentrum* TNC (The Swedish Centre for Terminology).⁹

4.1 Important prerequisites

A number of factors could be said to be essential prerequisites for a national centre for terminology; the most important ones include a favourable scientific tradition in the country – Sweden has a strong tradition of systematizing and categorizing, going back to Linnaeus and Berzelius – and strong financial support.¹⁰ The TNC, which started out as a non-profit member organization, later to become a private company without profit distribution, has received financial support from many sources through the years. Today, the main shareholder is the Swedish Standards Institute SIS, but the TNC also continues to receive a substantial grant from the Ministry of Enterprise, Energy and Communications.

Even though the existence of a national terminology centre with a long tradition and a high credibility is crucial in itself, it is certainly not the only factor to be taken into consideration. The importance of timing, raised terminological awareness in society and continued governmental financial support from the Swedish Ministry of Industry, Employment and Communications have already been stressed. To this, we should add the above raised awareness of language in general.

7. Saami, Finnish, Meänkieli, Romani chib, and Yiddish

8. The Language Act (Språklag, 2009:600) came into force on 1st July 2009. Of special importance for terminology is Section 12 where it is stated that: “Authoroties and agencies have a special responsibility for Swedish terminology within their respective domains so that such terminology is accessible, used and developed” (unofficial translation).

9. Referred to as ‘TNC’ in the following.

10. See Bucher & Kalliokuusi, 2000.

4.2 Terminology projects

Another important precondition not to be forgotten is the actual terminology projects; in a way, those projects – often co-ordinated by the TNC but initiated by other parties – have also contributed to raising the level of terminological awareness. Furthermore, new kinds of projects have also contributed to the increased competence of the terminologists themselves; several projects related to information modelling are currently being conducted within the Swedish health care system, for example. Parallel to this development of ‘modern’ terminology work, focusing on new ways of structuring information using new technology, there is also a return to ‘traditional’ projects stemming from a renewed interest in revising existing glossaries (TNC 98: Basic Technical Vocabulary, TNC 46: Glossary of Concrete Terms, TNC 86: Glossary of Geology, etc.) which indicates an awareness of the need for terminology which is up-to-date in many domains. The TNC has also worked with several MLIS-projects, such as TDCNet and especially Nordterm-Net, through which a joint term bank for the Nordic countries was planned and realized. That project in particular laid some important foundations for the creation of a national term bank.

5. The TISS programme (TERMINOLOGY INFRASTRUCTURE FOR SWEDEN) (2002–2004)

As in many other countries, the Swedish public administration is facing a great challenge – that of reinventing itself as a modern so-called e-government, i.e. using new technology and finding new electronic ways of filing, communicating with citizens, etc. That development will call for precise terminology, and different authorities need to start making inventories of existing terminology, harmonizing and co-ordinating their terminology work. This has been further emphasized by several public investigations. Specific people, responsible for terminology work, will also have to be appointed.

With the help of special funding, between 2002 and 2004, from the Ministry of Enterprise, Energy and Communications, the TNC was able to realize the TISS programme, aimed at providing Sweden with a terminology infrastructure. Overall, TISS aimed at following up the results of the Pointer project. It attempted to describe what organization was necessary for a terminology infrastructure, the basic prerequisites, and the needs for such an infrastructure, taking into consideration existing and planned terminological resources in Sweden. This visionary programme spanned a preliminary study (consisting of a number of preparatory investigations) and a main study. Both revolved around a number of key concepts:

development, intensification, inventory, tools, co-ordination, training. More concretely, it included the establishing of networks of people responsible for terminology work in different organizations, and the investigation of national terminological resources – all of it in order to arrive at a web-based terminology portal with a national term bank, *Rikstermbanken*.

More specifically, TISS comprised the following stages:

- A survey on terminological resources.
- The creation of a terminology co-ordination programme which would include people responsible for terminology co-ordination inside all Swedish authorities, agencies and companies.
- The continuation and expansion of the activities of the existing joint groups for terminology, especially the Joint Group for Swedish Computer Terminology, JOGSCOT (see below).
- A pilot study for the creation of a joint terminology group (applied to the terminology of *in vitro*-diagnostics).
- The drafting of terminology training programmes and the preparation of new teaching material, including a translation of the Guide to Terminology into Swedish.¹¹
- The creation of a terminology portal and a national term bank where existing term banks and terminology collections could be made publicly available.

According to the Pointer project, a fully implemented terminology infrastructure would¹²:

1. promote high-quality terminology work,
2. promote the profession of terminologist and joint efforts in terminology education,
3. facilitate access to information on all aspects of terminology and terminology work,
4. distribute terminological resources and support the development of new tools for terminology work,
5. intensify and support standardization efforts.

TISS aimed at achieving some of these goals:

1. work in terminology groups and networks is already a well-established way of working, thanks to 60 years of terminology work at TNC and especially through the joint terminology groups, the first of which, JOGSCOT¹³, was founded in

11. Nordterm 8, ISBN: ISBN 952-9794-14-2

12. Pointer, Final report, <<http://www.computing.surrey.ac.uk/ai/pointer/report/>>

13. Svenska datatermgruppen: Joint Group for Swedish Computer Terminology: <www.nada.kth.se/dataterm>

1996. In this group, domain experts, terminologists, media representatives and general language planners meet and discuss concepts and give recommendations for corresponding Swedish terms, which are later published on a website. Through TISS the working methods of those groups were further developed.
2. through an extensive survey, not only an overview of existing terminological resources was obtained, but also of existing terminology training and the needs for such training and training material.
 3. the idea of a terminology portal perfectly fits this goal.
 4. this portal and the creation of a national term bank, *Rikstermbanken*, will of course be an important and simple distribution mode.

TISS overall contributed greatly to TNC's overview of terminological activities in Sweden but especially the survey showed that, already at a response rate of 22 % (corresponding to more than 440 organizations), there are substantial terminological resources on a national level which could be included in a national term bank.¹⁴ The survey also revealed important facts about the in-house organization of terminology work, the tools used for developing, editing and storing terminology and the most important needs related to terminology work, e.g. training, support, tools. The creation of a new joint terminology group also demonstrated how this kind of terminology network can be organized and function.

6. Rikstermbanken – a Swedish national term bank

A central term bank, the national term bank, should be set up so that access to Swedish terms can be facilitated and that the quality of the terms can be assured. The responsibility for the formation of terms and concepts should be part of the language cultivation work of the [Swedish] authorities, within their respective fields of activity.¹⁵

The idea of creating a national term bank is not new; it was presented in TISS as one of the main ingredients of a terminology infrastructure (see 5. above), as a way of distributing, free of charge, large amounts of terminology in an easy and

14. The web survey was sent out to more than 2 000 organizations of different kinds.

15. IT-propositionen, prop 2005/06: 175: Från IT-politik för samhället till politik för IT-samhället: <<http://www.regeringen.se/sb/d/108/a/47411>>

accessible way.¹⁶ The TNC has indeed been working towards this goal ever since its foundation in 1941: from card register to in-house term bank, '*TNC-bas*'.¹⁷

6.1 Reasons for creating a national term bank

The Swedish Government recognizes that there is a great need for this kind of central terminology repository:

A well-functioning terminology is necessary within all fields of activity if we are to use the fast information flow and communication possibilities of modern society. Terminology work contributes to a well-functioning language within all areas of our society and increases efficiency within, and between various subject fields. The fast development of society requires constant work at creating and making accessible agreed-upon terminologies, within an increasing number of subject fields. Easy access to terms via the Internet in a national term bank endorses such a development.¹⁸

A national term bank like *Rikstermbanken* will

- make for simpler and quicker access to terminology from a large number of domains,
- facilitate the quicker dissemination of new terms,
- raise terminological quality since the contents will be continuously checked, edited and updated,
- make terminology work more efficient by making it easier to re-use existing terms and definitions.

Although *Rikstermbanken* has the potential of being an editing and storage tool (see 7. below), it will, in an initial stage, primarily function as a search and retrieval tool, where answers to different kinds of questions may be found, e.g.

- What terminology is used by different organizations?
- Is there already a definition of a certain concept? Could this definition, with some modification, be used by another organization, in another context?
- What term equivalents are there of a particular Swedish term?
- etc.

16. Inspiration has also come from Lithuania where a national term bank, Terminų Bankas, <<http://terminai.vlkk.lt:10001/pls/tb/tb.search>>, has already been created, by VLKK (Valstybinė lietuvių kalbos komisija: State Commission of the Lithuanian Language. Also the project EuroTermBank has been inspirational; see further <www.eurotermbank.com> and other contributions in this volume.

17. See further Dobrina in this volume.

18. IT-propositionen, prop 2005/06: 175

In addition to those general benefits and uses of a national term bank, several other benefits can be listed, depending on the different potential user groups:¹⁹ *Rikstermbanken* will help domain experts communicate, companies trade with other companies abroad, and, of course, translators in and outside of Sweden who need quick and easy access to up-to-date Swedish terminology. The media, technical writers, standardizers and other language professionals will no longer be in any doubt as to what is the established terminology within a particular field. For terminologists, as well as for officials and legislators inside the administration, *Rikstermbanken* will provide precise definitions of concepts and as such lead to further co-ordination and less duplication of effort – and in the end, to a smoother public administration and clearer-cut communication with the citizens. Although *Rikstermbanken* will not be another web-based general language dictionary, it will act as a contact facilitating tool between citizens and the administration, showing the former other ways of following and participating in the public debate, understanding research results, etc. From the beginning, it was decided that the service would be free of charge for anyone connected to the Internet, which is, of course, an important prerequisite for the above mentioned benefits.

Rikstermbanken will not exist in isolation, but rather form part of a national terminology portal that is the natural starting point for anyone looking for terminology-related information (events, courses, who's who, bibliographies, etc.).²⁰ A separate secretariat located at the TNC will be responsible for all maintenance of the future *Rikstermbanken*.

6.2 Rikstermbanken: Contents

Rikstermbanken should mainly reflect concepts of Swedish society; however, this does not mean that the term bank will comprise Swedish terms only. In order for it to function as planned, the term bank should also contain term equivalents in foreign languages – not only in English but also in various immigrant languages and the official minority languages of Sweden.²¹

Concepts related to Sweden and Swedish society have priority for inclusion in *Rikstermbanken*. This does not mean, however, that *Rikstermbanken* will only contain Swedish terms. Terms in the other Nordic and the major European languages,

19. The different needs of these groups will naturally be reflected in different search options in the interface of *Rikstermbanken*.

20. There are already such portals in use today, e.g. DTP (Deutsches Terminologie-Portal): <<http://www.iim.fh-koeln.de/dtp/>>, and Taalunieversum: <<http://taalunieversum.org/taal/terminologie/>>.

21. IT-propositionen, prop 2005/06: 175

as well as the official minority languages of Sweden will naturally all be included in the term bank as well, although Swedish terms will certainly dominate.

In an initial phase, the glossaries and term collections of TNC will constitute the basis for the term bank, but collection of material from other parties has already started, mainly from official bodies. The survey conducted within the framework of TISS revealed that there are extensive collections with a varying degree of availability, both in public and private organizations. This survey serves as the basis for continued scrutiny into existing term collections which could be included in *Rikstermbanken*.

Although information as to concept – such as definitions and comments – is considered a crucial part of any material to be included in *Rikstermbanken*, other kinds of material with other kinds of contents and structure will also be accepted – the kind of material often known by the names of taxonomies, ontologies, thesauri, nomenclatures, controlled vocabularies, concept catalogues, information models, concept(ual) models, lexicons, dictionaries, glossaries, name lists, etc. There are also plans to include all legal definitions and definitions contained in regulations from various authorities; for this purpose co-operation with computer linguists has been sought, so that tools for automatic term recognition can be developed and used.

On a general level, the material in *Rikstermbanken* should be

- extensive, considering the number of domains covered and the number of terms and concepts,
- representative, i.e. the most important domains should be covered,
- varied, i.e. the terminological material shall preferably be of different kinds and originate from different organizations,
- reliable,
- of high terminological quality.

Furthermore, each term collection should (preferably) fulfil the following requirements:

- it should not be general language, but LSP.
- it should preferably be structured in terminological entries, which should contain some obligatory fields (Swedish term, concept-related information, source, etc.).
- Swedish must be one of the included languages (although exceptions could be made here).
- it should be digitally available.
- any copyright issues should be managed.

The quality assessment of the material will be based on a number of weighted criteria which, in the end, will produce a reliability code of 1–3. To begin with, the

following criteria have been withheld as the most important ones: reliability/provenance/origin, up-to-dateness, originality; contents (number of languages, quality of definitions, etc.)

Legal matters, e.g. copyright issues, have been discussed with legal representatives and agreement templates been drafted. Those content-related issues are still being discussed and the above merely suggests possible solutions that may very well prove to need revision as more material is considered for inclusion.

6.3 Rikstermbanken: Technology

Already during the MLIS-project 'Nordterm-Net', TNC had the opportunity to evaluate existing terminology management systems (TMS).²² Combined with experiences gained from the continuous work of examining various terminology management systems, it was considered that no existing tool would probably be powerful enough for *Rikstermbanken*. Considering that the TNC had been involved in one of the biggest term bank development ventures in years – the development of the new inter-agency term bank within the EU (IATE) since 1999 – the choice of the IATE software as appropriate software to be used also for a Swedish national term bank was a natural one. Through a special grant from Vinnova²³, a formal evaluation of IATE was made possible. With the permission of the EU, the evaluation was conducted during 2004–2005, resulting in the report 'Ja till IATE?'.²⁴ This evaluation, which focused mainly on data entry, import and search facilities, largely showed positive conclusions: The terminological situation within the EU institutions is in many respects similar to that between different authorities in Sweden today, and the business logic of the software seemed especially well developed for those situations. Although a great many of the IATE features could easily be used for a national term bank also, other features were considered slightly too focused on translation or too complicated. Some features also needed to be adapted for the Swedish language. Following this evaluation, contacts were made with the European Commission for further technical co-operation. For various reasons (legal and others), the process was delayed and the TNC decided to opt for an open source-solution instead. The technical development started in mid-2006, towards a solution based on Linux, Apache Tomcat and MySQL among others. On the structural level, the same basic structure of terminological entries as in IATE

22. See further the final report (in Swedish): 'Nordterm-Net på plats' at <<http://www.nordterm.net/info/Nordterm-Netslutrappport.pdf>>

23. Swedish Agency for Innovation Systems: <www.vinnova.se>

24. Translation: 'Yes to IATE?'

has been adopted, i.e. a three-level structure embracing a language independent level, a language level and a term level.²⁵

7. Terminology co-ordination programme

As shown above, creating a national terminology infrastructure is mainly a question of developing a tool (to be used both in the actual terminology work and as a repository for the terminology which is the result of this work), but also a matter of identifying, creating new and enhancing existing co-operation networks for terminology. The other major component in a terminology infrastructure is therefore thought to be a terminology co-ordination programme between different kinds of organizations in Swedish society. Within TISS a model²⁶ for this co-ordination was put forward (see Figure 1).

The model can be applied to terminology co-ordination at two levels: at the organization/authority level or at the national level. On the authority level, each authority should employ a terminology co-ordinator (TC in Figure 1). The task of co-ordinating terminology at the authority may include the following activities:

- maintaining an overview of terminology within the authority
- making inventories of terms and concepts used within the authority
- responding to raised terminological queries
- finding Swedish equivalents for terms in other languages
- establishing internal terminology networks
- participating in network activities related to terminology and language issues
- presenting, publishing and disseminating terminology in an easy and accessible way, e.g. by creating an interactive tool which could provide a single point of access to various terminology collections and which could be used both in the terminology work itself and for the presentation of its results, e.g. a term bank
- documenting internal terminological practices
- recognizing terminological problems within the authority and initiating solutions to those (projects, etc.)
- co-operating with the TNC

25. This is also in accordance with the meta-model presented in ISO 16642: TMF (Terminological Markup Framework). See <<http://www.loria.fr/projets/TMF/>> for further information.

26. Based on a model of conceptual work by Pettersson, 2003.

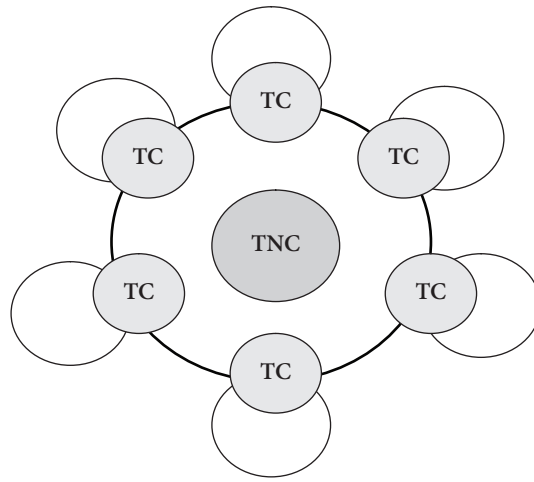


Figure 1. A model for a terminology co-ordination programme (TC = terminology co-ordinator)

A terminology co-ordinator should preferably be interested in terminology and language, but also be officially appointed, since this in itself will vouch for a certain status for his or her tasks. S/he should occupy a central position within the organization so as to have the necessary overview of ongoing terminological activities; most of those activities will probably take place in different working groups made up of experts and terminologists. By way of support for those activities, a kind of terminology council could also be established, in which management staff members could participate. This would ensure that the work is recognized and that the terminology does get used within the organization. This kind of responsibility for each authority has also been stressed in different government bills – and by the governments themselves:

Give the authorities a long term development mission, including responsibility for concepts and terminology.²⁷

Internal co-ordination of terminology between authorities and a mapping of responsibilities for concepts would be of great help when terms and concepts are revised.²⁸

On a national level, it becomes even more important that each authority should be given a formal responsibility for the standardization of terms and concepts within its own domain. Out of the internal networks and TCs of each authority, one could

27. 'Bredtjänster – perspektiv på 24-timmarsmyndigheten', translation: 'Key E-Services – a New Phase in ICT Policy'

28. Swedish Integration Board

then imagine interco-operative terminology networks and joint terminology groups, treating concepts relevant to many, or even all, authorities.

To this model should be added the TCs of private companies and other organizations; today Sweden already has several people calling themselves 'terminology co-ordinator' inside large companies, and there are even co-operative networks for terminology and translation between companies, e.g. within the automotive sector.

At the centre of the model, one finds the national centre for terminology, TNC to which all individual segments are connected and which helps with e.g. training, terminological queries, participation in, and management of terminology projects. In connection to this organizational network, *Rikstermbanken* and its secretariat will constitute an important asset for the work of the terminology co-ordinators, especially when it starts being used not only as a search and retrieval tool, but also as an editing and storage tool. The idea is to also allow different actors to use *Rikstermbanken* for editing and storing their own terminology, i.e. as a full-blown terminology management system. This raises new questions: Should organizations already maintaining a term bank (few today) continue to do so or should they hand it over to *Rikstermbanken*? Should those without a term bank still invest in a proper TMS? And so on. Those questions remain to be answered, and as shown above, the intention of the first phase of realizing *Rikstermbanken* is merely to provide a search and retrieval tool.

The above model can be seen as a generalization of the one used within *Socialstyrelsen* (the National Board of Health and Welfare), one of the authorities with the most elaborately organized terminology work.

7.1 THE NATIONAL BOARD OF HEALTH AND WELFARE – a model authority for terminology work

In order to further promote the idea of a terminology co-ordination programme, a seminar was arranged where representatives from 17 different authorities presented their current situation regarding terminology work. The TISS survey had already revealed interesting aspects of public and private terminology work, but during this seminar, further insights into public terminology work were gained. The situation proved very different within various authorities; some have people responsible for terminology work, others have published glossaries on intranets and on the web, others do not consistently work with terminology yet. At the seminar, the National Board of Health and Welfare presented their internal organization for terminology work and it proved to be a model authority in many respects – not only because it has the right to standardize and officially recommend terminology within the healthcare and social sectors, but also because of its elabo-

rate internal organization. In 2000, a separate entity (EpC) first started employing terminologists, alongside several other, inter-related entities: different working groups (where the actual terminology is developed), a language group and a terminology council.²⁹ Several measures have also been taken, to ensure the quality of the terminology work done within the organization: hundreds of employees have been given terminology training (in collaboration with the TNC), a term bank has been developed, and a handbook of terminology work been published. The importance of conducting a kind of ‘terminological consequence analysis’ has also been stressed from the top down. According to the National Board of Health and Welfare, a terminology should:

- reflect the knowledge of the field
- be as independent of aspects specific to the organization as possible
- be in accordance with the terminological usage on regional as well as local levels
- be based, in its elaboration, on terminology standards and established methods for terminology work.

The whole idea of a terminology co-ordination programme might seem easy enough in theory. In practice, however, it will take time, financing, tools, and training (special courses, specially adapted training material).

8. Training needed!

As shown above, there is a current interest in employing terminologists, and in all the new areas where terminology work becomes important – quality assurance, content management, knowledge processing, etc. – there is a clear need for competent people who know the basics of terminology work. The problem, however, is that they are few and far between. Terminology training was, and still is largely lacking from many higher-education level programmes in Sweden. Most of the current training is given by the TNC in both academic and non-academic contexts, either as separate lectures as part of longer curriculums (translators, interpreters, etc.) or as tailor-made courses for other groups (companies, trade associations, technical committees, etc.). Recently, the subject of terminology was introduced as a single subject course at TÖI and later developed into a web-based course.³⁰

29. EpC, enheten för klassifikation och terminologi; translation: Centre for Epidemiology, classification and terminology unit, <<http://www.sos.se/epc/>>

30. The Institute for Interpretation and Translation Studies, <www.tolk.su.se>

Currently there are also joint Nordic plans for a Nordic master programme in terminology.³¹ But more training is needed, as is new training material.³²

9. How did the TNC get this far?

The current state of affairs can be seen as the result of several conscious steps. The TNC has succeeded in getting the idea across and getting the continued governmental financial support (to develop TISS and *Rikstermbanken*) by constantly stressing the importance of terminology work, harmonization and the advantages of a shared national repository for terminologies. All this has resulted in those ideas being put forward in important documents (several government bills) which in turn helped to disseminate the ideas. This, not only through participation in relevant seminars and projects, but through continuous teaching and training, old and new collaborations – and thanks to suitable timing (the raised level of terminological awareness in society and especially in government agencies). Throughout the development of *Rikstermbanken*, co-operation with the European Commission and their IATE team proved pivotal. At a later stage, co-operation with the project EuroTermBank, as well as with the Lithuanian institutions which developed the Lithuanian national term bank, also proved very useful.

10. Conclusion

The final words of this paper are not about finality; *Rikstermbanken* has barely been launched³³ and many things remain to be done, before the national terminology infrastructure is fully realized:

- The quest for material to be included in *Rikstermbanken* needs to continue.
- The quality criteria for this material need to be developed further.
- Technological issues and availability issues related to *Rikstermbanken* (RTB) need to be dealt with.

31. A first on-line course (Terminology I, covering 7.5 ECTS) was started by the network TERMDIST in September 2009, and currently (November 2009) some 50 students are enrolled. See further <www.termdist.no>.

32. It could be mentioned here that a Swedish version of the Nordterm handbook Guide to Terminology (Nordterm 8) was produced during TISS.

33. *Rikstermbanken*, the national term bank, was officially inaugurated on 19th March 2009. Currently (November 2009) it contains almost 60 000 terminological entries (about 250 000 terms) from 300 various sources and with terms in some 20 languages. See further <www.rikstermbanken.se>.

- Plans for the continued financing of the updating and maintenance of *Rikstermbanken* need to be further developed.
- The organization of the *Rikstermbanken* secretariat needs to be settled.
- Ways of realizing the planned terminology co-ordination programme need to be found.
- The training of terminologists needs to continue and be expanded to more groups.
- The terminological awareness needs to be further raised: influencing the Government, participating in conferences, – briefly, spreading the term(s)!

References

- Bucher, A.-L. & Kalliokuusi, V. (2000). How to survive after many years as a national terminology centre? Views from two Nordic countries. In *Conférence sur la coopération dans le domaine de la terminologie en Europe*, Union Latine. ISBN 92-9122-005-1.
- Budin, G. & Wright, S.-E. (eds.) (1997). *Handbook of Terminology Management. Vol. II: Application-oriented terminology management*. Amsterdam: John Benjamins. ISBN: 90-272-2155-3.
- Galinski, C. (1998). Terminology infrastructures and the terminology market in Europe. In *Trans*, Internet-Zeitschrift für Kulturwissenschaften, September 1998, <http://www.inst.at/trans/0Nr/galinski.htm>, 2007-03-30.
- Joint Group for Swedish Computer Terminology: www.nada.kth.se/dataterm
- Nilsson, H. (2005). TISS & IATE. Svensk terminologisk infrastruktur och svensk rikstermbank. In *Nordterm 2005 Ord og termer*. Reykjavík: Íslensk málstöð. ISBN: 9979-842-85-7.
- Nilsson, H. (2004). Terminology work – the Swedish way. In *Eesti oskuskeel 2003*. Tallinn: Eesti Keele Sihtasutus. ISBN 9985-79-074-X.
- Nilsson, H. (2000). One-stop shop – two-aim game. Improving the availability and quality of terminology in the Nordic countries through Nordterm-Net. Nordic Term bank Services via Internet. In *Conférence sur la coopération dans le domaine de la terminologie en Europe*, Union Latine. ISBN 92-9122-005-1.
- Pettersson, R. (2003). *Ord, bild & form: termer och begrepp inom informationsdesign*. Lund: Studentlitteratur. ISBN 91-44-03177-7.
- Pointer, *Final Report*, (DIT, 16/02/96), <http://www.computing.surrey.ac.uk/ai/pointer/report/>, 2007-03-30.
- Rikstermbanken, national term bank for Swedish terms www.rikstermbanken.se
- Swedish Centre for Terminology: www.tnc.se
- Valeontis, K. (2006). *EPOS in brief (National Programme for Terminological Co-ordination)*. Unpublished presentation.

SECTION II

Best practices in terminology management

Terminology on demand

Maintaining a terminological query service

Claudia Dobrina

With all due deference to the emerging e-society and its needs, the acquisition and exchange of knowledge in a plain human-to-human way, remains highly relevant. In these Google times, when you can wash out a plenitude of information with just a few clicks, many still prefer to ask a professional terminologist for help with their terminological problems; at least in Sweden, where a terminological query service operated by TNC, the Swedish national centre for terminology, has been providing users with terminological information on demand since the 1940s. This paper gives an overview of TNC's terminological query service and discusses the sine qua nons for its effective functioning. It also describes query processing procedures and challenges that terminologists face in their aspiration to meet users' terminological needs.

Keywords: terminology, subject field, terminology database, terminological query service, query, response, query processing procedures, terminologically reliable resource, terminological awareness, national centre for terminology

1. Introduction

A telephone rings. A terminologist on duty answers:

- *Terminologicentrum*, TNC (The Swedish Centre for Terminology)
- Can you help me with a definition?
- Yes, we help with all kinds of terminological problems.
- It is about photovoltaics, the subject field is energy transformation.
- Good. I'll get back to you as soon as I am ready with the response.

That's how it all starts. A telephone call, an e-mail, or a web request arrives at TNC, the Swedish Centre for Terminology, announcing some urgent terminological need, and the wheels of TNC's terminological query service start turning. The requested information is sought, hopefully found, a response is compiled and (within 24 hours) delivered to the enquirer. The summary of the query and

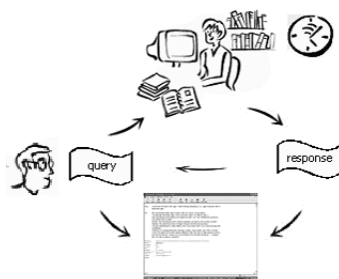


Figure 1. TNC's terminological query service in a nutshell

the response are stored in TNC's internal terminology database to be used for future queries or other terminological undertakings.

TNC's terminological query service (henceforth *the query service*) has been offering solutions to its users' terminological problems for more than 60 years (starting soon after TNC's foundation in the 1940s). The explosive growth of freely available electronic terminological and other knowledge resources notwithstanding, the query service is still going strong. Why it still is, how it functions today, what challenges the query service staff faces and how they are dealt with are the issues touched upon in this paper. The query processing procedures are discussed in more detail.

2. TNC in the service of terminology

The query service is just one of a wide range of products and services offered by TNC to users who need information and assistance with issues relating to terminology and knowledge organization. TNC's activities are centred on the following main areas:

- Compilation of terminological and special language resources: terminological vocabularies, terminology databases, style manuals for technical writing, etc.
- Teaching terminology theory and terminology practice: at universities, as open lectures, through custom-made courses for project teams, companies, etc.
- Consulting services varying from operating the query service to leading long-term terminology projects, participating in language planning activities, development of ontologies, information modelling, etc.
- Terminology standardization on the national and international level
- Promotion of the development and use of terminology in all spheres of professional activities and public life through participation in international and national terminology networks.

The query service together with TNC's other small-scale consulting services is accessible on a subscription basis. Non-subscribers can choose to be billed. First-time users are usually helped free of charge, provided their queries do not require extensive investigation. TNC also receives a grant from the Swedish Ministry of Enterprise, Energy and Communications aimed to support TNC in its function as a national centre for terminology and special languages.

The whole of TNC's terminological activities and the infrastructure created to support them contributes greatly to the effective operation of the query service and ensures the high quality of terminological assistance the query service renders its users. A long-standing effort of raising terminological awareness among subject-field experts, language professionals and the general public brings about a never dwindling stream of queries.

3. The query service: An overview

Services and products should, as is generally known, be developed to meet users' needs. Terminological services and products are no exception to this rule. Terminological vocabularies are compiled to meet the needs of subject-field experts for a comprehensive and harmonized terminology. Terminology databases are created to provide easy access to terminological information on a wide range of subjects. Terminological consulting services are to help with all kinds of terminological problems. The query service, this terminological 'emergency ward', has emerged to meet urgent and not too vast terminological needs.

3.1 Users

Query service users stand somewhat apart from consumers of TNC's other services. An employee of a large company, a freelance translator, a university student or a librarian contacting the service usually seeks a solution to his/her own terminological problem but the information produced by the query service in response to a query usually benefits a wider community. The query service users live in Sweden and abroad; they practise a variety of professions and have a variety of terminological needs. In the past, the majority of users had backgrounds in science and engineering. Not any longer: humanities and social sciences, economics and management are now widely represented. When, in the middle of the 1990s, Sweden joined the European Union, Swedish translators of the European Union's documents became avid users of the query service. And now subject-field experts and government agency officers, students, and teachers, journalists and the public at large are among the users' ranks.

3.2 Queries

Queries posed to the query service show a great diversity in many respects: as to type of terminological information requested, languages involved, subject fields covered, degree of complexity, purposes (in which way the information to be produced is going to be used), etc.

A brief account of queries viewed in various dimensions follows.

3.2.1 *Types of queries*

The following main categories of queries can be distinguished:

- Concept-related queries

Information requested: definitions, concept descriptions, explanations of the difference between two or more related concepts, etc.

Example

A construction engineer asks for a definition of *golvbjälklag* ('floor joist'). An employee of the National Food Administration wants to know the difference between the concepts designated by the terms *teknisk* ('technical') and *teknologisk* ('technological').

- Term-related queries

Information requested: term equivalents, new terms to designate a new or a borrowed concept, synonym differentiation, etc.

Example

A journalist working on an article for a popular science journal requests a Swedish term to match the English *e-science*. A translator looks for an English equivalent of the Swedish *gyllene standard* ('gold standard').

- Language/style queries

Information requested: grammatical forms, etymological information, abbreviations, word formation, etc.

Example

A subject-field expert wonders why 'k' in the abbreviation for *kilowatt* (kW) is written in lowercase, unlike TW (*terawatt*) and GW (*gigawatt*). A senior citizen asks for the plural form of *medium* in Swedish.

Matching users to types of queries is not an easy task, but some preferences can nevertheless be distinguished: translators are prone to be more interested in term equivalents, subject-field experts in definitions, journalists in new terms.

3.2.2 *Subject fields concerned*

A query may concern practically any subject field: engineering, medicine, law, chemistry, construction, management, etc. Of some 6 000 queries recorded in TNC's query database (more about that below), the largest group is constituted by queries relating to the field of information and documentation, closely followed by queries concerning life sciences. The number of queries dealing with new and rapidly evolving subject fields with an explosive development of terminology certainly exceeds that of queries relating to old established domains with a more stable terminology.

3.2.3 *Languages involved*

Swedish is represented in one way or another in the majority of queries. English overwhelmingly dominates the multilingual queries concerning, e.g. translation equivalents, but also monolingual queries requiring information on English definitions or English style rules. German, French, Danish and Finnish come next, Italian, Russian, Icelandic and a number of other languages are only occasionally requested.

3.2.4 *Origins and frequency*

The subjects of some queries originate from current events: in January 2005 one user wondered whether the Swedish media's indiscriminate use of the terms *floedvåg* ('tidal wave') and *tsunami* had any basis. The Swedish equivalent of *wild game birds* was sought in the winter of 2006, when the bird flu menace was front-page news. Other queries have to do with older issues like the difference between *skruv* ('screw') and *bult* ('bolt') in Swedish, which remains an issue, despite the devices' long history. Between 1991 and 2001, a f.a.q(uey) concerned the Swedish equivalent for *e-mail* (before 1991 the Swedish users probably had no scruples about accepting the English term).

3.2.5 *Complexity*

Some queries can be solved within minutes, e.g. advising on the plural form of *skanner* ('scanner') in Swedish, others – actually, the majority of queries – may call for an extensive investigation. For example, finding a Swedish equivalent for the English *gender* proved to be a rather demanding task given the rapid development in the subject field of gender studies and researchers' conflicting opinions on the subject. Some queries drive a terminologist into a theoretical jungle, like addressing the issues of dimension criteria or types of referents and queries where a pragmatic solution is the most appropriate.

3.2.6 *Queries today and yesterday*

A vast collection of old query records in TNC's archive leads to the conclusion that no dramatic change has occurred over the years regarding users' terminological needs. Users are still after the same types of terminological information: definitions, term equivalents, new terms, orthography information, etc. In 1952 a user asked TNC to veto the use of the English abbreviation *TV* in Swedish (the abbreviation is still used, and the poor soul's hopes have been crushed). TNC's staff still has to occasionally explain that TNC does not act as a 'language police force'. The number of subject fields concerned in present-day queries is, as already mentioned, much bigger than in years gone by and multilingual queries now prevail. Complicated queries clearly dominate, which may be viewed as an obvious result of the growing availability and accessibility of web resources (Sweden is considered one of the most computerized countries in the world). People obviously cope with less complex terminological problems without external help.

4. Meeting users' needs: who and how

The diversity and complexity of information requested by users, a plethora of subject fields to consider, a vast amount of data to sieve through, etc. presents the query service with many challenges. The main challenge, however, lies in providing users with high-quality terminological information within short time limits and without jeopardizing the financial feasibility of the service.

TNC meets this challenge through the following accomplishments:

- extensive competence and experience
- reliable terminological resources
- established procedures.

4.1 Competence

The competence required for coping with users' requests has been accumulated through many years of involvement in a very wide range of terminological undertakings (from running huge terminological projects like complementing a European term bank with 140 000 Swedish terms, defining a hundred basic concepts in the field of terminology to information modelling projects, preparing terminology standards, etc).

The following types of competence are regarded as necessary for the query service to function effectively:

- Expertise in terminology
- Command of query processing procedures

- Mastery of search methods
- Linguistic knowledge and skills
- Subject-field expertise.

The first four belong to the in-house competence. There are also a number of subject-field experts on TNC's staff: mathematicians, biologists, translators, a computer engineer, an information and documentation specialist, but this is by no means sufficient for coping with all the subject fields a query may span. Long experience has given the query service staff the ability to quickly familiarize themselves with a query's subject field, but if that does not suffice, TNC's wide network of subject-field experts may be contacted. These include participants in TNC's many completed terminological projects, language planning experts, members of the Swedish joint groups for terminology, etc. The joint groups for terminology (JGT) are associations of subject-field experts and terminologists engaged in compiling the new terminology emerging in their respective subject fields. There are at present four JGTs in Sweden which cover the subject fields of computer terminology, building and construction, life sciences and optics. TNC is taking an active part in each of the JGTs.

4.2 Terminological resources

TNC's wide collection of terminological resources, both printed and electronic, has been amassed to be used in various terminological undertakings. In query processing, three main types of terminological resources are used:

- TNC-bas (TNC's internal terminology database) containing over 250 000 terminological entries, including terminology from terminological vocabularies compiled by TNC and other terminology centres in the Nordic countries, excerpts from Swedish standards, collections of terminology compiled by the JGTs, etc. It also includes TNC's query database of nearly 6 000 queries starting from 1971.
- TNC's reference library containing around 10 000 volumes: terminological vocabularies, dictionaries, handbooks, monographs, catalogues, the query service archives starting from the middle of the 1940s, etc.
- Web resources, in the form of a collection of links to special language dictionaries, terminology databases, etc.

4.3 Query processing procedures

The stages of query processing are much the same as they were at the dawn of the query service. The used search methods and resources, however, have greatly

evolved since the days when performing a search was limited to looking up library sources and consulting subject-field experts. Hand-written and machine-typed query records in the query service archives provide some very instructive insights into the terminological plights of the past.

Query processing, starting with the arrival of a query, includes the following stages:

- Reviewing a query
- Searching for information
- Preparing and delivering a response
- Storing the information for future use.

4.3.1 *Reviewing a query*

When reviewing, a query a terminologist has to check if the information delivered by a user is sufficient, if the subject field has been indicated, if the purpose for which the requested information will be used has been specified, etc. A number of administrative issues must also be settled, like checking the subscription status of the enquirer, discussing the pay mode with non-subscribers, etc.

4.3.2 *Searching for information*

Given the number of terminological resources which may be necessary to consult and the time-constraints (a maximum of 24 hours for the response delivery) it is important to start with determining search priorities. If a query concerns, e.g. the genitive form of a Swedish initialism, *Skrivregler för svenska och engelska från TNC* (TNC's Style Guide for Swedish and English) will be the first source to be consulted. A Swedish equivalent for the English term *clothing for light use* is more likely to be looked for in *Kompass*, an extensive electronic product directory or in *Textilordlista* (Glossary of textiles), one of TNC's own terminological vocabularies.

The general rule is that the most reliable terminological sources should be consulted first. The majority of the query service searches therefore start with consulting *TNC-bas*. If no quick remedy is found, there are other alternatives, such as the reference library, web resources and colleagues who may have considerable expertise in a particular issue, etc. The assistance of subject-field experts is sought for queries belonging to very special domains and for complicated issues requiring profound subject-field knowledge. A Google search might also be of use, e.g. to check the occurrence, frequency or acceptability of requested terms in the respective subject field.

Another challenge at this stage, besides the number of resources to be consulted and the time limits, is the varying terminological reliability of the available resources. The quality of the query service output largely depends on the degree of the terminological reliability of the resources used. In the context of query

processing, a resource is considered terminologically reliable if it meets the following principal requirements:

- it has precise and coherent terminological content, built up on the basis of both subject-field and terminological expertise
- it is up-to-date (in older resources, especially those concerning rapidly evolving subject fields, the most recent terms are not captured and the definitions may prove outdated).

Unfortunately, not many resources satisfy both requirements. TNC's own meticulously compiled terminological vocabularies easily meet the first requirement, but many of them have been around for quite a time and the information they contain has to be checked by consulting more modern resources. Precision, homogeneity and coherence of terminological information found in terminological vocabularies compiled on the basis of terminological methods are often in short supply in other resources. Dealing with many up-to-date web resources requires great caution, since their content 'spans the spectrum from the good, to the bad, and extending to the downright ugly' (Pacifi 2002). This is why both old and new resources usually have to be consulted in query processing.

The reliability of terminological resources used in query processing is evaluated both on a permanent basis and in the process of a particular search for a particular query. That is actually how TNC's extensive collection of web resources mentioned in 4.2 came into existence.

Notwithstanding the plethora of available resources, there are still cases when a search brings no result. This may be because the information requested is very specific, e.g. it may be about some very peculiar concept in a subject field not even the web search could help with, or no consultable expert on the subject field in question may be found, or the information requested concerns a definition of a new concept in an emerging subject field and experts are yet to reach a consensus on the issue.

4.3.3 *Preparing and delivering the response*

This is by far the most challenging processing stage where new terminological information often has to be created or some ingenious solution to a complicated problem suggested. But, as Cabré (2002: 157) resignedly observes when discussing a terminological query service: 'Not all questions posed have a solution'.

This stage comprises a number of other steps, not forgetting that responding to queries is an iterative process and many shifts between the steps as well as between this and the preceding stage may take place before the response is ready to be delivered. The following steps can be distinguished:

- a. Inventorying the results of the search

- b. Evaluating the quality of the information obtained
- c. Modifying and complementing the information obtained
- d. Compiling a response report
- e. Preparing the supplementary information
- f. Delivering a response

The sequence of steps may vary depending on the outcome of the search: if the search produces no results the terminologist will proceed directly to step d.; if the information obtained does not require any further processing, step c. is left out; if no supplementary information has to be prepared, step e. is superfluous.

In case of complicated queries concerning, e.g. concept differentiation or finding a Swedish term for a new concept, a terminologist on duty usually asks colleagues for assistance. A brainstorming session, say, during a coffee break seldom fails to produce an acceptable solution.

A. *Inventorying the results of the search*

If the search has produced no results, the terminologist proceeds to compile a response report (step d.). A response should always be delivered, whatever the search's outcome. The more information is obtained in the search, the more work is required at this stage and it is important to select the information which best fits the enquirer's need. For example, if a request for a definition is made by a layman who merely wishes to understand what the concept is about, a 'lighter' version of a definition (e.g. with less intension depth) would be selected, while a subject-field expert would be offered a more cogent definition. In case of multiple choices, a terminologist has to select the most appropriate solution, e.g. the best equivalent in another language, the most fitting term candidate to designate a new concept or the best-formulated definition of a concept (see also b. below).

Example

A user requests a definition of the concept designated by *rabitzvägg* ('Rabitz wall'), subject field: construction. The search produces no exact match. The terminologist, however, does find definitions of *rabitz* ('plastering') and *rabitzputs* ('plastering on metal lathing') in one of the construction vocabularies in the library. On analysing the information contained in the vocabulary entries, the terminologist is able to provide an explanation of the concept *rabitzvägg*.

A translator asks which of the two existing Swedish equivalents *viltlevande fågelvilt* or *frilevande fågelvilt* best matches the English term *wild game birds*. A terminologist analyses the concept designated by *wild game birds* and discovers that two

different Swedish terms (*hägnat vilt* or *frilevande vilt*) correspond to the English term, depending on the context in which the latter occurs.

B. *Evaluating the quality of the information obtained*

Each type of terminological information to be delivered should meet terminological requirements which have been elaborated for those types. For example, an intensional definition is considered acceptable if it includes a superordinate concept, lists the delimiting characteristics, is not too narrow or too wide, is not circular, etc. A new term should be precise, unambiguous and transparent; it should fit into the Swedish language system, etc. A term equivalent in a target language should refer to the same concept as a source language term, etc.

If the information obtained in the search meets the above requirements a terminologist proceeds to step f. Otherwise (s)he has to go through step e.

C. *Modifying and complementing the information obtained*

If the information obtained in the search is to be modified or complemented, a terminologist may attempt another search (this time with modified parameters) or draft the suggested modification and check it with a subject-field expert.

Example

A new product called *energy bar* in English makes its appearance on the food market. A user wonders if there is a matching Swedish term. The search shows that several alternatives are being used, none of them satisfactory. A terminologist sums up the results of the search and discusses the problem with colleagues. A smart term proposal (*energikaka*) ensues, which soon finds its way into the media and general use.

D. *Compiling a response report*

If the search has produced no results, the fact is indicated in the response report, complemented (if possible) by an explanation why the information has not been found, as well as a recommendation on a practicable course of action.

Example

A translator asks for a Swedish equivalent of a French term in the field of mechanics. There seems to be no established Swedish term in use. A terminologist suggests giving a short description of the concept in Swedish in the enquirer's text instead.

If the search has been successful, a response report includes:

- the requested information which hopefully matches the user's need¹,
- a brief account of how the information was obtained (if deemed necessary),
- argumentation in support of a terminologist's decision in case of multiple choices,
- examples of usage, relevant considerations, recommendations, etc.,
- a list of references.

Example

A user asks for an opinion on her proposal for a new Swedish term *prolongivitet* to match the English *prolongivity* (subject field: medicine). In her response report, a terminologist recommends to use the already existing term *livsförlängning* instead. Her arguments are that *livsförlängning* is more explicit than *prolongivitet* and that the introduction of a new term may lead to unnecessary confusion. The terminologist also mentions that there is a synonym of *livsförlängning* in Swedish, namely *livstidsförlängning*, which is rather common. The list of references points to a number of web resources as well as discussions with colleagues.

E. *Preparing the supplementary information*

A response report may be complemented by supplementary information to support a terminologist's suggestion or to give a better overview of the terminological problem investigated. The supplement may include concept diagrams, copies of relevant documents, illustrations, contexts, a copy of correspondence with an enquirer or subject-field experts, etc.

F. *Delivering a response*

A response report is delivered orally or in written form. Non-subscribers are usually encouraged to consider subscription to TNC's consulting services and referred to TNC's website for information on the conditions. Feedback from users on a delivered response, more often voicing gratitude than disappointment, may follow, thus concluding this lengthy processing stage.

1. Users' needs vary from person to person. Of several users asking for a Swedish equivalent of a term in another language one would be satisfied with the information that no such equivalent exists or that there is a borrowed term already in use, another wants a new term to be suggested and a third would ask if TNC cannot advocate the use of a term (s)he personally prefers, instead of some other widely established term. Users' belief in TNC's absolute authority in matters of Swedish special languages is sometimes quite touching.

4.3.4 Storing the information for future use

The information acquired and created in query processing is a valuable asset for future queries and other terminological ventures. All the queries, together with response reports and the supplementary information, are therefore stored in TNC's query database and the query service archives. Summaries of queries and response reports and the pertinent administrative information are entered into a special query record format. A copy of supplementary material delivered together with a response, is stored on paper medium. At intervals of one to two months, the latest queries processed are printed out and circulated to all terminologists at TNC for review and comments. The query material is then revised on the basis of the comments, converted and entered into the query database (see an authentic example of a query record in Figure 2 and a translated one in Appendix).

TNC's query database is not open to the public, but a selection of the most interesting queries and responses is regularly published on TNC's website.

The screenshot shows a web-based interface for a query record. On the left, a list of labels is connected by lines to specific fields in the record:

- Question** points to the 'Fråga:' field.
- Response** points to the 'Svar:' field.
- References** points to the 'Referens:' field.
- Comments** points to the 'Kommentar:' field.
- Date** points to the 'Datum:' field.
- Subject field** points to the 'Fackomr:' field.
- Responder** points to the 'Svarare:' field.
- Response time** points to the 'Tid:' field.
- Enquirer** points to the 'Förstag & d:' field.

The record data is as follows:

Fråga: Jag undrar vad den engelska termen "triple play" har för svensk motsvarighet? Det handlar om Internet, TV och telefoni genom samma anslutning.

Svar: I tidningen Computer Swedens ordlista anges "trippeljänst" som ekvivalent till "triple play" med följande förklaring: abonnemang på Internet, tv, radio och telefoni från en operatör och genom en bredbandsanslutning. Jag har också diskuterat frågan med mina kollegor och vår gemensamma uppfattning var att "trippeljänst" var OK.

Referens: Computer Swedens ordlista, diskussion vid fikabordet, UH.

Kommentar: Karin visste mycket väl att DI inte hade abonnemang, hon hoppades dock på att få svar därför att "frågan var så aktuell". Hon var mycket nöjd med förslaget: "Det var väl en mycket bra översättning! Den ska vi använda" (ett citat ur Karins svarbrev). Frid och Frid men två dagar senare hittade Ulfson ett e-brev från Anders Lotsson om "triple play" där denne föreslog "trippeljänst" och "trippelanslutning".

Datum: 2005-05-09

Fackomr: inf

Svarare: KD

Tid: 30

Förstag & d: Frågare: E-post:

Figure 2. An example of a query record in TNC's query database

5. Conclusion

With all due deference to the emerging e-society and its needs, plain person-to-person knowledge transfer continues to offer many advantages. After all, a terminological query service is but a win-win knowledge transfer situation that benefits the participants as well as a range of other interested parties: enquirers' terminological problems are solved, responders sharpen their terminological wit

and augment their knowledge, terminological information in the form of new terms, definitions, etc. accumulates, and terminological know-how widens. The terminological information acquired through TNC's query service's efforts is recorded and re-used in many ways thus reaching a wide public. New terms made up in response to queries enter the Swedish specialized languages and contribute to their development. For example, Swedish terms for park-and-ride (*infartsparkering*), plastic (*plast*), air bag (*krockkudde*), home page (*ingångssida*) and many others were once suggested in response to queries.

A terminological query service together with other professional terminological activities contributes to the creation of high-quality terminology, to the promotion of terminological awareness and thus to better professional communication.

References

- Cabré, M.T. (1999). *Terminology: Theory, Methods and Applications*. Amsterdam & Philadelphia: John Benjamins.
- Dubuc, R. (1997). *Terminology: A practical approach*. Brossard, Québec: Linguattech éditeur inc.
- ISO 704 (2000) *Terminology work – Principles and methods*.
- Pacifici, S. (2002). *Verifying sources on the net* <http://www.llrx.com/features/verifying.htm>.
- Pavel, S. and Nolet, D. (2001). *Handbook of terminology*. Québec: Public Works and Government Services.
- Suonuuti, H. (1997). *Guide to terminology*. Tekniikan Sanastokeskus. Helsinki (Nordterm 8). *The Pavel Terminology Tutorial*. http://www.termiumplus.translationbureau.gc.ca/site/didacticiel_tutorial_e.html.

Appendix

Example of a query record in TNC's query database²

QUESTION	I've discovered that quite a number of Swedish terms are used as equivalents for the English <i>clear GIF</i> (subject field: IT): <i>webbsignal</i> , <i>åtgårdsmärke</i> , <i>en-pixel gif</i> , <i>webbfyr</i> , etc. Which of them would TNC recommend?
RESPONSE	There is no established Swedish term for the concept designated by <i>clear GIF</i> (<i>single-pixel GIF</i>). If there are several synonyms used, none of which can yet be considered an established term, it is reasonable to choose the one which gives a better intimation of a concept in question. I find <i>webbsignal</i> more comprehensible than <i>webbfyr</i> or <i>åtgårdsmärke</i> and would therefore recommend it. To support my point I'd like to refer to <i>Paginas IT-ordbok</i> (Pagina's IT-dictionary) on the web, where <i>webbsignal</i> is listed.
REFERENCES	http://www.pagina.se/itord/default.asp?Id=4634 , discussion with colleagues
DATE	2002-06-11
SUBJECT FIELD	IT
RESPONDER	UH
RESPONSE TIME	30 min
ENQUIRER	
AFFILIATION	
EMAIL	
REVIEWED BY	MG

2. Translated from Swedish and somewhat adapted.

Frames, contextual information and images in terminology

A proposal*

Mercedes García de Quesada and Arianne Reimerink

In this paper, it is our aim to analyse and evaluate possible applications of the Berkeley FrameNet project (Fillmore et al., 2003) and its Spanish counterpart (Subirats, 2004, 2007; Subirats and Petruck, 2003) to recent advances in terminographic definitions, context and visual representations (Faber et al., 2006; Montero and García, 2004). In our proposal, following a typology of visual contexts (Cook, 2006; Mayer 2001) within the domain of Coastal Engineering, images and texts are manipulated and grouped together following a frame-like structure, where a specific syntactic realization of a Frame Element (FE), within a given frame, is explicitly associated to an image (or part of it). As a result, we will have information about the syntactic contexts in which the term occurs, and the semantic properties of the key word's syntactic companions, together with information about the membership of the term in classes of semantically similar words (Atkins, Fillmore and Johnson, 2003).

Keywords: FrameNet, context, definition, multimodal corpus

1. Introduction: Contextualizing context

The notion of context has been extensively debated in disciplines such as pragmatics, linguistic anthropology or Artificial Intelligence (AI), among others. Despite its complexity, context can be regarded as encompassing external (situational and cultural) factors and/or internal, cognitive factors, all of which can influence one another (House, 2006: 342). In fact, although it has been conceptualized in a

* This research is part of the projects Spanish FrameNet: a lexical resource for automatic semantic language processing, FFI2008-00875, funded by the Spanish Ministry of Science and Innovation and MARCOCosta: Multilingual knowledge frames in the integrated management of coastal areas, PO6-HUM-01489, funded by the Andalusian Regional Government (Spain).

variety of ways in different disciplines, in most of the approaches context is regarded as dynamic rather than static; all of them share a common view of language as a kind of action, where the meaning of the linguistic forms is fully understood in their use.

Lexicography and Terminology have not lagged behind current thinking, since the fields of knowledge and research are also explicitly or implicitly engaged with a dynamic and data-driven notion of context. A semasiological approach has been recently advocated for terminography, more in the line of modern-day Lexicography (Cabr , 1999ab). Far from a word-driven notion that excludes any conceptual analysis, many authors conceive corpus-based terminology as an integrated top-down and bottom-up approach, in which the texts of the corpus bring terminologists closer to the underlying conceptual structure and help them develop terminographic resources based on the underlying conceptual structure on the one hand and the syntactical and collocational behaviour of the terminographical units in the texts on the other.

In the approach taken in this article, following Mitkov (1998, 2002) and Meyer (2001), we distinguish between so-called *knowledge poor contexts* (KPCs) and *knowledge rich contexts* (KRCs). Using FrameNet terminology, where the semantic arguments of a predicating word correspond to the Frame Elements (FEs) of the frame or frames associated with that word (Johnson and Fillmore, 2000), KPC are those in which none of the FEs, be they core or non-core FEs, are lexicalized, nor any of the siblings or neighbour lexical units. These contexts do not provide any domain knowledge.¹

KRCs are contexts which indicate at least one item of domain knowledge that could be useful for conceptual analysis. In other words, these contexts should indicate at least one conceptual characteristic, whether it be an attribute or a relation (Meyer, 2001: 281). In FrameNet, such KRCs include at least one core FE. However, some KRCs are richer in domain knowledge than others. For example, a meaningful context is one that includes at least one item of domain knowledge. A defining context, on the other hand, is a context which includes all or most items of domain knowledge that are necessary to understand a concept.

In the termbase, it is in the description of lexical units and their underlying conceptual frame that the notion of knowledge rich context intermingles with that of definition. We conceive those KRCs as definition-like metalinguistic utterances, based upon multimodal corpora, which may well include images, audio and video

1. The coreness of an FE is established in terms of how central it is to a particular frame. For more detailed information on FE coreness see Ruppenhofer et al. (2006).

data.² It is our aim to go beyond the well beaten track of *a priori* structured terminographic definitions and call for a more dynamic, customizable notion which would also comprise other metalinguistic data, together with visual and audio information.

In the following sections, a brief summary is given of the research carried out by the Berkeley FrameNet (FN) and FrameNet related projects. Then our current terminological research on Coastal Engineering is touched upon. Thirdly, we will show how the methodology and lexical resources created by the FN project can be applied to terminology. Finally, we will show how multimodal corpora can be used in definition construction.

2. Recent research in lexical semantics: FrameNet and Spanish FrameNet

In a special issue of the *International Journal of Lexicography* devoted to the FN project, Fillmore, Johnson and Petruck (2003: 235–250) claim that the linguistic basis for Frame Semantics (FS) can be traced back to Fillmore's theory of Case Grammar (Fillmore, 1968), where the underlying idea is that syntactic deep structures are best expressed as configurations of 'deep cases'. The theory has evolved since then and semantic roles have been substituted by frame elements, designated in terms of frame specific situational roles. According to Frame Semantics (Fillmore, 1977, 1982; Fillmore and Atkins, 1992; Petruck, 1996) then, each word, in a given meaning, is defined in relation to a particular schematic conceptual scenario, or semantic *frame*. These frames underlie the use and interpretation of the lexical items and their general complementation and modification properties (Johnson and Fillmore, 2000; Fillmore et al., 2003: 241).

FrameNet (<http://www.icsi.berkeley.edu/~framenet>) is a computational lexicography project that extracts information about the linked semantic and syntactic properties of English words from large electronic text corpora, using both manual and automatic procedures, and presents this information in a variety of web-based reports (Fillmore et al., 2003: 235). The data included in the resulting relational database, always based upon corpus evidence, achieve a level of depth which no other print or computational lexical resource has been able to obtain. The goal is to provide, for a significant portion of the vocabulary of contemporary English, a body of semantically and syntactically annotated sentences from which reliable information can be reported on the valences or combinatorial possibilities of each item included.

2. For more information about extraction of definition-like sentences, Explicit Metalinguistic Operations (EMOs), see Rodríguez Penagos (2004ab).

The FN project, its theoretical framework and methodology, have been successfully applied to other languages apart from English, such as German (the SALSA project at Saarbrücken, <http://www.coli.inu-saarland.de/projects/salsa> and German FrameNet at Austin, <http://gframenet.gmc.utexas.edu/>), Japanese (Japanese FrameNet, <http://jfn.st.hc.keio.ac.jp/>) and Spanish (Spanish FrameNet, <http://gemini.uab.es/SFN/>), all focused on general language study. Some applications have been developed to apply FrameNet-like descriptions to specialized languages, namely BioFrameNet (Dolbey et al., 2006) and Kicktionary, (Schmidt, 2007) accessible from the FN webpage. In the same vein, our proposal is a step forward in the research carried out by Faber et al. (2006, 2007) on the applicability of FN notions to specialized language.

The Spanish FrameNet Project, henceforth SFN, is a project currently under development, headquartered at the *Laboratorio de Lingüística Informática* (LaLI) of the Autonomous University of Barcelona (Subirats, 2004, 2007). The project was launched in 2002, and since then all the outcomes of previous projects in LaLI have been devoted and, consequently adapted, to the existing FN resources to create an online lexical resource for Spanish, based on Frame Semantics and supported by corpus evidence (<http://gemini.uab.es/SFN>). This resource is already accessible from the FrameNet website (<http://framenet.icsi.berkeley.edu/>) in a platform independent format. It contains 305 frames and 575 fully annotated lexical units adding up to over 10,000 annotated sentences representative of a wide range of semantic domains.

3. Recent research in terminography: The case of MarcoCosta

The theoretical mainstays that underlie our research are rooted in the modern trends in terminology, more specifically in the Communicative Theory of Terminology (Cabré, 1999ab), the Sociocognitive Theory of Terminology (Temmerman, 2000) and the approach taken by Specialized Lexicography (Faber and Jiménez Hurtado, 2002). In the line of those new proposals, which leave aside a de-contextualized study of terms, we agree that specialized lexical units are part of particular communicative settings and thus subject to pragmatic constraints (Cabré et al., 2001: 306).

The research project we are currently involved in is called MarcoCosta (MC). Its principal objective is the generation of terminological resources for the representation of specialized concepts in the domain of Coastal Engineering in three languages: Spanish, English and German. The project focuses on three areas: (1) the conceptual organization underlying any knowledge resource; (2) the multidimensional nature of conceptual representations; and (3) knowledge extraction through the use of multilingual corpora (Faber et al., 2006).

In our research, an Event can be defined as the underlying conceptual framework of a specific knowledge domain (Faber et al., 2006). The most generic categories of a domain are configured in this prototypical domain event or action-environment interface (Barsalou, 2003) providing a template which is applied to all levels of information structuring in the resources resulting from our terminographical research. In doing this, we obtain a structure which facilitates and enhances knowledge acquisition, as the information contained in term entries is internally as well as externally coherent (Faber et al., 2007).

The Coastal Engineering Event constructed as part of the project is based on the evidence provided by the corpus we have compiled with English, Spanish and German texts related to the field of Coastal Engineering (see Figure 1).

The Coastal Engineering Event (CEE) is conceptualized as a process initiated by an agent, and which affects a specific kind of patient and produces a result. These macrocategories (AGENT→PROCESS→PATIENT/RESULT) are the concept roles characteristic of the domain, and provide a model for representing their interrelationships. Additionally, there are peripheral categories which include instruments that are typically used during the CEE, as well as a category in which the concepts of measurement, analysis and description of the processes are grouped together (DESCRIPTION TEMPLATE) (Faber, Márquez and Vega, 2005).

As can be seen in Figure 1, the processes in the CEE can be divided into natural processes and artificial processes caused by human agents. Coastal Engineering deals with highly dynamic processes that take place in the coastal environment

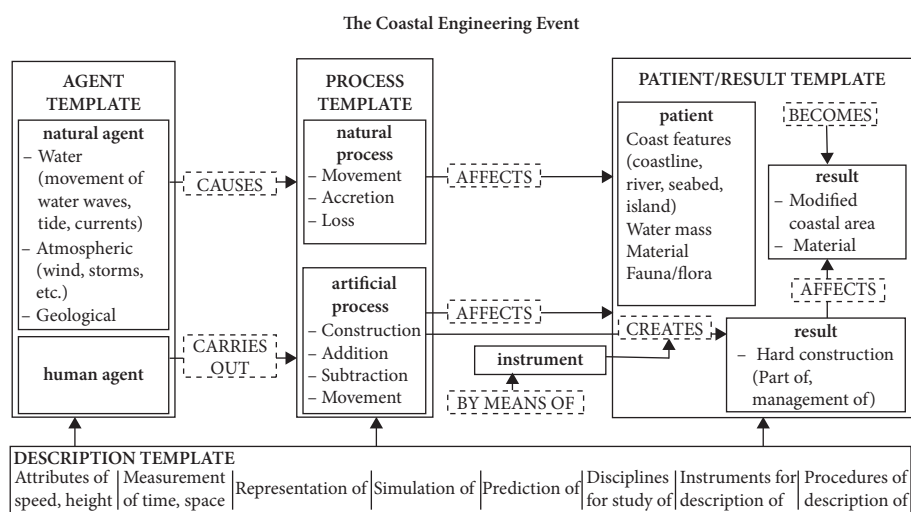


Figure 1. Coastal Engineering Event (Faber et al. 2006: 194)

(Faber et al., 2006). For example, human processes result in constructions which become part of the coastal environment and, as such, can be affected by both natural processes and other human processes. That is why both the affected and the result categories are included in the same template (PATIENT/RESULT TEMPLATE).

Although Faber and colleagues have referred to the work by Fillmore and the research staff of FN (for example, Faber and Jiménez, 2002; Faber et al., 2006, 2007), the conceptual framework Event is somewhat different from the notion of frame applied in the FN project. For the purpose of the present research, however, we will focus on what unites both projects. Both projects take on the communicative situational perspective (Fillmore et al., 2003; García de Quesada, 2001; Montero-Martínez and García de Quesada, 2004; Seibel, 2004): both apply a research approach based on corpus analysis (Fillmore, 1994; Ruppenhofer et al., 2006; Pérez Hernández, 2002; Faber et al., 2006); and both study and provide information on the semantics as well the syntactical behaviour of the items under analysis (Ruppenhofer et al., 2006; Faber et al., 2006).

4. Lexical semantics meets terminography: Definition in MarcoCosta and FrameNet

Within a dictionary entry, be it specialized or general, the definition is undoubtedly one of the most important parts. Leaving aside the need to include in the definition the necessary and sufficient characteristics (Fodor et al., 1970), the notion of prototype from cognitivism is applied in MC as well as in FN, always within a corpus-based approach.³

In recent projects within our research history in Terminography, using a semasiological corpus-based approach, we take on García de Quesada's (2001, 2002, 2003ab, 2004 [together with Montero-Martínez]) and Seibel's (2004) notion of terminographic definitions. According to those authors, definitions are considered dynamic entities with (i) an underlying structure formed by a set of conceptual properties following categorial templates; and following a controlled vocabulary (see Figure 2), (ii) a lexical formalization of such properties on the terminographic definition proper (see Figure3).

3. It must be said that, within this framework, definitions are taken as a meaning representation method. Nowhere in MC or FN is it claimed that the conceptual structure of the human mind follows a definitional structure.

BODY-PART	
A	→ ANIMAL-PART
ELEMENT-OF	→ COLLECTION
HAS-FUNCTION	→ EVENT
LOCATION-OF	→ HAVE-NEOPLASM
THEME OF	→ DIAGNOSTIC TEST

DIAGNOSTIC-DEVICE	
IS-A	→ MEDICAL-DEVICE
HAS-FUNCTION	→ EVENT
HAS-PARTS	→ OBJECT
INSTRUMENT-OF	→ DIAGNOSTIC-TEST
USED-BY	→ HEALTH-ROLE

MEDICAL-ROLE	
A	→ HEALTH-ROLE
AGENT-OF	→ EVENT
ACTIVITY-FOR- ROLE	→MEDICAL-SERVE
AREA-OF-INTEREST	→MEDICINE
MEMBER-OF	→ HEALTH-CARE-ORGANIZATION
WORKS-IN	→ HEALTH-CARE-ORGANIZATION

HEALTH-CARE-ORGANIZATION	
IS-A	→ SOCIAL-OBJECT
HAS-MEMBER	→ MEDICAL-ROLE
LOCATION	→ PLACE
ORGANIZATION-INVOLVED-IN	→ WORK-ACTIVITY
PLACE OF WORK	→ HEALTHROLE

Figure 2. Conceptual templates of fout different categories from the medical event (García de Quesada 2001)

ANTICANCER DRUG	
bleomicina	
[IS-A]	→ antibiótico antineoplásico
[HAS-TRADENAME]	→ Bleoxane
[USED-IN]	→ quimioterapia
[USED-IN-THE-TREATMENT-OF]	→ tumores germinales, osteosarcoma, enfermedad de Hodgkin sarcoma de Kaposi, carcinoma epidermoide de cabeza y cuello, pene y cérvix.
[HAS-SIDE-EFFECT]	→ toxicidad pulmonar; fiebre
[WAY-OF-ADMINISTRATION]	→ intravenosa; intramuscular, intrapleural: intraperitoneal

↓

antibiótico antineoplásico, cuyo nombre comercial es Bleoxane, utilizado en la quimioterapia de tumores germinales, osteosarcoma, enfermedad de Hodgkin, sarcoma de Kaposi y carcinoma epidermoide de cabeza y cuello, pene y cérvix. La toxicidad característica es la pulmonar; muy frecuentemente produce fiebre.

Administración intravenosa, intramuscular, intrapleural o intraperitoneal.

Figure 3. Lexical formalization of a conceptual template (García de Quesada 2001)

Along a similar line, Montero-Martínez’s notion of terminological phraseme (TP) (2002, 2003 [together with García de Quesada]), based on a transcategorical and ontological approach, is poured onto the above mentioned idea of terminographic

definition, resulting in a theoretical and practical scheme for the design of a corpus-based grammar aimed at the construction of pragmatic terminographic definitions (Montero-Martínez and García de Quesada, 2004). Indeed, the propositional realization of the most productive relations and attributes of the different categories depends on a restricted group of phrasemes extracted from corpora with different levels of specialization. These most productive relations and attributes, spotted at a conceptual level, constitute a definitional grammar, a controlled language which leaves aside the distributional criteria based on grammatical category description and follows the transcategorial approach proposed by Lexical Syntax (Subirats, 2001).⁴

This type of definition is based upon a pragmatic scheme, and results in a product that satisfies both the user's communicative and cognitive expectations (see Figure 4).

Going one step further, more recent research within MC has shown that definitions which effectively combine both visual and verbal information have a great deal of potential as meaning-making resources (Faber et al., 2007, see Figure 5). Following a typology of images, structured in terms of their most salient functions or in terms of their relationship with the real-world entity that they represent, the inclusion of images in term entries is based upon the conceptual relations activated in the definition of the concept (Faber et al., 2007; Prieto Velasco, 2008).

- (20) malignant neoplasm [IS-A] located in the epithelial tissues and mucous membrane [LOCATION], carcinoma affects squamous cell, basal cell, transitional cell and glandular epithelium [AFFECTS], it is the most common malignant tumor with an incidence of 80 to 90% [FREQUENCY-RATE], it is a rapidly progressive tumor [PROGRESS-RATE] treated with chemotherapy, radiotherapy, biological and hormone therapy [TREATED-WITH], among others, carcinomas in children and adults [AFFECTED-POPULATION-AGE], both males and females [AFFECTED-POPULATION-SEX] are found, prognosis in carcinoma [prognosis] depends on staging [CANCER-STAGING] at diagnosis.
- (21) kind of cancer [IS-A] in the epithelial tissues and mucous membrane [LOCATION], treatment in carcinoma includes chemotherapy and radiotherapy [TREATED-WITH], prognosis of carcinoma [PROGNOSIS] depends on staging of the tumor [CANCER-STAGING].
- (22) malignant disease [IS-A] arising from the skin or the surfaces of other structures [LOCATION] with a tendency to spread to other parts of the body [METASTATIZES-TO], conventional therapy includes treatment with drugs. X-rays and surgical removal [TREATED-WITH], the expected outcome of the disease [PROGNOSIS] varies greatly.

Figure 4. Three possible definitions of the same term, triggered by different user's needs. Activated conceptual relations are also included (Montero-Martínez and García de Quesada 2004)

4. In this approach, meaning is not the result of chains of LUs belonging to certain grammatical categories, but of dynamic dependency relations among predicates and arguments which establish hierarchies in which predicates operate over their arguments (Subirats, 2001: 23).

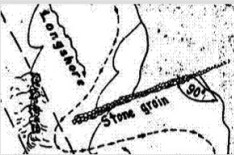
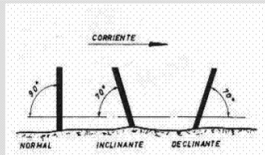
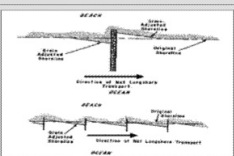
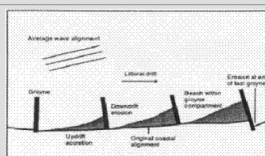
GROYNE			
	hard coastal defence structure [IS_A],		
	default value (concrete, wood, steel, and/or rock) [MADE_OF]		
	perpendicular to shoreline [HAS_LOCATION]		
	protect a shore area, retard littoral drift, reduce longshore transport and prevent beach erosion [HAS_FUNCTION]		

Figure 5. The convergence of linguistic and graphic descriptions of groyne (Faber et al 2007)

Nevertheless, much remains to be done to fully exploit visual representations in Terminography. It is our belief that incorporating research into this issue from other somewhat related fields, such as Educational Psychology (Cook, 2006; Pozzer and Roth, 2003; Mayer 2001, 2002) will prove very useful. Indeed, browsing a terminological database is a way of acquiring knowledge, a way of learning; we could, therefore, draw a parallel between a terminological database user and a learner in the science classroom. Along the same line, if visual representations, with a proper instructional design, have shown their effectiveness, we could conclude that terminological databases which take advantage of both visual and verbal media (in the form of text or audio) are also more beneficial than those which rely on either visual or verbal information alone. However, all that glitters is not gold and visual representations do not always lead to better understanding (Pozzer and Roth, 2005). Thus, the effectiveness of multimodal representations is dependent on several variables, such as whether both visual and verbal modalities are coherently linked in space and time (Mayer, 2002).

The use of a structured typology of images and video needs to be further investigated. More specific directions, such as arrows and coloured areas, as suggested by Pozzer and Roth (2005: 237), could be exploited in order to spot the right detail to be observed and link it to the appropriate frame-like text, consequently enhancing expert knowledge understanding of the database user.

In contrast, in the case of FN, no definition is provided for specific lexical units (LUs) but for frames. As we can see in Figure 6, FN lexical units come with

ignite.v

Frame: Setting_fire

Definition

COD: catch fire or cause to catch fire. no obj.

Figure 6. A LU definition

definitions, either from the Concise Oxford Dictionary, 10th Edition (courtesy of Oxford University Press) or a definition written by a FrameNet staff member.

This frame-by-frame analysis of the lexicon rather than lemma by lemma is precisely what makes FN interesting for our terminographical purposes. This will allow us to focus, not on the definition of the LU, but on the frame definition.

A frame definition conceived as a schematic representation of a situation type which includes Frame Elements (FEs), participant roles specifically defined for the frame (Fillmore et al., 2003: 305, see Figure 7), becomes a powerful tool to study how the FEs are given linguistic expression in sentences containing our specialized LU. Besides, it will serve to study how those FEs are linked to a visual representation taken from our previously structured multimodal corpora. In other words, instead of using traditional lexical relations such as synonymy, antonymy and meronymy in combination with conceptual ontological information, our aim is to link parallel lexicon fragments, i.e. contexts, with multimodal corpora by means of semantic frames and their FEs. In this way, by working on a frame level, phraseology, collocations and selection criteria of the specialized LUs contained in a specific frame can be foreseen and later instantiated within the multimodal contexts.

This notion of definition within MC and other terminographical projects could be enriched by focusing on frames as an effective way to provide detailed syntactic and semantic valence information. In our proposal, both projects meet in the definition of frames and the syntactic realizations of different FEs (FN) and in the lexicalization of predicative structures typical to the conceptual categories of a given domain (MC).

Setting_fire

Definition:

This frame describes either the creation of a **Flame** by a **Kindler** or the igniting of **Flammables** by a **Kindler**.

Figure 7. A frame definition, including the FEs

5. Frames, LUs and contextual information in coastal engineering

According to Subirats (2004: 8), we can distinguish four main stages in the construction of a semantic network of frames (SFN), namely, (1) identification of the frames and their FEs together with the LUs which are to be studied in each of them, (2) identification within the corpus of the different syntactic constructions which convey the meaning of the FEs of each predicate from a specific frame, (3) semi-automatic syntactic and semantic tagging of the sentences automatically extracted from the corpus and (4) web query of the tagging results to verify and to examine the semantic characterization of the target predicate.

We have used the FN resources freely available on the web, as well as our own data and software, namely the MC database, the MC textual and visual corpus and the application WordSmith Tools®, a suite of lexical analysis tools.

It is our aim to focus on the first three stages in the construction of a semantic network of frames, leaving verification through the web for future research. Our two main objectives are frame identification and definition, on the one hand, and textual and visual subcorpora selection and tagging, on the other. Thus, our three tasks will be to find the right frames (SFN task 1), to find the right contexts (SFN tasks 2 and 3) and to put the images into contexts (a further development of 2 and 3).

5.1 Finding the right frames

To analyse the existing conceptual framework of this field of Coastal Engineering, extract those phrasal constituents which will be subsequently tagged, and select and manipulate the images which will best serve our purposes, we used the English part of the trilingual corpus (English, Spanish and German), and the images collected and manipulated specifically for the MarcoCosta project.⁵

In an inductive way, within the scope of the above mentioned field, we decided to choose a group of words that follow similar patterns of conceptual behaviour, belong to the same frame and, therefore, also share the same FEs. We opted for the predicative nouns *precipitation*, *infiltration* and *percolation*, with the core meaning of water falling down, leaving aside, that is, other senses such as *precipitation* within the frame of Cause_change (the events which precipitated his downfall), *infiltration* within the frame Placing (they had infiltrated agents into our organization), or *percolation* within the frame Dispersal (her enthusiasm has percolated through to her team).⁶

5. For more information on the corpus, see Faber et al. (2006).

6. Examples taken from Gran Diccionario Oxford. OUP. 3rd edition.

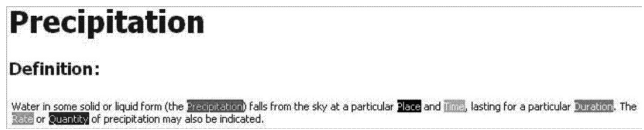


Figure 8. Definition of the frame *Precipitation*, including core and non-core FEs

The issue of placing our selected LUs into an already existing frame, creating a new one from scratch, or splitting an already existing frame into smaller ones was solved by FN in the case of *precipitation*, since a frame already exists with the same name Precipitation (see Figure 8).

Though neither percolation nor infiltration have been studied so far in FN, an initial glance shows that several frames, such as Dispersal or Filling could be said to have something in common with parts of the meanings of percolation or infiltration.

In the case of the Dispersal frame, for example, a CAUSE, *precipitation*, disperses INDIVIDUALS, (drops of) *water*, from the SOURCE, *soil upper layers*, to a/the GOAL_AREA, *soil deeper layers* (in the case of *infiltration*) or *deeper aquifers* (in the case of *percolation*, see Figure 9). However, neither *percolation* nor *infiltration* really fit into this frame because of the mere notion of dispersal and, to some extent, the fact that INDIVIDUALS are being dispersed here, i.e. water seen as drops and not as a continuum or a fluid.

Neither of the two could be inserted in the Filling frame either, since the notion of ‘filling a container/covering an area’ is not really within the core meaning of *infiltration* nor *percolation*, which is water movement through soil (see Figure 10).

After this previous conceptual analysis and having looked at the kinds of information conveyed by the phrases that are syntactically in construction with both LUs, we have found that the tagged semantic roles coincide best with that of the frame Fluidic_motion (see Figure 11).

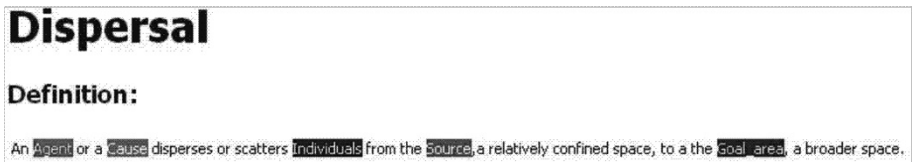


Figure 9. Definition of the frame *Dispersal*, including cora and non-core FEs

Filling

Definition:

These are words relating to filling containers and covering areas with some thing, things or substance, the **Theme**. The area or container can appear as the direct object with all these verbs, and is designated **Goal** because it is the goal of motion of the **Theme**. Corresponding to its nuclear argument status, it is also affected in some crucial way, unlike goals in other frames.

Figure 10. Definition of the frame *Filling*, including core and non-core FEs

Fluidic_motion

Definition:

In this frame a **Fluid** moves from a **Source** to a **Goal** along a **Path** or within an **Area**.

Figure 11. Definition of the *Fluidic_motion* frame

The suitability of the Fluidic_motion frame for the LUs *infiltration* and *percolation* becomes clear when we look at how the different FEs defined for the frame are present in the contexts where those LUs are instantiated in the MC corpus. According to the FrameNet database, the Core FEs for this frame are AREA, FLUID, GOAL, PATH and SOURCE. Furthermore, the non-core FEs provided by FN can also be found in the contexts of the LUs in the MC corpus of texts on Coastal Engineering: CAUSE, CONFIGURATION, DEPICTIVE, DISTANCE, DIRECTION, DURATION, MANNER, PLACE, RESULT, SPEED and TIME.

5.2 Finding the right contexts

A context is more than a chain of words which includes the entry term, but is not too big so that it fits in one sentence. In our proposal, a context is an instantiation at sentence level of an LU and its FEs configuration, embedded in a frame. Moreover, we could select a set number of contexts per entry term, or select as many contexts as there are syntactic constructions that follow the conceptual pattern of the frame within our corpus.

The question is: contexts, what for? For the purpose of this study, contexts will have two objectives. The first one, which will be dealt with in this section, is to identify and describe the possible syntactic instantiations of the FEs corresponding to a specific target predicate, which in turn belongs to a specific frame. This kind of information has proven very useful not only for the depiction of the different syntactic realizations of a given conceptual pattern, but also to make transversal queries, for instance: which conceptual patterns, all along FrameNet, are real-

ized with one specific preposition (Subirats, 2004: 183). The second purpose of contexts is to mine and process metalinguistic predication (Rodríguez Penagos, 2004ab; Malaisé, Zweigenbaum and Bachimont, 2005; L'Homme, 2003) from the MC corpus to obtain KRCs, i.e. meaningful and defining contexts (Meyer, 2001). In a parallel fashion, images are selected according to quality criteria, aimed at promoting meaningful learning (Mayer, 2002; Pozzer and Roth, 2004; Cook, 2006) and related to the above mentioned KRCs.

In order to extract the possible syntactic instantiations of the FEs corresponding to the predicate *infiltration*, we have followed the methodology proposed in Montero and García (2004), where the emphasis is put not on the most frequent lexemes but on predicative structures typical of the concept category. Once the term has been assigned to its concept category, we establish the lexical chains comprising semantically related lexemes of *infiltration*, identifying its most typical neighbours. Following the frame description, we then spot the core and non-core FEs and their lexical formalizations.

In a bottom-up and top-down approach, that is going from the corpus to the lexical information included in the MC database and vice versa, we have found the following LUs which correspond to the core and non-core FEs, occurring with our target predicate *infiltration* (see Table 1).

As we can see, the core FE FLUID is normally realized as a noun phrase (NP) that acts as an External argument (Ext) of *infiltration*. The other FEs are usually realized as prepositional phrases (PPs) headed by the prepositions shown in Table 1.

It is worth mentioning that GOAL and SOURCE are the core FEs which have fewer attestations in our corpus, quite probably because both are taken for granted and they are part of the shared knowledge, though this would have to be further investigated. Some non-core FEs, however, are more frequent than the above mentioned core FEs; such as DIRECTION (very frequently in the form of a preposition preceding PATH) or SPEED.

In Table 2 we can see some of the non-core FEs, their corresponding LUs and some syntactic and grammatical markers.

Some annotated sub-corpora are shown in Figure 12.

Table 1. Core FEs of *Fluidic_Motion* and their possible instantiations with *infiltration*, and syntactic markers

Core FE and FN definitions	Corpus-attested LUs for each FE	Syntactic markers
Area the setting in which the Fluid 's movement takes place on an unspecified Path .	soil (hillslope soils), surface	in of
Fluid the entity that changes location and moves in a fluidic way.	water, liquid, nonaqueous liquid, pressurized liquid	---
Goal the location the Fluid ends up.	zone of saturation, soil profile	to into
Path the trajectory along which the Fluid moves.	porous medium, intermediate zone	down, across, along, through
Source the location the Fluid occupies initially.	stream, well, canals	from

Table 2. Some non-core FEs of *Fluidic_motion* and their possible instantiations with *infiltration*

Non-core FE and FN definitions	Corpus-attested LUs for each FE	Syntactic and grammatical markers
Cause the reason for the fluidic motion.	high permeability	---, caused by, due to
Depictive identifies the Depictive phrase describing the Fluid throughout the action.	cumulative, enhanced	--- Adjective
Duration The amount of time for which the Fluid moves.	days, months,	---, for
Result This FE identifies the result of the Fluid moving.		is a prime factor in, results in
Manner The Manner in which the Fluid flows.	continuous	Adjective
Speed the rate at which the Fluid flows.	rapid, slower	Adjective
Place describes the location of the motion which has a specified Source, Path or Goal.	name of a place	in, at
Time The Time when the Fluid flows.	heavy rain storms, fixed times, rainfall	in, during, at

- 1... decreased INFILTRATION capacity of hillslope soils.
- 2... so something else must be impeding downward INFILTRATION of water during heavy rain storms.
- 3... continuous INFILTRATION and percolation of surface water removed the cementing material along the seepage zone towards the base
4. There were also strong empirical functional relationships between cumulative INFILTRATION at fixed times
5. Sands and gravels allow more rapid INFILTRATION due to their high permeability.
6. Enhanced INFILTRATION caused by groundwater extraction is suspected to be both a prime factor in the disappearance of a lagoon system at the Toluca Basin and a potential risk to long-term groundwater quality.
7. In the last few years, the pattern has been variable, but tending towards a split of nine months river discharge and three months INFILTRATION.
8. A rapid INFILTRATION system can handle a large amount of waste water on a relatively small land area.

Figure 12. Snapshot of some annotated subcorpora examples

At this stage of the project, FEs are semi-automatically detected and semantic relations are given priority over syntactic realizations. This is the reason why, though *hillslope soils* in sentence 1 is syntactically attached to capacity, it is marked as a core FE of the target predicate, *infiltration*. The same goes for the prepositional phrase *at the Toluca Basin*, which is syntactically dependent on *disappearance*, and (partially) semantically dependent on the target predicate, *infiltration*.

The aim of those annotated sub-corpora is to see the most productive syntactic constructions for the target predicate. Different realizations for the core and non-core FEs of *infiltration* are also attested. With those contexts, the emphasis is put on how a specific predicate within a frame selects its own arguments, leaving aside any specification about the conceptual features of the arguments. In this way, as seen in Examples 1 and 5, *infiltration* selects *hillslope soils* and *sands and gravels*, respectively, as the NPs to be included within the core FE AREA, where no syntactic marker is present. Similarly, the non-core FE CAUSE is expressed, in Examples 5 and 6, by means of two different syntactic markers, namely, *due to* and *caused by*, followed by the NPs, *their high permeability* and *groundwater extraction*, respectively. For more information about those and other FEs, we refer the user to the MC database, where hierarchical and non-hierarchical relations are explicitly established.

5.3 Putting the images into context

As opposed to the contexts used for 5.2, the contexts here will be aimed at constructing an event (Faber et al., 2006), normally a partial one, where more than one LU and more than one frame are normally involved. The reason is to complement the syntactic and frame-like information given before and enrich it with the co-occurrence of interrelated LUs against a common background, usually in the form of an image, since it is by comparison and opposition that many concepts acquire their true meaning.

In Terminology, one of the main criteria for selecting corpora is term density. For the purpose of our study, we also include another criterion: that of density of knowledge patterns (KPs), understood as explicit metalinguistic information on how those terms, and their underlying conceptual structures, relate to each other. Bearing in mind that not all corpora rich in term density are rich in KPs (Halskov, 2005), we have constructed two different but interrelated corpora.

In this section, we therefore not only look for term density, as we did in 5.2, but for domain-independent KPs, spotting the metalinguistic markers which include typographical, syntactic, and lexical clues (Rodríguez Penagos, 2004a; Halskov, 2005). In other words, we look for the intersection of contexts rich in knowledge patterns and contexts rich in knowledge. The result is a compendium of defining KRCs that are especially useful to relate LUs within one frame as well as between two or more frames within a domain event. Finally, we relate the intersection of KRCs and KPs to visual resources which, pertaining to the same event, encompass as many related terms as possible.

Let us take as an example the Water Cycle Event, where *infiltration* would be inserted. With the help of the MC database, we find other terms which would co-occur with *infiltration*, such as *percolation*, *precipitation*, *runoff*, as predicates, and *water*, *surface*, or *water vapour*, as arguments.

Table 3 shows several examples of KPs found in our corpus that include one of the specialized LUs, *precipitation*, *infiltration* or *percolation*. These KPs provide explicit information on the concepts and their relation to a specific frame and, since it is usually a definition that it is provided, mainly hierarchical relations are portrayed.

Some other interesting KPs are shown in Table 4 where, in one KP, two or more specialized LUs are related. In this case, one aspect of the conceptual structure of the specialized LUs is highlighted. This table mainly shows non-hierarchical and complex relations.

Table 3. Corpus-attested KPs for *precipitation*, *infiltration* and *percolation*

LU	Corpus-attested KPs	KP Markers
precipitation	Precipitation means solid or liquid moisture falling from the sky.	bold print means
	Annual total precipitation refers to rain, snow and other forms of moisture such as hail.	refers to
	Before the rain, there had been 32 days without measurable precipitation (defined as one hundredth of an inch or 0.01") in the Little Rock area	parenthesis defined as
infiltration	In this study, "infiltration" is synonymous with "ground-water recharge", unless otherwise specified.	quotation marks synonymous with
	In this context, infiltration is the water that seeps into the ground and migrates downward below the root zone to reach the water table.	in this context is
	Infiltration, as the term is used by many surface-water hydrologists, includes all water that seeps into the ground, regardless of whether or not it ever reaches the water table.	as the term used (by many) as
percolation	Deep percolation is defined as the drainage of water beyond the root zone (Hillel, 1982).	defined as
	In a process called percolation, water passing through the unsaturated zone may reach the saturated zone and become groundwater.	called

The KPs in Table 4 give explicit information on how different specialized LUs are related to each other and through those relations a (partial) event is constructed. Contrary to what occurs in the corpora selected according to term density criteria, these KPs do not always share the same meaning, and not all KPs have the same explanatory power. In this table the KP, X means Y, in *'Infiltration' means the entry of precipitation or runoff into or through the soil*, gives a definition of the LU *infiltration*, but does not explain the difference between the two LUs used for this definition, *precipitation* and *runoff*. The last KP in the table, *The hydrologic cycle involves evaporation, condensation, run-off, infiltration, percolation, and transpiration*, names the processes involved in the Water Cycle Event, but this does not imply that it is exhaustive and does not give any information on what those processes are or how they relate to each other.

If we want the end users of our database to really exploit the information contained in the contexts provided, we have to carefully select specialized KPs, once KP and KRC corpora have been intersected. We believe that images, depicting a common background to the related LUs, provide a very useful resource for this. However, very often definitions and images are not really meaningful, since the contrast and differences with other specialized LUs are not made explicit either. This applies to the following text and its corresponding image in Figure 13.

Table 4. Corpus-attested KPs with more than one specialized LU

Related LUs	Corpus-attested KPs
infiltration precipitation	In hydrology, infiltration refers to the maximum rate at which a soil can absorb precipitation
infiltration runoff	Potential run-off occurs when the application rate exceeds the soil infiltration rate.
runoff erosion	Surface runoff is one of the causes of erosion of the earth's surface.
infiltration precipitation runoff	"Infiltration" means the entry of precipitation or runoff into or through the soil.
infiltration precipitation runoff evapotranspiration	precipitation – runoff + evapotranspiration + infiltration, which means that the water that falls to the ground in the form of rain, snow, etc will either soak into the groundwater, runoff into surface streams, or be evaporated from the surface or transpired through plant leaves
infiltration precipitation runoff precolation evaporation condensation transpiration	The hydrologic cycle involves evaporation, condensation, runoff, infiltration, precolation, and transpiration

Despite the fact that the image helps us to infer that *infiltration* and *percolation* refer to a downward movement as opposed to *transpiration*, which is an upward one, it is impossible to deduce the difference between *infiltration* and *percolation*.

In our corpus we also find other texts (see Table 5), KRCs with KPs such as *refer*, *defined*, *used*, *difference is (not similar to)*...where the difference between both terms is made explicit.

The images in Figure 14 depict what differentiates both LUs.

Images, such as the ones in Figure 14, explain things more clearly than the type of image in Figure 13. Our aim is to link those kinds of images to KRCs of the type shown in Table 5, so that we can see how LUs from different frames relate to each other in a domain specific event.

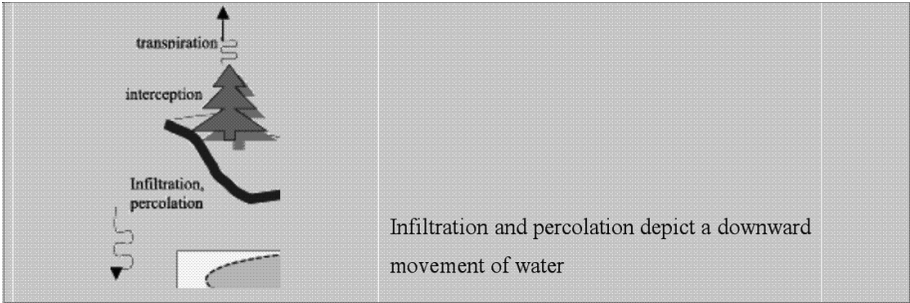


Figure 13. A multimodel context for *infiltration* and *percolation*

Table 5. Sentences where *infiltration* and *percolation* are compared

Although percolation and infiltration are very similar words, we use percolation here to refer to the transfer of water from the soil to the deeper aquifer – the ground water reservoir.

Source:
http://www.carleton.edu/departements/geol/DaveSTELLA/Water/watershed/watershed_model.htm

The terms infiltration and percolation are often used interchangeably. Infiltration is the entry of water into the soil surface. Infiltration constitutes the sole source of water to sustain the growth of vegetation and it helps to sustain the groundwater supply to wells, springs and streams. Percolation is the downward movement of water through soil and rock. Percolation occurs beneath the root zone. Groundwater percolates through the soil much as water fills a sponge, and moves from space to space along fractures in rock through sand and gravel, or through channels in formations such as cavernous limestone.

Source: <http://www.water-research.net/Watershed/hydrologicalcycle.htm>

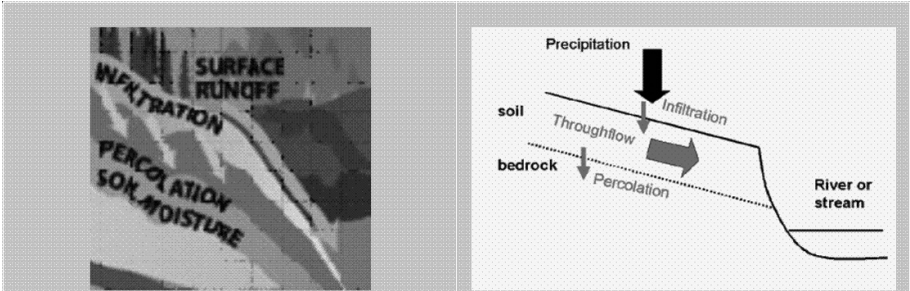


Figure 14. Images including and contrasting *infiltration* and *percolation*

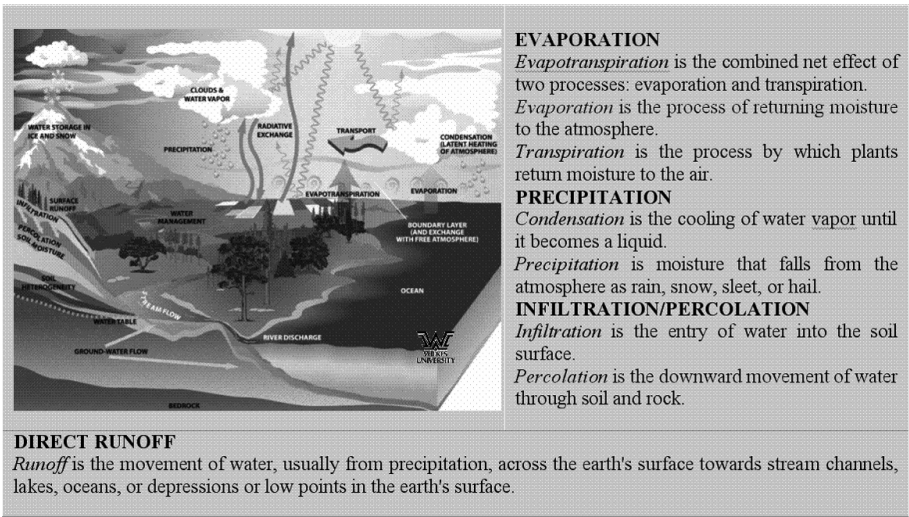


Figure 15. Visual representation of a domain specific event where the relations among different frames and their LUs are shown

The multimodal context in Figure 15 includes (1) a visual representation of the water cycle, where the different symbols, such as arrows and spirals, explain different processes, direction and phase change and (2) definitional contexts chosen and manipulated depending on what has been profiled in the image.⁷

From the perspective of the database user, if the different types of information are well and coherently integrated, the frames, contexts and images can provide a multimodal learning experience. The frame information will help the user understand the underlying conceptual structure of the domain and the relationship of each LU with one or more specific frames. It will also provide the user with the necessary information to predict the syntactic behaviour of the LU. Textual information, selected according to the criterion of knowledge richness, not only shows the syntactical behaviour of the LU but also provides the user with the necessary information on how the LU relates to other LUs of the same frame and other frames. The images contained in the database, selected according to the same criterion of knowledge richness, provide the user with information on how the LU relates to its frame and to other frames in a specific event, as defined by MC.

7. www.water-research.net/powerpoint/water_cycle2.ppt [Access date: October, 2008].

6. Conclusion

FrameNet resources have a great deal to offer to lexicographers and dictionary publishers (Atkins, Rundell and Sato, 2003), as well as to terminographers and domain specific research projects which account for the continuum between general and specialized language, at both the cognitive and linguistic realization level. Indeed, the fact that (1) there is an ongoing FN project for Spanish and its outcomes, (2) data, publications and some tools from the Berkeley FN project are already available through the web, and (3) those projects are a work-in progress, makes it plausible to apply FN rational, data and tools to Terminology projects. Besides, in MarcoCosta and other previous projects, Lexical Semantics and Frame Semantics have been permeating our research; this article pretends to be a step further in a possible cooperation between both groups.

So far, contexts have been studied and applied in an isolated way, usually selected according to syntactic criteria leaving aside conceptual or meaning criteria. We suggest that two kinds of contexts be offered to the end user: (1) the FN like contexts, that is, subcorpora chosen to depict the different syntactic constructions, conveying the different realizations of the FEs and the target predicate, and (2) the multimodal contexts, comprising, on the one hand, KRCs, that is, corpus-attested metalinguistic texts which mirror the dynamic evolution of domain specific discourse, and, on the other, images which encompass a comprehensive view of all possible related specialized LUs, following the criteria stated in Science Education literature. Both context types are embedded in a frame and in an event, which gives the end users the possibility to find information at different levels of complexity and will provide them with a learning environment flexible enough to adapt to their needs.

References

- Atkins, S., M. Rundell and H. Sato (2003). The Contribution of FrameNet to Practical Lexicography. *International Journal of Lexicography*, 16 (3), 333–357.
- Barsalou, L. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, 18 (5/6), 513–562.
- Cabré, M. T. (1999A). *Terminology. Theory, Methods and Applications*. Amsterdam: John Benjamins.
- Cabré, M. T. (1999). *La terminología: representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos*. Barcelona: IULA.
- Cabré, M. T., M. Domènech, J. Morel and C. Rodríguez (2001). Las características del conocimiento especializado y la relación con el conocimiento general. In M. T. Cabré and J. Feliu (eds.), *La terminología técnica y científica* (173–186). Barcelona: IULA.

- Cook, M. P. (2006). Visual Representations in Science Education: The Influence of Prior Knowledge and Cognitive Load Theory on Instructional Design Principles. *Science Education*, 90 (6), 1073–1091.
- Dolbey, A., M. Ellsworth and J. Scheffczyk (2006). BioFrameNet: A Domain-specific FrameNet Extension with Links to Biomedical Ontologies. In *Proceedings of KR-MED*. Baltimore, Maryland, USA. [Available at <http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-222/krmed2006-p10.pdf>].
- Faber, P. and C. Jiménez Hurtado (eds.) (2002). *Investigar en terminología*. Granada: Comares.
- Faber, P., C. Márquez Linares and M. Vega Expósito (2005). Framing Terminology: A process-oriented approach. *META*, 50 (4), CD-ROM.
- Faber, P., S. Montero-Martínez, R. Castro, J. Senso, J. A. Prieto, P. León, C. Márquez, and M. Vega (2006). Process-oriented terminology management in the domain of Coastal Engineering. *Terminology*, 12 (2), 189–213.
- Faber, P., P. León Araúz, J. A. Prieto Velasco, and A. Reimerink (2007). Linking images and words: the description of specialized concepts. *International Journal of Lexicography*, 20, 39–65.
- Felber, H. (1984). *Terminology Manual*. Paris: UNESCO and INFOTERM.
- Fillmore, C. J. (1968). The case for case. In E. Bach and R. Harms (eds.), *Universals in Linguistic Theory* (1–88). New York: Holt Rinehart and Winston.
- Fillmore, C. J. (1977). Scenes-and-frames semantics. In A. Zampolli (ed.), *Linguistics Structures Processing* (55–81). Amsterdam & New York: North Holland Publishing Company.
- Fillmore, C. J. (1982). Frame semantics. In The Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm* (111–137). Seoul: Hanshin.
- Fillmore, C. J. (1994). Starting where the dictionaries stop: The challenge of corpus lexicography. In B. T. Atkins and A. Zampolli (eds.), *Computational Approaches to the Lexicon* (349–393). Oxford: Oxford University Press.
- Fillmore, C. J. and B. T. Atkins (1992). Towards a Frame-based organization of the lexicon: the semantics of RISK and its neighbours. In A. Lehrer and E. Kittay (eds.), *Frames, Fields, and Contrasts: New Essays in Semantics and Lexical Organization* (75–102). Hillsdale: Lawrence Erlbaum.
- Fillmore, C. J., C. R. Johnson and M. R. L. Petruck (2003). Background to FrameNet. *International Journal of Lexicography*, 16 (3), 235–250.
- Fodor, J. A., M. F. Garrett, E. C. T. Walker and C. H. Parkes (1970). Against Definitions. *Cognition*, 8, 263–367.
- FrameNet and Frame Semantics (2003). *Special Issue International Journal of Lexicography*, 16, 3. Oxford: Oxford University Press.
- Gangemi A., D. M. Pisanelli and G. Steve (1999). An Overview of the ONIONS Project: Applying Ontologies to the Integration of Medical Terminologies. *Data Knowledge Engineering*, 31 (2), 183–220.
- García de Quesada, M. (2001). *Estructura definicional terminográfica en el subdominio de la oncología clínica*. Madrid: CSIC/Elies. [Available at: <http://elies.rediris.es/elies14/>].
- García de Quesada, M. (2002). Propuesta de estructura definicional terminográfica en OntoTerm. *Terminology*, 8 (1), 59–92.
- García de Quesada M. and S. Montero-Martínez (2003a). Hacia una gramática de la definición terminográfica. In N. Gallardo San Salvador (ed.), *Terminología y traducción: un bosquejo de su evolución* (243–254). Granada: Atrio.
- García de Quesada, M. and S. Montero-Martínez (2003b). Documentación y adquisición terminográficas basadas en el conocimiento: el caso de la Interpretación. *Hermeneus*, 5, 107–130.

- Halskov, J. (2005). *Exploring Knowledge Patterns for building knowledge-rich corpora*. Power-Point presentation given at the University of Montreal, November 30th, 2005. [Available at http://www.halskov.net/files/Montreal_nov2nd_v1.4.pdf].
- House, J. (2006). Text and context in translation. *Journal of Pragmatics*, 38, 338–358.
- Johnson, C. R. and C. J. Fillmore (2000). The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics (ANLP-NAACL 2000)*, April 29–May 4, 2000, Seattle WA. 56–62.
- L'Homme, M. C. (2003). Capturing the lexical structure in special subject fields with verbs and verbal derivatives. A model for Specialized Lexicography. *International Journal of Lexicography*, 16 (4), 403–422.
- Malaisé, V., P. Zweigenbaum and B. Bachimont. (2005). Mining defining contexts to help structuring differential ontologies. *Terminology*, 11 (1), 21–53.
- Mayer, R. E. (2001). *Multimedia Learning*. New York: Cambridge University Press.
- Mayer, R. E. (2002). Cognitive Theory and the Design of Multimedia Instruction: An Example of the Two-Way Street Between Cognition and Instruction. In M. D. Hakel and D. F. Halpern (eds.), *New Directions for Teaching and Learning: Applying the Science of Learning to University and Beyond* (vol. 89, 55–71). San Francisco: Jossey-Bass.
- Meyer, I. (2001). Extracting knowledge-rich contexts for terminography: A conceptual and methodological framework. In D. Bourigault, C. Jacquemin and M. C. L'Homme (eds.), *Recent Advances in Computational Terminology* (279–302). Amsterdam: John Benjamins.
- Mitkov, R. (1998). The latest in anaphora resolution: going robust, knowledge-poor and multilingual. *Procesamiento del Lenguaje Natural*, 23, 1–7.
- Mitkov, R. (2002). *Anaphora resolution*. London: Longman.
- Montero-Martínez, S., P. Fuertes Olivera and M. García de Quesada (2001). The Translator as a Language Planner: Syntactic Calquing in an English Spanish Translation of Chemical Engineering. *META*, 46 (4), 687–698.
- Montero-Martínez, S., M. García de Quesada and P. A. Fuertes-Olivera (2002). Terminological Phrasemes in OntoTerm[®]: A new theoretical and practical approach. *Terminology*, 8 (2), 177–206.
- Montero-Martínez, S. and M. García de Quesada (2003). Terminological Analysis for Translation. In *Perspectives: Studies in Translatology*, 11 (4), 293–314.
- Montero Martínez, S. and García de Quesada, M. (2004). Designing a Corpus-Based Grammar for Pragmatic Terminographic Definitions. *Journal of Pragmatics*, 36 (2), 265–291.
- Pérez Hernández, C. (2002). *Explotación de los corpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento*. Madrid: CSIC/Elies. [Available at: <http://elies.rediris.es/elies18/>].
- Petruck, M. (1996). Frame semantics. In J. O. Östman, J. Blommaert and C. Bulcaen (eds.), *Handbook of Pragmatics* (1–13). Amsterdam: John Benjamins.
- Pozzer, L. L. and W. M. Roth. (2003). Prevalence, Function and Structure of Photographs in High School Biology Textbooks. *Journal of Research in Science Teaching* 40 (10), 1089–1114.
- Pozzer, L. L. and W. M. Roth (2005). Making Sense of Photographs. *Science Education*, 89 (2), 219–241.
- Prieto Velasco, J. A. (2008). *Información gráfica y grados de especialidad en el discurso científico-técnico: un estudio de corpus*. Ph.D. dissertation. University of Granada. Granada (Spain).
- Rodríguez Penagos, C. (2004a). *Metalinguistic Information Extraction from Specialized Texts to Enrich Computational Lexicons*. Ph.D. dissertation. University Pompeu Fabra. Barcelona

- (Spain). [Available at http://www.tdx.cesca.es/TESIS_UPF/AVAILABLE/TDX-0228105-114717/tcrp1de1.pdf].
- Rodríguez Penagos, C. (2004b). Mining metalinguistic activity in corpora to create lexical resources using information extraction techniques: the MOP system. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics 2004*. Morristown, NJ: ACL. 215–222.
- Ruppenhofer, J., M. Ellsworth, M. R. L. Petruck, C. R. Johnson and J. Scheffczyk (2006). *FrameNet II: Extended Theory and Practice*. [Available at: <http://framenet.icsi.berkeley.edu/book/book.html>].
- Schmidt, T. (2007). The Kicktionary: A Multilingual Resource of the Language of Football. In G. Rehm, A. Witt and L. Lemnitzer (eds.), *Data Structures for Linguistic Resources and Applications* (189–196). Tübingen: Gunter Narr.
- Seibel, C. (2004). *La codificación de la información pragmática en la estructura de la definición terminológica*. Granada: Universidad de Granada.
- Subirats Rüggeberg, C. (2001). *Introducción a la sintaxis léxica del español*, 13. Madrid: Iberoamericana.
- Subirats Rüggeberg, C. (2004). FrameNet Español. Una red semántica de marcos conceptuales. In E. Serra and G. Wotjak (eds.), *Cognición y Percepción Lingüísticas* (182–196). Valencia: University of Valencia and University of Leipzig. [Available at: http://gemini.uab.es/SFNpub/papers/Leipzig_Paper.pdf].
- Subirats Rüggeberg, C. (2007). Relaciones semánticas entre marcos en FrameNet Español. In J. Cuartero and M. Emsel (eds.), *Vernetzungen. Bedeutung in Wort, Satz und Text. Festschrift für Gerd Wotjak zum 65. Geburtstag* (357–366). Frankfurt: Peter Lang. [Available at: http://gemini.uab.es/SFNpub/papers/FrameNet_Espanol-Leipzig.pdf].
- Subirats Rüggeberg, C. and M. Petruck (2003). Surprise: Spanish FrameNet. In *International Congress of Linguists. Workshop on Frame Semantics*, Prague (Czech Republic), July 2003. [Available at: <http://gemini.uab.es/SFNpub/papers/subirats-petruck.pdf>].
- Temmerman, R. (2000). *Towards new Ways of Terminology Description: The Sociocognitive-Approach*. Amsterdam: John Benjamins.

How much terminological theory do we need for practice?

An old pedagogical dilemma in a new field

Vassilis Korkas and Margaret Rogers

The relationship between theory and practice is a particularly pertinent one for newly-emerging academic subject fields such as Translation Studies and Terminology Studies, as translation and terminology / terminography / specialized lexicography are activities originally rooted in practice over millennia. In this paper we would like to address a number of issues related to this relationship in the context of advanced specialized translation programmes (postgraduate or year 5). Starting from a consideration of how 'theory' as a concept can be variously interpreted by students and practitioners alike, we go on to review the goals of courses in terminology management for translation purposes with a view to identifying the potential contribution of terminological theory. In this, we will argue that certain aspects of theory can help to inform decisions of practice and that the crucial question may be one of pedagogical presentation rather than 'does theory matter?'

Keywords: terminology theory (and practice), form and content, term-concept relations, problem-solving, synonyms, polysemes, definitions, contextual examples

1. Introduction

The relationship between theory and practice is one which has implications for both, i.e. it is a bi-directional relationship. Theory can impact on practice in many ways, e.g. by providing consistent principles which can be applied in order to solve problems arising in practical work. And practice can throw up new questions for theory to answer. In this paper we intend to focus on the first perspective. The second, in itself more theoretical perspective, must remain a topic for another paper. Our focus is terminology management rather than any other language- or translation-related software.

Academic programmes focusing on specialized translation provided one of the earliest forums in which terminology studies and one of its applications, terminology management, were taught and in which issues of the theory-practice relationship (and which aspects of theory) therefore arose. While Wüster's original work on a General Theory of Terminology dates back over 70 years, to name one of the most well-known theories, training in terminology is more recent. Discussions of syllabuses for the teaching of terminology, particularly in the context of specialized translation, began to appear in the literature in the late 1970s (cf. for instance, Auger, 1979; Sager, 1979; Sager, 1981; Arntz, 1981; Picht, 1981; Picht & Draskau, 1985; Bühler, 1987; Arntz & Picht, 1989; Sager, 1992). One of the first courses in terminology to be offered was at the University of Vienna by Wüster himself in the early 1970s (Picht & Draskau, 1985: 248). Another early course was given at the University of Saarbrücken in 1979 (Arntz, 1981). In the francophone world, courses were first offered during the 1970s to students of translation at the *Université du Québec à Trois-Rivières* and the *Université Laval*. Early courses in francophone Europe were reported in Belgium, and in Nanterre, France (Auger, 1979).

In the design and delivery of any syllabus in any subject field, decisions have to be made on approach as well as scope and content; and in considering which pedagogical approach to adopt – e.g. deductive (top-down) or inductive (bottom-up) – direct light is cast on the complex relationship between theory and practice. It is therefore from this pedagogical perspective that we have chosen to address this relationship. Our aim is to present a view of how theory can support practice in the context of a contemporary Masters programme in specialized translation in the UK, based on our experience of teaching terminology skills for translation in this context to many cohorts of students.

It has always been our aim to approach our task in a way which goes beyond an instrumental, technology-driven development of software manipulation skills. While such skills are clearly important and necessary, so is an active engagement with the contents of any term-base, if it is to be fit-for-purpose. Courses on terminology management software in the UK context are most often offered as part of university programmes in translation, mostly at 'postgraduate' (i.e. graduate) level following a four-year Bachelors degree. Such programmes usually consist of a combination of practice-based translation (often in two language pairs), some translation theory, and other subjects such as technical writing, professional aspects of translation, and translation and technology, often including terminology management. The programmes follow a pattern of two taught semesters followed immediately by a dissertation semester, the whole programme covering a full calendar year. Terminology management is a regular topic for Masters dissertations.

Our paper starts with the practical consequences of working with software (Section 2) and the expectations which this generates in students, compared with

those of the tutors concerned (Section 3). The tricky question of what is meant by 'theory' in this context is then discussed (Section 4), before the scope of so-called theoretical issues relevant to a course with applied aims is addressed (Section 5). We then propose a model in which 'problem-solving' is a key notion in linking theory and practice pedagogically, followed by three illustrations of how aspects of terminology theory can support decision-making in terminology management for translation (Sections 6 and 7). Finally, we offer a conclusion (Section 8).

2. Dealing with the 'itchy fingers syndrome'

Unlike many other specialist areas in higher education, where software is used alongside traditional teaching media, when it comes to translation training the use of software presents one inherent disadvantage: the fact that the terminology management software in question is designed and marketed commercially for professional use means that it has not been designed for educational purposes. Therefore, the first and biggest challenge educators are faced with is to design and develop training models that would adequately reflect standard practice with software that is not normally used for training. The fact that even software developers themselves are now offering training courses on the software they produce comes to show that the dissociation between the tool and the user is a clear and present danger when no training aspect is present (in-built) in the software itself.

As is the case for most software applications, most new users will want to start using the tool straight away without worrying about any of the background or the concept behind the operation of the software. This of course limits the potential scope of application of the software (as users cannot or do not care to see the limitations or, on the contrary, the full capabilities of the tool) when it comes to real-life professional scenarios or classroom simulations of the same. Therefore, it is often the case that students of translation who are trained on terminology management software are 'sufferers' of what we call the 'itchy fingers syndrome', i.e. enthusiastic beginners who want to get straight to the practical aspects of the software.

Especially in such cases, broader issues that have to do with the design and the scope of the tool itself, but also with principles of good practice in terminology management, become even more pertinent. One such issue is differentiating between form and content in a terminology database. Whereas a terminologist would see the terminology management system (TMS) as an empty shell that can be shaped into a particular form according to the demands of the given project, and the data (and its structure) as the content which can be stored in the customized shell, a new user would find it hard to make that distinction and, in fact, might not even care to make that distinction, at least at first. To a beginner, both data and the

shell that contains them are just some form of input that goes into a term-base – no regard for structure is really necessary.

At a more advanced stage, when principles of terminology management become more apparent and relevant, real issues of practice emerge and the aforementioned first-level distinction (between form and content) becomes very relevant indeed. For example, in terms of form: creating new entries, checking and editing term records, creating pre-set menus and cross-references (between elements of entries and between whole entries in the same term-base), searching and filtering – all these are operations that need to be rationalized in the eyes of any user of a TMS, more so for beginners. On the other hand, in terms of content, there are emerging issues of theory that relate to the above practical issues: identifying candidate terms, validating data with the help of subject experts, providing elaboration (in the form of definitions and contextual examples), structuring and relating fields within an entry record, and so on. In this process of identifying and processing data, theory can help develop a sense of relevance and motivation.

3. A difference in perspectives

The issue of motivation is an interesting one when considered from the point of view of a student of terminology theory. Particularly when a practical element is included in the training programme, i.e. a module that involves training on a terminology management tool, then students are likely to find difficulty in making the connection between that practical element and the terminology theory element. Why is terminology relevant? Why is *terminology theory* relevant? Relevant to what? When the expectations of students are essentially restricted in that they expect to learn how to use the software, the motivation behind learning about and actively studying terminology theory is unclear, at least in their eyes.

On the other hand, the perspective of teachers and trainers is significantly different in this respect. Terminology theory can be used as a framework which will provide the necessary guidelines in practice (and that is essentially the link between theory and practice in the field of terminology studies). It can also provide qualitative criteria for evaluation and assessment purposes, which means that these criteria can be used to tell the difference between cases of good practice and cases of bad practice.

Let us then examine where (and if at all) these two perspectives meet.

4. What can be understood by theory?

For many students, even students who are in their fifth year of university education, 'theory' is seen as an exclusive alternative to practice or as an irrelevant side game. This point of view weakens the concept of theory, as it encompasses anything which is not 'practical'. An example can serve to illustrate this perceived distinction: in the final stage of their Masters programmes, our translation students write a dissertation. This can consist of a translation with commentary, a term-base with commentary or a topic-based dissertation. The latter has become known among the students as a 'theoretical' dissertation, regardless of its aims, approach, scope and subject matter. The two former options are set in some kind of false opposition to this, disregarding on the one hand the commentary accompanying the translation or term-base, in which similar arguments may be presented and similar scholarly sources analysed as those used in the topic-based option, and on the other hand the knowledge which underpins decisions of practice.

Rather than the binary opposition which tends to shape student expectations about theory and practice, we would prefer to think in terms of a set of interrelated elements: principles, approaches, and procedures. In this context, 'theory' can be understood as a coherent set of principles (see also Budin's (2001) third 'meaning cluster' below), which is perhaps closest to Wüster's *Allgemeine Terminologielehre*, which was developed in response to what Wüster saw in 1930s Austria as a basic problem of specialized communication from his perspective as an engineer. Principles are task-neutral and comprise the basis for systematic work of a theoretical or practical kind. In the case of terminology as originally understood by Wüster, these principles are drawn from a range of disciplines, including linguistics, epistemology, computer science (*Informatik*) (Wüster, 1974). These principles may be more or less relevant depending on the application. For instance, the documentalist will draw more on the principles of information or computer science than the specialized translator, and the translation-oriented terminologist will draw more on linguistic principles than either the domain terminologist or the documentalist. Here, approaches could be distinguished from principles. They are not neutral, but indicate an orientation, such as descriptive or normative, semasiological (word-based) or onomasiological (concept-based).

An approach can be seen as a way of tackling a particular task. Hence, the task of the terminological documentalist (to produce codified unambiguous systems of subject 'descriptors' or terms for information retrieval) and that of the domain terminologist (to produce codified unambiguous systems of terms for professional communication) can be seen as normative in approach. The tasks of the translation-oriented terminologist and the specialized translator are both descriptive in their orientation, since they have to deal with language as it is used by authors in a range of texts in all their variations.

Procedures are even more concrete than approaches. In all cases, 'terminologists' (whatever the task) need to know how to pursue their tasks in real time. Procedures indicate actions for executing particular sub-tasks related to an overall orientation and task. While certain procedures may be common to more than one terminological task, others may be confined to just one. So, for example, documentalists and domain terminologists will not be concerned with procedures for selecting a contextual example to illustrate the linguistic usage of a term, but the translation-oriented terminologist and the specialized translator will be.

In a 'critical evaluation' of terminology theory, Budin (2001) proposes four different understandings of 'theory'. The first is extra-scientific, indicating a 'vague conjecture' or opposing 'useless theory' and 'common sense action and practice' (2001: 9). Our students' understanding of theory as described above would fall under the second part of this 'meaning cluster'; 'theory' is also commonly used in student essays as an inflated way of referring to an idea. The second 'meaning cluster' proposed by Budin is the philosophical meaning derived from the Ancient Greek *θεωρία* meaning 'observation' or 'contemplation' starting with Plato's separation of 'the world of ideas [...] from the world of practice and from practical knowledge' (2001: 9). Again, this has some resonance with the simplified binary opposition of theory and practice but not in the evaluative sense of 'theory bad', 'practice/common sense good'. The third cluster takes into account different degrees of discipline-based theoreticity, ranging from those which 'describe a certain practice in a systematic way' and those that describe 'abstract entities and constructions of human thought' (2001: 10). Terminology theories are placed by Budin at a 'relatively low level of theoreticity' (2001: 19). The fourth and final cluster concerns usages in contemporary philosophy of science, which is marked by a 'methodological pluralism' in so far as many means of reaching the theoretical goals of explanation and prognosis are envisaged (2001: 10–11).

In this last sense, terminology theory is still pre-theoretical. And so we return to the understanding of 'theory' in our present context as a set of internally-consistent principles drawn from different fields which make explicit their 'methods of description' (Budin, 2001: 19).

5. Which bits of theory?

Given our aim of linking theory and practice through problem-solving, our syllabus is shaped by the type of problems which students typically encounter when trying to build a term-base. The following items form the framework from which solutions can be derived in a way which is in our view likely to encourage a reflected, consistent and motivated approach. The alternative is a series of *ad hoc*

decisions which ultimately leads to problems in database structure, coverage, integrity and quality of content. We stop short of implying a direct and deterministic relationship here between what we have called rather grandly the 'theoretical framework' and the solutions to particular problems, as decisions usually contain an interpretive element e.g. when deciding degrees of equivalence.

What is a term? What is a concept? What is a domain?

While each of these questions is in itself highly challenging from a linguistic or philosophical perspective, our more modest aim is to highlight the term-concept distinction (important for understanding the structure of a concept-based term-base) on the one hand, and the term-domain perspective of special languages on the other hand. This we understand in the sense that terms belong only to the linguistic resources of specialist domains and express their meaning in relation to specific domains.

Organizing knowledge – onomasiology

Mapping knowledge spaces is an activity undertaken by knowledge engineers building expert systems and by terminologists building term-bases. In the latter case, however, the goals are still modest, as terminology management systems – at least those on the commercial market – are not structured as knowledge bases which operationalise inheritance and other types of relation between concepts. In learning to map knowledge spaces, students' attention is drawn to the need to delimit a domain, as well as to the need to sort out concept-concept relations such as genus-species and part-whole relations (important for definitions), and those on the same level of abstraction (co-hyponyms in linguistic terms). Term-concept relations can then be understood as a different kind of mapping which underlies the onomasiological organization of terminologies as opposed to the semasiological organization of most dictionaries. The fact that term-concept relations are much more likely to be one-to-many (polysemy) and many-to-one (synonymy) than one-to-one (monosemy, mononymy or Wüster's *Eineindeutigkeit* of monosemy with mononymy) is crucial to a systematic approach to populating the database.

Definitions: Identification, evaluation, structuring, matching concepts

Definitions – understood here as definitions of the concept (for other understandings, see Robinson, 1954) – are central to mapping the way in which concepts and their associated terms are related. As students of translation building a term-base are not engaged in a process of normalization as subject experts, they are unlikely to be writing their own definitions. But, having understood the importance of the relationships between concept-definition-term(s) for the structure of a terminological record in a term-base, the focus can be on identifying and evaluating definitions.

Definitions can also, as an extension of this from monolingual to bilingual terminology building, serve to establish equivalence or degrees of equivalence.

Contextual examples: Identification, evaluation

As translators are text-creators, the nuancing of meanings when terms are used in conjunction with other terms and words in text (see, for instance, Rogers, 1999) is an aspect of term use which is not normally acknowledged from a normative perspective. Furthermore, the co-text of terms, even in the mother tongue normally has to be learnt. These two dimensions of term use can be supported by the selection and meaningful evaluation of contextual examples according to explicit criteria, informed by an understanding of system-use differences.

6. Linking theory and practice

Having established what terminology theory is and what aspects of terminology theory would be relevant in the context of practical terminology management training, the next step is to identify the common ground where theory and practice can come together. We propose that a model that would link terminology theory and practice, in the eyes of both students and educators (and, why not, practitioners as well), is that of problem-solving (Fig. 1). By their own nature, both theory and practice in terminology are oriented towards the description, comprehension and analysis of terminological problems that arise mainly in two areas: firstly the creation of new terms as new fields in science and technology emerge, and secondly, translation. Regardless which direction we take, that of terminology theory providing the background for solving a practical problem or that of terminology practice giving rise to issues which terminology theory will need to adjust to, the common denominator in both cases is the fact that a problem needs to be solved.

This is particularly so in Language for Special Purposes (LSP) texts, where issues such as lexical density, compounding and neologization can pose significant challenges to translators and terminologists alike, including the experienced ones. It is therefore important to introduce this concept of problem-solving to students, from early on, so that they become familiarized with potential problems they may have to tackle in practice and also with some ways in which terminology theory could help them deal with those problems effectively. In an ideal situation, this contextualization of terminology theory in real problems of practical terminology work can make the link between theory and practice more obvious in the eyes of students and in fact give rise to more in-depth investigation, interestingly more often in aspects of theory.

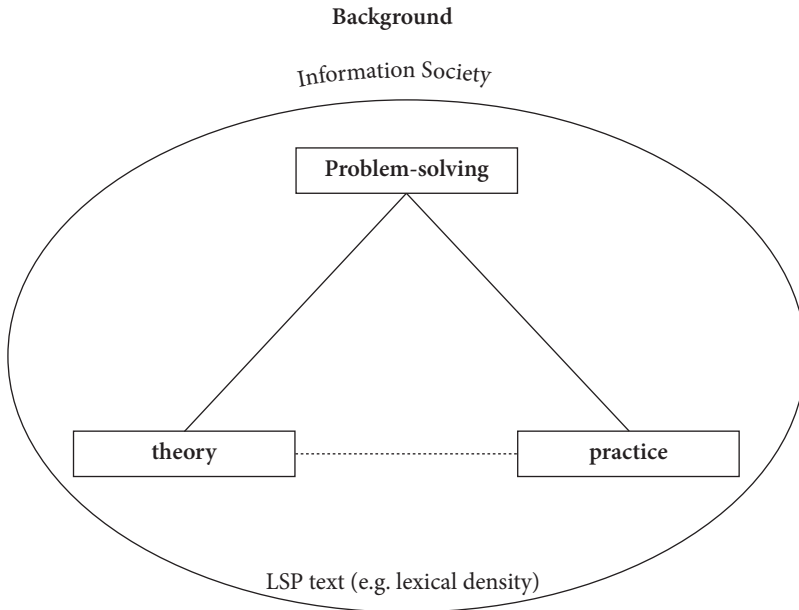


Figure 1. Linking theory and practice

In order to illustrate the possible ways in which this model could be applied, we will present three issues of terminology management in translation in some more detail. Among other issues, it is particularly in synonymy, polysemy and the identification of definitions and contextual examples where we believe terminology theory can provide a basis for supporting the decision-making process.

7. Problem-solving

Students usually start terminology classes with a lexicographical model as their base. Hence, they are accustomed to polysemes and homonyms being part of a single entry (perhaps marked as sub-parts of the same entry) and synonyms being scattered alphabetically across the dictionary (if paper) or being ignored in searches of online or other electronic resources unless a link has been hand-crafted by the lexicographer.

In our experience, common questions include:

- What exactly *is* a synonym?
- Is a new record necessary for each synonym?
- How do I deal with multiple synonyms in each language?
- Can I create cross-references?

- Do I need a different definition for each synonym?
- Do I need a contextual example for each synonym?
- What exactly *is* a polyseme?
- How do I decide which polysemes should be in my term-base?
- Should polysemes be included in the same record?
- Do I need a different definition for each polyseme?
- Do I need a contextual example for each polyseme?

Drawing on what we have presented here as ‘theory’, both synonyms and polysemes can be usefully understood in this context as term-concept relations, building here for expedient reasons on the system-focused Wüsterian claim that ‘the world of concepts is viewed in terminology as independent of the world of designations’ (Wüster, 1974: 67).¹ So in the domain of haematology ‘red blood cell’ and ‘erythrocyte’ can be related to the same concept through a single intensional definition: ‘cellular component of blood, millions of which in the circulation of vertebrates give the blood its characteristic colour and carry oxygen from the lungs to the tissues’ (Britannica Online Encyclopedia, 2008). A well-known principle of good information management – information should not be repeated in a database – can also be harnessed here as a shared definition facilitates the updating of term records.

Conversely, polysemes can be distinguished by linking each sense with a different definition. Hence, ‘plasma’, for example, can be distinguished according to its sense in haematology as ‘the clear, yellowish liquid that forms the fluid portion of blood and lymph’ (MSN Encarta Encyclopedia, 2008) and in physics as ‘an electrically conducting medium in which there are roughly equal numbers of positively and negatively charged particles, produced when the atoms in a gas become ionized’ (Britannica Online Encyclopedia, 2008).

Principles of term-base structure clearly indicate that one record represents one concept or meaning within a particular domain, and hence it follows that polysemes (two or more concepts) are entered in different records, whereas synonyms (a shared concept) are entered in the same record. Where multiple meanings are covered by a single but polysemous form, the domain becomes the deciding factor, with only one definition per language per concept, allowing for the fact that the knowledge mapping may not be isomorphic across all languages or cultures. Even within the same domain, different senses can be distinguished according to the intension of the concept such as in the case of the polysemous ‘emission’ in automotive engineering as a substance on the one hand and a process on the other hand.

Understanding the different functions of definitions (system level ‘idealization’ of concept and conceptual relations) and contextual examples (actual use of

1. Translation by MR: ‘Das Reich der Begriffe wird in der Terminologie als unabhängig vom Reich der Benennungen angesehen’ (Wüster 1974: 67).

language: formal and meaning aspects) can guide decisions on the number of definitions and the number of contextual examples needed in relation to synonyms and polysemes:

- Synonyms:
 - one *concept*, therefore one definition and one (monolingual) record;
 - several *terms* (forms), therefore several contextual examples within one record.
- Polysemes:
 - several *concepts*, therefore several definitions distributed across different records, possibly in different domains;
 - single *term* (form), therefore single contextual example within one record in each domain

If all synonyms are included in the same record linked with a single updatable definition, then the need for cross-referencing synonyms which occupy different parts of the alphabet falls away, as does the need to include more than one definition per polyseme, once these are recorded in separate records. Having understood these principles of concept-based term-base organization, students should no longer entangle themselves in multiple cross-referenced entries for each synonym, each with its own often different definition, and multiple definitions and domains for single polysemous entries.

Looking now beyond synonyms and polysemes, there are a number of questions that students ask when trying to understand what makes a good definition or contextual example and how best to use it in a term record. The problem of choosing a good definition, for example, is a real one and on many an occasion even all good available references may be insufficient in that respect. Here are some popular questions:

- Can I write my own definitions?
- How do I identify definitions in other sources?
- Do I need a definition in each language?
- Should I translate my definitions?
- Where do I place the definition in the hierarchical structure of a term record?

As we can see, some of these questions have to do with form, some with content. In both cases, however, it is evident that there is a need for guidance, particularly as students are eager to move on quickly to the data entry stage where they actually start populating the term-base. This is where terminology theory can provide a framework for evaluating—and in some cases editing—extant definitions. To this end, students are provided with good examples of the classical Aristotelian model of *genus-differentiae* as well as examples of deficient intensional definitions (e.g. circular, negative, too broad, too narrow, and so on). The use of extensional definitions (and of mixed types) is also critically evaluated.

Of course, in a structured term-base, form is also important and therefore one should expect to receive questions on that aspect as well. For example, deciding in advance (at the stage of designing a term-base) whether a definition should be available in each of the languages present is something that will affect the presentation and the structure of all the entries. It will also affect what a potential user might expect to find in this term-base, so this alone is a decision with far-reaching implications, perhaps beyond what a new user would have anticipated. Here, some input regarding equivalence relations is useful, showing the range of equivalence relations which can pertain (e.g. exact, intersection, superordination, cf. Felber, 1984: 153), even in technical domains, through the use of examples of terms in different languages linked to their definitions. So, for example, the scope of the job of a 'civil servant' in the UK can be compared to that of a *Beamte(r)* in Germany through their respective definitions, including some extensional examples, showing that some positions, such as that of school teachers, are classified as civil service jobs in the German system, but not in the UK system. Such partial equivalence draws attention to the fact that concepts, and hence definitions, are often culturally specific, indicating the need for a definition within each language which is represented in the term record. Technical examples such as those discussed by Schmitt (1999) are also illuminating in this respect; his discussion of the difficulties of matching different types of hammers and their functions in Germany and Great Britain (1999: 244–7) further reinforces the point that technical objects can also be culturally specific.

Similar considerations need to be taken into account for contextual examples and their role in a term-base. Some relevant questions here would be:

- What is the difference between a contextual example and a definition? Do I need both?
- Where do I place the contextual example in the structure of a term record?

The first question in particular is a very common one and normally it takes some time for students to realise that in a good contextual example you expect to see the term used in actual text but in a good definition you do not expect to see that term at all. However, over time and with practice, students come to see the functional difference between these two elements of term records more clearly. It is then easier for them to apply the relevant theoretical guidelines that will help them in identifying an effective definition/contextual example in the future. Exercises focused on identifying selection criteria for a given set of potential contextual examples support the link between term (as a particular form) and example. Repeating the exercise for synonyms within the same entry can strengthen the term-example link (in contrast to the term-concept link in the definition).

8. Conclusion

Terminology studies as a subject taught in the context of postgraduate translation programmes is a rather complex affair, as the right balance between theory and practice needs to be found. Our own pedagogical experience has shown that the expectations which students and trainers have of terminology theory can be quite different; in many cases, reconciling these different perspectives can be one of the biggest challenges syllabus designers are faced with. A balanced approach that reveals the value of terminology theory in terms of guidelines for good practice whilst also bringing to light the potential shortcomings in practical terminological work when terminology management software is used in an *ad hoc* manner, is perhaps the most effective way to highlight the relationship between terminology theory and practice. It could perhaps also demonstrate that, when it comes to choosing between theory and practice, there is in fact no dilemma at all.

References

- Arntz, R. (1981). Terminology as a Discipline in the Training of Translators and Interpreters in Languages for Special Purposes. In M. Krommer-Benz (ed.), *Theoretical and methodological problems of terminology* [Infoterm Series 6. Proceedings of an International Symposium, Moscow, November 1979] (524–530). München: Saur.
- Arntz, R. and Picht, H. (1989). *Einführung in die Terminologearbeit*. Hildesheim: Olms.
- Auger, P. (1979). L'enseignement de la terminologie (aspects théoriques et pratiques) dans le cadre des études en traduction et en linguistique. In *Actes du 6e colloque international de terminologie*. Pointe-au-Pic (Québec), 2 au 6 octobre 1977, 445–484. Gouvernement du Québec, Office de la langue française.
- Britannica Online Encyclopedia (2008) <http://www.britannica.com/> Last accessed: 28 October 2008.
- Budin, G. (2001). A critical evaluation of the state-of-the-art of terminology theory. *IITF Journal* 12/1–2: 8–23.
- Bühler, H. (1987). Terminologieseminar für zukünftige Sprachmittler – das Wiener Modell. *Lebende Sprachen* 1987/2: 52–56.
- Felber, H. (1984). *Terminology Manual*. UNESCO & INFOTERM: Paris & Vienna.
- MSN Encarta Encyclopedia (2008) <http://encarta.msn.com/> Last accessed: 28 October 2008.
- Picht, H. (1981). La Section de terminologie de l'École des hautes études commerciales de Copenhague, *Terminogramme, Bulletin de la Direction de la terminologie, Office de la langue française, Gouvernement du Québec* 9: 1–4.
- Picht, H. & Draskau, J. (1985). *Terminology: an Introduction*. Guildford: University of Surrey.
- Robinson, R. (1954). *Definition*. Oxford: Clarendon Press.
- Rogers, M. (1999). Translating terms in text: Holding on to some slippery customers. In G. Anderman and M. Rogers (eds.), *Word, Text, Translation* (104–116). Clevedon: Multilingual Matters.

- Sager, J. (1979). Training in Terminology: Needs, achievements and perspectives in the world. In H. Felber, F. Lang and G. Wersig (eds.), *Terminologie als angewandte Sprachwissenschaft* (149–163). München: Saur.
- Sager, J. (1981). Approaches to Terminology and the Teaching of Terminology, *Fachsprache* 3/3–4, 98–106.
- Sager, J. (1992). The translator as terminologist. In C. Dollerup and A. Loddegaard (eds.), *Teaching Translation and Interpreting* (107–122). Amsterdam & Philadelphia: John Benjamins.
- Schmitt, P.A. (1999). *Translation und Technik*. Tübingen: Stauffenburg.
- Wüster, E. (1974). Die allgemeine Terminologielehre – Ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sachwissenschaften. *Linguistics* 119, 61–106.

Ontological support for multilingual domain-specific translation dictionaries

Rita Temmerman and Sancho Geentjens

In this article, we will show how cross-linguistic and cross-cultural information can be structured in ontologically-underpinned specialized translation dictionaries, by applying the Termontography method. Examples are taken from retailing vocabulary and from the automotive domain. An ontology-based specialized translation dictionary in the automotive domain is being compiled by CVC Brussels. The terminology extracted from texts in five languages (English, French, German, Dutch and Italian) is structured by means of a cognitive/ontological model. We show how this model could contribute to the resolution of translation problems caused by cross-linguistic differences and similarities.

Keywords: Termontography, translation dictionaries, knowledge management, automotive domain, isomorphisms

1. Introduction

The need for conceptual frames in developing terminology-intensive systems has been widely acknowledged (Meyer, 2001). Methods in ontology engineering have been applied to domains such as medicine (Rector, 1999; Herre & Heller, 2006), molecular biology (Schulze-Kremer, 1998, 2002) and financial fraud (Kerremans et al., 2003; Temmerman & Kerremans, 2003) in an effort to model those conceptual frames. However, the transfer of those insights and practises to the compilation of terminological dictionaries remains problematic. In this paper, we will discuss some possibilities and limitations of ontological support for multilingual terminological dictionaries.

We base ourselves on the Termontography method. This method combines theories and methods for multilingual terminological analysis (Temmerman, 2000) with methods and guidelines for ontological analysis (Gómez-Pérez et al., 1996; Fernandez et al., 1997; Sure & Studer, 2003). In Termontography, the categories as well as their inter-categorical relations are structured in a shared knowledge

scheme or categorization framework. The termontographer starts from a predefined, language-independent framework of domain-specific categories and inter-categorical relationships, set up in collaboration with field specialists, to which lexicalizations from a domain-specific corpus are mapped (Kerremans et al., 2003). Lexicalizations referring to categories are listed in a termontological database offering not only a language-independent description of each category, but also specifications of possible culture-specific and/or language-specific items.

The English-to-French *Dictionnaire Analytique de la Distribution. Analytical Dictionary of Retailing* (2000) by Dancette and Réthoré was one of the first fully worked-out examples of an ontologically underpinned translators dictionary. The key concepts of this translation dictionary are not only described by means of a traditional definition and a context, but also by means of additional information under subheadings such as *précisions sémantiques*, *compléments d'information* and *relations internotionnelles*. The translator, a native speaker of French, is immersed in a wealth of semantic information in French about English terms.

This paper is structured as follows: In Section 2 we will describe the bilingual *Analytical Dictionary of Retailing* by Dancette and Réthoré in more detail. We will illustrate the necessity of cognitive framing in a translation dictionary and discuss problems of quasi-synonymy and non-isomorphism. Section 3 deals with our efforts at compiling a machine-readable translation dictionary. We have chosen the automotive domain because we expect fewer non-isomorphism problems in this much more technical field, as compared to the retailing vocabulary dealt with in the dictionary of Dancette and Réthoré. This hypothesis is inspired by the fact that the retailing domain is structured according to a culture-specific (legal) system, which differs from country to country (even from region to region). The automotive domain, however, concerns an artefact: the automobile and less variation in categorization and terminology is expected.

2. Analytical dictionary of retailing

2.1 Conceptual foundation

The *Analytical Dictionary of Retailing* (ADR) is compiled for a user group of translators from English to French on the subject of retailing. The authors believe that a translator with French as his native language will benefit from being submerged in a wealth of discursive information, i.e. information on how the term or phrase which has to be translated is related to other terms in the same field or semantic network of related terms. This information is given in the target language (French)

to stimulate the ‘discursive autonomy’ of the translator in phrasing the target language text (Dancette, 2000; Temmerman, 2003).

The advantage of a cognitively underpinned translation dictionary over a traditional translation dictionary can be illustrated by the North-American concept ‘superstore’. In the ADR, the literal translation of ‘superstore’ into French is *supermagasin*. This literal translation is only used in Canada. The English term *superstore* and the French term *supermagasin* both refer to the same reality in bilingual areas in Canada. The cross-cultural translation into French spoken in France, however, is much more complicated. The ADR distinguishes two meanings of ‘superstore’. In the first meaning, ‘superstore’ is ‘une formule de distribution qui, sur le plan de l’assortiment et de la taille, se situe à mi-chemin entre le supermarché et l’hypermarché français []’ (Dancette & Réthoré, 2000: 220). In the second meaning, ‘superstore’ means ‘une grande surface spécialisée dans toute catégorie de produits autre qu’alimentaire’ (Dancette and Réthoré, 2000: 220). Translators need a solid cognitive framework of the retailing sector both in the Anglo-Saxon countries and in France in order to translate ‘superstore’ correctly in any specific context. The English-French general dictionary *Le grand Collins-Robert électronique* only gives *hypermarché* for *superstore* (Figure 1).

The specialized *Dictionnaire de la comptabilité et de la gestion financière, anglais-français* by Louis Ménard et al. (2004) does not offer enough information in order to allow for a correct cross-cultural translation in all contexts (Figure 2).

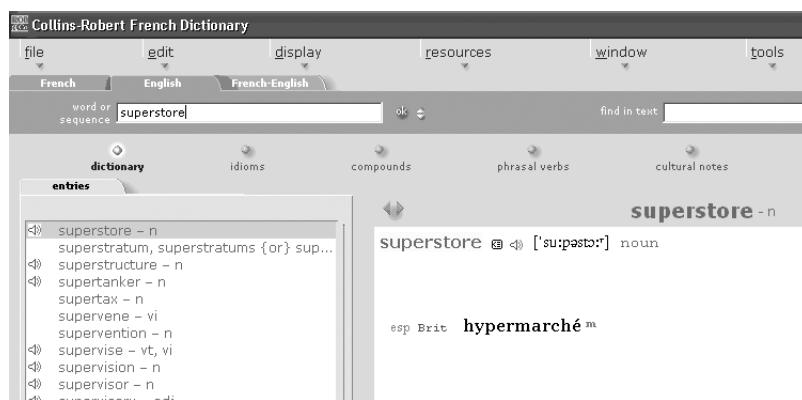


Figure 1. Superstore in *Le grand Collins-Robert électronique, anglais-français*

SUPERSTORE (MAGASIN À GRANDE SURFACE) <i>Commerce.</i> Magasin de détail qui dispose d'une surface de vente importante, et qui vend en libre-service et à des prix compétitifs un assortiment très large et très varié de produits alimentaires et non alimentaires.

Figure 2. Superstore in Ménard et al. (2004)

As shown in this example, term, translation and definition provide insufficient information in order to solve cross-linguistic and cross-cultural terminological problems. Conceptual links (in order to improve the understanding of the specialized domain) and extralinguistic/encyclopaedic information (in order to improve the understanding of terms and categories of source and target language in the specialized domain) often prove essential. The ADR provides much of this vital information:

- *Relations internotionnelles*: semantico-conceptual links to establish a cognitive structure of the specialized domain,
- *Compléments d'information*: extralinguistic/encyclopaedic information e.g. diachronic information.

Under the subheading *relations internotionnelles* in the entry SUPERSTORE 1 the ADR gives essential information for the cross-cultural translation of 'superstore' in its first meaning.

As shown in Figure 3, the translator has to choose from three French shop concepts: *supermarché*, *hypermarché* and *hypérette*, depending on the context (assortment and sales area of the shop). A conceptually framed terminological dictionary like the ADR makes explicit the differences and similarities between concepts as well as the relations with other concepts. This information facilitates the translator's decision making efforts.

SUPERSTORE 1 RELATIONS INTERNOTIONNELLES
Le supermagasin 1 est une formule de distribution qui, sur le plan de l'assortiment et de la taille, se situe à mi-chemin entre le supermarché (=> SUPERMARKET) et l'hypermarché (=> HYPERMARKET) français ou le magasin combiné (=> COMBINATION STORE) américain. De par sa taille, entre 25 000 et 50 000 pi² (2 323 à 4 654 m²), le supermagasin 1 se rapproche du très grand supermarché français ou hypérette dont la surface de vente varie entre 2 500 et 3 500 m².

Figure 3. Relations internotionnelles in Dancette and Réthoré (2000)

2.2 Problems with formalization

Dancette (1998) gives an account of the problems encountered while formalizing the ADR. The author states that she has arrived at a deadlock, due to problems with quasi-synonymy and non-isomorphism.

Quasi-synonymy, the multitude of terms for the same reality, makes it very difficult to conceptually structure a domain (Dancette, 1998). The English terms ‘superstore’, ‘category killer’, ‘big-box store’, ‘destination store’ and ‘chain store’ for example, are all used in the specialized press when referring to a shop like IKEA. The associative relations between the English terms are difficult to establish. The terms share a great number of semantic features, as shown in Table 1 (Dancette, 1998).

Non-isomorphism i.e. different structuring of reality in different languages and cultures, causes translation mismatches. Dancette (1998) indicates three types of terminological equivalence problems in the bilingual multicultural setting of retailing vocabulary in English and French:

1. partial equivalents, e.g. *grande surface spécialisée* for category killer;
2. converse equivalents, e.g. *electronic shopping* for *vente électronique*,
3. terminological gaps due to the disparity of hierarchical structures, e.g. the generic French term *vente à distance* does not have an English equivalent.

The comparison of the structure of retail stores in Canadian English to the structure in Canadian French shows perfect one-to-one equivalence relations. In this case of isomorphism, one model of language-independent categories can be built.

The comparison of the terminological structure of retail stores in Canadian English to the terminological structure in French spoken in France, however, shows a lot of partial equivalents. In case of non-isomorphism, two frameworks are particularly useful. The associative relations between partial equivalents are as difficult to establish as the relations between quasi-synonyms. The formalization of those relations should also resort to an analysis by means of semantic features (Table 2).

Table 1. Semantic features shared by ‘category killer’, ‘superstore’ and ‘big-box store’

Semantic features	<i>category killer</i>	<i>Superstore</i>	<i>big-box store</i>
big surface	+	+	+
self-service	+	+	+
specialized assortment	+	+	Ø
non-food products	+	+	Ø
reduced prices	+	+	+
sober facilities	Ø	Ø	+

Table 2. Feature analysis of the categories ‘superstore’ and ‘supermarché’

Superstore	Supermarché
– assortment: – food: very wide assortment – non-food: wide assortment (household products, clothing, kitchen utensils, gardening tools...) – area: 2300 to 4600 m²	– assortment: – food: wide assortment – non-food: fairly wide assortment (household products, clothing...) – area: 400 to 2500 m²

3. Towards an ontology-based translation dictionary for the automotive domain

Within the framework of the OOV (*Ontologisch Onderbouwde Vertaalwoordenboeken*, Ontology-based Translation Dictionaries) project, CVC is developing a specialized translation dictionary for the automotive domain. The terminology was extracted from specialized texts in five languages (English, French, German, Italian and Dutch) by four Master students of *Erasmushogeschool Brussels*. The terminological information, often with definition and context, was then structured in a language and culture independent categorization framework.

Our categorization framework is constructed by means of the Termontography Workbench (TW). This software tool was developed by CVC Brussels in order to support the different methodological phases of the Termontography method (Kerremans et al., 2004). Figure 4 shows part of the categorization framework of the automotive domain in the TW.

The TW can be used to compile and manage a domain-specific corpus in different languages, to build a categorization framework and to extract relevant terminological information from the corpus using the CF as a reference point.

The automotive domain has far more isomorphism (the structure of an automobile is the same across languages/cultures) and far less near-synonymy/partial equivalence (since terms refer to the same physical reality/phenomena) than the retailing domain. We could therefore benefit from the power of a cognitive frame, as described in Section 1, without having to deal with the more intricate terminological problems described in Section 2. However, we still encountered numerous problems when structuring and aligning the terminology, such as terminological overlap in different engineering branches.

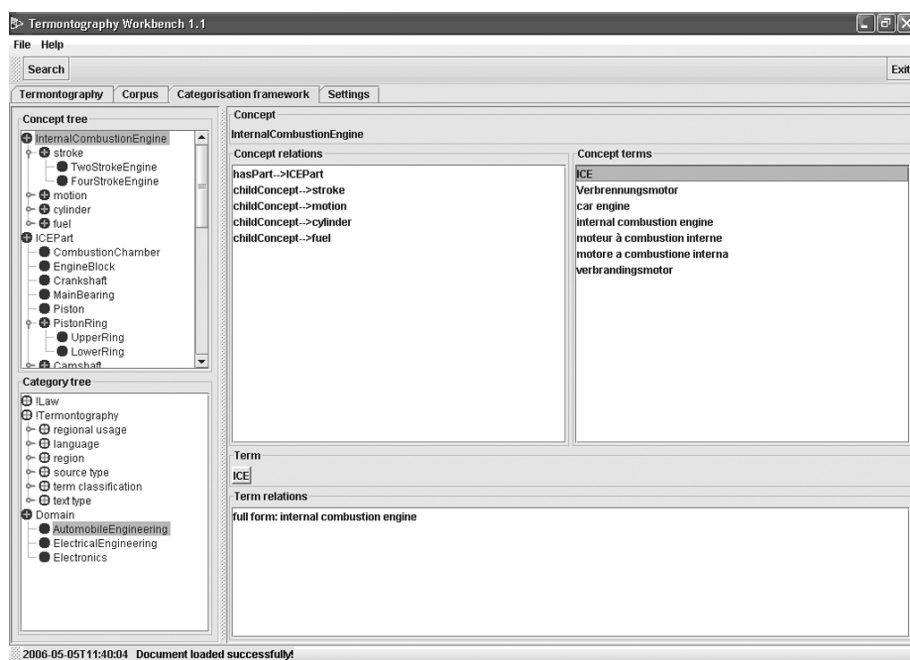


Figure 4. Part of the categorization framework

Automobile engineering originated as a by-product of the mechanical engineering industry. It now involves many other fields of technology. Complex electrical and electronic devices are needed for tuning the engine, monitoring fuel consumption, adjusting the air-conditioning or operating the windscreen wipers. Chemical and metallurgical processes are required for material-strengthening, as well as rust-protection of the bodywork, chassis and suspension system. Sophisticated robot technology is employed for purposes such as production assembly, welding, testing, cable-positioning. Each field of technology came with its own terminology and has made the automobile terminology rather inconsistent and unpredictable.

An example of this is the terms 'capacitor' and 'capacitance'. Both terms are firmly embedded in all fields of electrical technology. The older terms *condenser* and *capacity*, which date from the very early days of electrical engineering, have become virtually obsolete; but they persist as fossils in small subfields cut off from the main streams of science and technology, e.g. hair dryers or model railways, and occasionally crop up in connection with automobile and other ignition systems. 'Capacity' is not to be confused with 'capacitance'. The capacity of a battery is the amount of charge it can hold, a figure measured in coulombs or the more usual unit ampere-hour (A.h). Capacitance is measured in farads, a unit defined in

Table 3. Five categories related to automobile engineering with lexicalizations in five languages

Category	En	Fr	Ge	Du	It
amount of charge	Capacity	capacité	Kapazität	capaciteit	capacità
amount of charge per volt	Capacitance	capacité	Kapazität	capaciteit	capacità
amount which a tank can contain	Capacity	capacité	Füllmenge	capaciteit	capienza
opposition to electric current	Impedance	impedance capacitance	Impedanz	impedantie	impedenza
rate of doing work	power capacity	puissance	Leistung	vermogen capaciteit	potenza

terms of coulombs/volt. The two concepts denote different physical quantities with different units and different dimensions. However, this distinction is only made in English, and not in German (Kapazität), French (capacity), Dutch (capaciteit) or Italian (capacità). French does have a term ‘capacitance’, but it means ‘impedance of a capacitor’, defined in ohm. ‘Capacity’ can also be used in the sense of ‘power’, defined in watt, but only in English and Dutch. German has the term ‘Füllmenge’ for fuel tank capacity. Italian makes the difference between the electrical phenomena and maximum volume too: ‘capacità’ and ‘capienza’. The five categories and their lexicalizations in the five languages are shown in Table 3.

The categories in Table 3 are mutually exclusive. This means that there is no quasi-synonymy and that the terms are fully equivalent across the five languages. Therefore, in a dictionary structured according to a language-independent categorization framework, what may initially seem intra- and inter-linguistic terminological confusion can be overcome

4. Conclusion

Cognitive frameworks underpinning terminological dictionaries are an asset for the translator. Submersion in a wealth of interrelated information helps him/her to make fast and accurate translation decisions. Moreover, the dictionary becomes more easily adaptable and reusable. Formalizing the cognitive framework, however, remains problematic, due to problems of overlapping terminology between related fields and languages and because of non-isomorphism. Building a separate model for each language could resolve some of those problems, but it is, of course, highly labour-intensive and difficult to reuse and to adapt.

The OOV project taught us that in a technical domain it is feasible and desirable to develop a translation dictionary related to a cognitive framework. In our

future work, we want to tackle the terminology of other more culture-dependent domains such as senior care, in order to further explore the possibilities of our methods and software. In a pilot study for the Belgian Federation of Independent Senior Care (Federatie Onafhankelijke Seniorencare, FOS), we have already developed a limited resource for the senior care domain (De Baer and Temmerman, 2006b). Setting up a language-independent categorization framework for this domain, which holds a lot of legal terminology, was found to be harder than for the automotive domain.

References

- Dancette, J. (1998). Le potentiel du dictionnaire spécialisé bilingue électronique. *Actes Sélectifs – Euralex98 Proceedings*. Université de Liège, 387–396.
- Dancette, J. and C. Réthoré (2000). *Dictionnaire analytique de la distribution. Analytical dictionary of retailing*. Les Presses de l'Université de Montréal.
- De Baer, P., K. Kerremans and R. Temmerman. (2006). Bridging Communication Gaps between Legal Experts in Multilingual Europe: Discussion of a Tool for Exploring Terminological and Legal Knowledge Resources. *Proceedings of the Euralex conference 2006, Turin*, 813–818.
- De Baer, P., K. Kerremans and R. Temmerman (2006). Facilitating Ontology (Re)use by Means of a Categorization Framework. *Proceedings of the AWeSOMe workshop*, Montpellier, France, October 2006, 126–135.
- Fernandez, M., A., Gómez-Pérez and N. Juristo (1997). METHONTOLOGY: From Ontological Art Towards Ontological Engineering. *Spring Symposium Series. AAAI97*, Stanford, USA.
- Gómez-Pérez, A., M. Fernandez and A. De Vicente (1996). Towards a Method to Conceptualize Domain Ontologies. *Workshop on Ontological Engineering. ECAI96*, 41–52.
- Herre, H. and B. Heller (2006). Semantic foundations of medical information systems based on top-level ontologies. *Knowledge Based Systems* 19 (2), 107–115.
- Kerremans, K., R. Temmerman and J. Tummers (2003). Representing multilingual and culture-specific knowledge in a VAT regulatory ontology: support from the termontography approach. *Lecture Notes in Computer Science*, vol. 2889 / 2003, 662–674.
- Kerremans, K. (2004). Categorization Frameworks in Termontography. In Temmerman, R. and U. Knops (eds.), *The Translation of Domain Specific Languages and Multilingual Terminology Management. Linguistica Antverpiensia New Series* 3/2004. Antwerp: Hogeschool Antwerpen, 263–277.
- Kerremans K., R. Temmerman and J. Tummers (2004). Discussion on the requirements for a Workbench supporting Termontography. *EURALEX 2004 proceedings*, Lorient, France, 6–10 July 2004, 559–570.
- Ménard, L. et al. (2004) *Dictionnaire de la comptabilité et de la gestion financière*, 2e édition.
- Meyer, I. (2001). Extracting knowledge-rich contexts for terminography. In D. Bourigault, C. Jacquemin and M.C. LHomme (eds.) *Recent advances in computational terminography*. Amsterdam: John Benjamins.
- Rector, A.L. Clinical Terminology: Why is it so hard? *Meth Inform Med* 1999; 38 (4–5): 239–252.
- Schulze-Kremer, S. (1998). Ontologies for Molecular Biology. In *Proceedings of the Third Pacific Symposium on Biocomputing. AAAI Press*, 363–383.

- Schulze-Kremer, S. (2002). Ontologies for molecular biology and bioinformatics. In *Silico Biology* 2, 17, 179–193.
- Sure, Y. and R. Studer. (2003). A methodology for Ontology-based Knowledge Management. In John Davies, Dieter Fensel and Frank Van Hamelen (eds.), *Towards the Semantic Web. Ontology-Driven Knowledge Management* (33–46). New York: John Wiley and Sons, 33–45.
- Temmerman, R. (2000). *Towards New Ways of Terminology Description. The Sociocognitive approach*. Amsterdam: John Benjamins.
- Temmerman, R. and K. Kerremans (2003). Termontography: Ontology Building and the Sociocognitive Approach to Terminology Description. *Prague CIL17-conference proceedings*.
- Vossen, P., W. Peeters and P. Diez-Orzas (1997). The Multilingual design of the EuroWordNet Database. In *Proceedings of the IJCAI-97 workshop Multilingual Ontologies for NLP Applications*. August 1997, Nagoya, 23–29.

SECTION III

Possibilities of terminological databases for different applications

In praise of effective export terminology

Danielle Dubroca Galin, Ángela Flores Garcia,
Valérie Collin Meunier and Marc Delbarge

The aim of this paper is to describe the terminology used to market certain local products in their country of origin and the transfers this yields in translated international trade documents in Spain and French-speaking European countries. Terminology used in the field of marketing, unlike that used in other areas, is in a state of constant change, and this adds to its complexity. The paper discusses the introduction and acceptance in the importing country of terms used in the exporting country and examines in particular the cases of Armuña lentils, Iberian ham and cured meat products, foie, and Belgian chocolates and a typical Spanish confectionary product from Avila. The importance of teaching future professionals an international exchange terminology to make them aware of the key role played by efficient marketing communication is also examined.

Keywords: terminology; international trade; translation; marketing communication

1. Local products in international trade

Our project: *Traducción y marketing: Exportar los productos y servicios de nuestras tierras*¹, revolves around local produce in the broad sense of the term, since it embraces both farm produce and tourist products (mutual promotion of the cultural heritages of France and Spain). Even this may be seen as little more than a provincial effort, the research project does have a European dimension, in that it places the problem inside an import as well as an export context. It may only be an attempt at simulation, but the findings – however modest – have proven to be of use to a number of enterprises which understand the linguistic importance of communication. Its main aim is to redress a perverse situation in the fields of communication and translation and to make language users aware of the key role played by efficient marketing communication.

1. This project was funded by the regional government of Castile and Leon, Spain.

The project is related to terminology issues (amongst others) on three different levels: As a scientific research project, it starts from specialized literature and the study of authentic documents written in (export) jargon. Classroom experience gives it a strong educational dimension and thus opens new didactic roads for others. The project also fits inside a professionalization context in that it deals with sales, more specifically with how regional produce and tourist services are marketed.

Its European nature adds yet another dimension to the project and situates terminology issues inside the framework of specialized translation.

Marketing edible consumer goods (farm produce) or local services (tourist services) relies on a terminology that goes from mere name-giving to the phraseology associated with each individual product – an observation that applies to all products associated with a specific cultural group.

In the case of export, three questions impose themselves:

1. How eager is the receiving country to adopt terms coined in the producing country?
2. How does the receiving country welcome such terms?
3. What is the adjustment process, considering the linguistic/cultural elements within each group?

The research project examines many products, but in this paper we will focus on lentils, Iberian ham/salt meats, and confectionery (a selection that allows us to work from French into Spanish and vice versa).

Disclosing information about edibles through terminology starts as an internal (local) issue. Upon arrival in the new (importing) country, however, products must fit an existing terminology mould for marketing purposes (linguistic terminology transfer, possibly even translation) and respect the existing distribution norms.

The problems occur on different levels. For the purpose of this exercise, a number of Spanish-French/French-Spanish examples have been studied, ranging from the strictly linguistic (phonetics, spelling, syntax, semantics) over the discursive (the term may have an argumentative trade value) to the cultural (mental stereotypes, gastronomy, diet, etc.).

When exposed to a consumer public, terminology comes under great strain. Having left its initial habitat – the purely scientific realm (in the broad sense of the word) – it starts a new life in forms and shapes that can be as varied as they are unexpected. Alongside everyday words, the products/services under scrutiny have also generated a scientific terminology: botanists will discuss lentil plants, cocoa beans and the acorns fed to Iberian pigs, chemists make their linguistic contributions to the food industry, lawyers add their own legal jargon, while art historians do their linguistic bit for tourism.

The best way for those words to make a quick entry into the consumer sphere is via a marketing pitch.

2. Terminology and marketing

New research has changed the views on a terminology that is itself the direct result of myriad efforts to communicate more efficiently. Terminology is no longer the exclusive realm of linguistics and specialist knowledge spreads its message through various resources. Terminological units form a strong but by no means the sole component of specialist discourse (Cabr , 2002), terminology indeed being part of a communicative and discursive approach in a cognitive, linguistic and social context.

Translation, as a theoretical and practical discipline, has evolved along the same lines. During the sixties, the first attempts at categorizing translation problems focused on systematically describing each language and on separately comparing lexical/grammatical elements and structures, in search of existing equivalents. By the seventies, however, foreign languages were viewed as communication systems. The transmitter/receiver concept was replaced by the interlocutor perspective, which states that semantic information is directly linked to pragmatic intention. In other words: text is viewed as a product of human activity, itself under tribute to a number of social constraints.

As translation practitioners, we approach texts (source and target ones) within the framework of applied linguistics. From a socio-linguistic point of view, however, we also see language in its social dimension 'as a structure and a communication tool, as a system and an answer to society's need to communicate and to inform.' (Cabr , 1998)

Within our research context (translating texts with a view to exporting regional products and services) translations are really new texts geared to a different linguistic community, the main idea being to incite people to buy what are often entirely new products for them.

Within today's client-oriented economic context, the market sees itself as the marketing reference par excellence. Product transfer must, in other words, take the various possible aspects of the new context into account: linguistic, cultural, ideological, etc. As a result, translators must acquire a thorough knowledge of both product *and* recipients/consumers – their preferences, linguistic/gastronomic predilections, but also their buying habits, etc. Differences between various consumer groups continue to play an important part in how purchases are triggered.

Recent research into consumer attitudes to edibles has shown that the food market is far from 'globalized'. Eating is a biological activity, but also a social one that clearly places individuals inside a community. (Muchnik, 2003, 2006) One way of

keeping up with today's constantly changing world, is to consume products that place you inside a certain group, a way of eating and thus help determine your 'identity'.

This is where various national/European labels of origin come in, each of them guaranteeing a food's pedigree by giving it certain quality labels (in France: *rouge*, *NF*, European), competition or exhibition medals, an *IGP* (geographic indicator), *Appellations d'origine contrôlée* AOC, *Appellations d'origine protégée* AOP, AB certificates for organic produce or prizes awarded to groups of producers.

In Spain this includes specifiers such as: *Alimentos de Calidad Diferenciada*: *Denominación de Origen Protegida* (DOP), *Indicación Geográfica Protegida* (IGP), *Tradiciones Garantizadas* (E.T.G.), *Producción ecológica* (the latter has not yielded an acronym yet, it being a relatively recent phenomenon).²

Marketing national or international products also transfers a social image, itself based on a number of trump cards that feed on public perception:

1. Origin, in a double role: unlike mass-produced goods, products with a label of origin are never anonymous and their local roots make them easy to trace. Identifying a region of origin also relies on the underlying power of 'distinct taste registers, a particular relief'. (Giraud, 1999)
2. Quality: itself revolving around health – the top EU consumer concern (Giraud, 1999) – as well as excellence (these are all natural products, or at least considered to be such, and invariably of superior quality).

In short: the idea is to *sell better* rather than *sell more*.

Of old, each professional sector and product has generated its own, typical terminological field, with terms gradually entering the public field, sometimes even becoming quite common. In the case of ham, for example, this includes the French words *séchage*, *affinage*, *salaison*, *flaveur*, *à consommer chambré*; for chocolate, they include *ballotin*, *fouillage*, *ganache*, *truffe*, *nougatine*, etc. – all of which have been relegated to the realm of the banal.

Thanks to an increased standard of living, consumers can now purchase better quality products. Choosy customers with increased purchasing power appreciate luxury and are eager to assimilate new, increasingly scientific culinary terms. Those same customers, within the same market segment, demand immensely detailed information about their purchases and expect maximum scientific information. The producer or distributor must try and meet those demands and give information about production circumstances (the famous 'traceability'), nutritional value, chemical characteristics, dietary benefits, etc. The following French example was taken from an Internet site promoting Iberian ham:

2. Taken from MAPA Spanish Agriculture, Fisheries and Food Ministry (www.mapa.es).

*[Le jambon ibérique] apporte des vitamines B1, B6, B12 et acide folique, très bienfaisants pour le système nerveux et pour le bon fonctionnement du cerveau. Il est riche aussi en minéraux, comme le cuivre, essentiel pour les os et les cartilages, fer et phosphore. Le jambon ibérique [] contient de l'acide oléique dans presque la moitié de ses graisses. Celui-ci est le meilleur pour combattre les maladies cardio-vasculaires.*³

Better informed than ever, consumers want to find out more and feel they can relate to a terminology that concerns them directly. Once picked up, the terminology will be used for their own benefit and removed from its traditional, specialized, limited context in the process.

3. Examples

3.1 Armuña lentils

French consumers swear by Le Puy lentils. In fact, the word 'lentil' barely seems to exist in its own right in France (except perhaps for botanists). In a sales context everyone automatically links it with the phrase '*lentille verte du Puy*'.

Today a number of lesser known Spanish lentil varieties are coveting their own place on the market. The challenge consists in making people – in Spain and elsewhere – aware of the existence of this top quality lentil whose light green colour makes it easy to recognize, while its origin, la Armuña, is easy enough to pinpoint.

Both elements provide Spanish with a perfect, specific word for distinguishing between *la Lenteja de la Armuña* and other varieties such as *la lenteja pardina*. Sadly, everyday language is not always eager to withhold such distinctions. Supermarkets may praise the qualities of the *Lenteja de la Armuña* to justify its price, gourmet menus may incorporate *Lentejas de la Armuña*, but housewives are yet to start thinking in terms of serving lentils to lure their families to the dinner table.

The difficulty with finding a French equivalent lies in the phrase's spelling and phonetic configuration. Most non-native users are sufficiently familiar with the Ñ, so keeping the reference to the product's origin should be easy. But is it in the product's best interest to keep the article in French, or would an apostrophe be better?

As far as we know, Spanish producers are yet to decide this issue, independently of their excellent relationship with the producers of Puy lentils. Usage will no doubt offer a solution, and probably push Spanish towards the handiest French equivalent – possibly even the one advocated by us: *lentilles d'Armuña*.

Terminological transfer from one language to another is not always a smooth process, sometimes for reasons outside terminology (in this case: phonetics and syntax).

3. For what this web translation is worth...

3.2 Foie gras

The difficulty with this term lies in the transfer from French to Spanish. As a product, *Foie gras* has 'Romanian' roots, even though it has been culturally claimed by France (various regions, but mainly the South West, which even got itself an IGP). It is now also being produced and marketed in Spain, more specifically in Castilla y Leon. An extensive glossary of French terms has already been compiled at Pau. Finding equivalent Spanish terms, however, is proving much harder.

The term *magret* (the white meat of a force-fed duck) is a first example. Hardly a problem, at first sight, since the word entered the Spanish language years ago – even though consumers may not always understand its exact meaning. The word is derived from Gascon dialect and refers to very lean meat (*magre*) that is not too dry, and sweet-tasting (affective diminutive in *-et*). The term *magret*, rightly associated with the word *pato* (duck), is found on labels, but also in all leading Spanish gastronomy publications.

The term *aiguillette*, on the other hand, more specifically of duck (fillets that remain on the carcass after the *magrets* have been removed on either side of the wish-bone), is by no means established in Spanish. Existing terms are used only for chicken. Spanish also has *pechuguita* and *aguja*, which may cause confusion, since cockerels producing small whites are also marketed in Spain (*pechuga/pechuguita* have nothing to do with duck). When asked the question, a Salamanca shopkeeper spontaneously came up with *solomillín*, a local diminutive typical of that part of Spain, itself related to a word borrowed from the pork meat trade (small *filet mignon*). It will be interesting to see which of the various solutions triumphs in the end.

The price of genuine *foie gras*, produced to very strict specifications, calls for a highly precise sales terminology that is binding for the seller and satisfies the consumer. French, for example, makes a difference between *Foie gras d'oie* and *foie de canard*, a distinction that is carried through in both animal camps with terms such as: *Foie gras entier* / *bloc de foie gras* / *galantine de foie gras* / *mousse de foie gras* / *foie mi-cuit* / *foie en conserve*.

Our research also shows that some producers refrain from using common marketing speak and prefer words such as *bloc de foie gras*, to which no connotation is attached. The term refers to a middle-range to low-range product which less well-off consumers love to eat and supermarkets love to sell (quality not necessarily being a top priority). Producer/manufacturer *Le Château de Bellevue* (in the Pyrénées Atlantiques), for example, calls his product *foie gras reconstitué* to steer clear of *bloc de foie gras*, which has been over-used and, as a result, lost its prestige. The same goes for *Foie gras 2ème catégorie avec 30% de morceaux de foie 1ère catégorie*, a sales term that is hard to manipulate but does avoid the banal *galantine de foie* or *ballottine de foie*.

Some manufacturers wishing to outdo the competition, go for terms such as: *Foie gras aux fruits*, *Foie gras aux kiwis*, *Foie gras truffé*, *Foie gras à l'armagnac*, etc., or *canards gavés au maïs non-OGM*, which is becoming more and more frequent and hovers over the thin line between phraseology and terminology.

Sales terminology clearly doubles up into, on the one hand, terminology codified by a particular professional group, and on the other, looser terminology launched by professionals via individual promotion efforts and recognized, understood and manipulated by consumers.

Spanish has several corresponding terms, such as *Foie de pato*, *Hígado de pato*, *Hígado confitado*, to name but a few. This reluctance to pick one term negatively influences the product. Spaniards know the root of this problem: one of last century's new pork delicacies was a spread containing pork and other ingredients. In an effort to give this very popular product a prestigious ring, it became known as *Foiegras*, a gem of a Gallicism ('foreign' and 'better' were part of the same equation in those days...).

Had the first maker of this sad ersatz product settled for the term *pâté* (much closer to the product's reality) the term *foie gras* would have remained unburdened. Gallicisms being widespread in Spanish gastronomy, *pâté* would have easily been adopted. Today, confusion reigns, and Spanish makers of real *foie gras* are having to make superhuman efforts to get this luxury product recognized and to distinguish it from what the French call *mousse de foie* or *pâté de foie*.⁴

3.3 Yemas de Ávila

The moniker *Yemas de Ávila* refers to what must be one of the best known Spanish confectioneries. Confectioneries themselves, however, are still treated in a step-motherly fashion in Spanish trade circles. The poor, restricted vocabulary for sweet products in general testifies to this. The word *dulces* (*douceurs* would be a possible French equivalent) embraces what in French includes *sucrieries*, *bonbons*, *petits-fours*, *biscuits*, *galettes* as well as *gâteaux* or *pâtisserie*. Sweet foods are also spontaneously associated with greed, and thus have a rather negative ring to them.

French and Belgian nationals view confectionery as a full-fledged sector, revolving around products that tend to be expensive because hand-made, perishable and sold in attractive but dear packaging.

Yemas of Ávila, are associated with the souvenir trade. In their marketing efforts, cities or regions indeed increasingly interweave a product's history, gastronomy and peripheral tourist activities. In France, this is the case with the *calissons*

4. Terms from the French pork meat trade that made their way into everyday usage.

from Aix en Provence, *nougat* from Montélimar, *cannelés* from Bordeaux and many other products that immediately conjure up a tourist/cultural association.

Every year, an estimated 250 000 boxes of *Yemas* are produced in Ávila. The origin of this confectionery is both unclear and uncertain⁵. We do know that the 'Flor de Castilla' first commercialized and distributed the sweets as *Yemas de Santa Teresa*. Their *yema*, made to an ancient recipe, is not the only one on the market, though. Other makers use the name *Yema de Ávila* and sell a product that closely resembles the former. What was called *Yemas de Ávila* has become a sales trump and yielded a number of marketing variants. *Yemas de Santa Teresa*, however, has been registered as a trademark and is implicitly associated with the city. To Spaniards, Saint Theresa equals *Ávila* and vice versa. Tourists, on the other hand, would have a hard time distinguishing between various brands. As a result, *Yema de Ávila* has become a generic term, with individual makers stressing possible, tiny differences.

What would be an acceptable French equivalent of the term *Yemas*? Bilingual dictionaries translate it as 'egg yolk', the product's main ingredient.⁶ Associating sweets and egg yolk doesn't make the French lick their chops, but rather conjures up visions of saturation and nausea. It might be a better idea, in other words, to stick to *Yema*, which is both easy to pronounce and remember ('egg yolk' clearly being an unhappy choice in French).

And what French category of sweets does it belong to? The word *dessert*, used in most brochures and texts about the city, as well as on its official website, is an unfortunate choice. Calling *las Yemas* a dessert plainly leads to confusion and limits their consumption. The confectionery can, in fact, be eaten at any time (also for dessert, albeit in limited quantities). The French might be tempted by the thought of *yemas* over coffee or at tea time and prefer a single reference along the lines of *douceur*. The hardest hurdle to take, however, remains that of getting the manufacturers to agree...

To impose a previously unknown product in a receiving country, a precise term that can easily be adopted by its new consumers is needed. Bringing in other factors, especially cultural ones (tastes, eating habits) may also make the product more attractive. A term certainly does not have to be translated to cross borders.

5. It is most certainly a convent product, typical of wine-making regions: in the past, people clarified wine with egg whites while the yolks were recycled, often by religious communities, and made into sweets. The Portuguese make a similar confectionary from 'dôce de ovos'.

6. Flor de Castilla's website actually starts with a photo of four eggs...

3.4 Pralines

French-speakers in France know *pralines* as sugar-coated, caramelized almonds and will use the word *chocolat belge* to refer to the chocolate-based sweet that in no way resembles the French *praline*. We are clearly in the presence of a divulgatory term that does not correspond to the original.

The best solution for Spanish, or the lesser of two evils, was to opt for *chocolate belga*. The confectionery is indeed slowly making a name for itself in Spain, thanks to the popularity of market leaders such as Sven franchises, Leonidas, Neuhaus and a number of other brands known mainly by chocolate lovers. *Chocolates belgas*, complete with Flemish descriptions, can now be found on supermarket shelves (especially around Christmas). There is no reason, however, why the loan translation *praline* should not have made its way into Spanish, since the language does make a distinction between chocolate (the ingredient made from cocoa beans) and *bombóns* – the Hispanic take on Gallicisms – i.e. a *bouchée au chocolat*, or *chocolat*, to French-speakers in France. The future will tell how this issue is resolved.

3.5 Ham

Fifty years ago all Spanish ham was by definition Iberian, only one pig species being bred locally (black or dark brown, depending on the region). Tales of rooms at Irun station, stocked to the rafters with hams (complete with black trotter and tiny hanging cord) confiscated from immigrant workers travelling to Northern Europe, are still being told. The term *pata negra*, however, was never used in the past.

Years went by and other culinary habits took hold: the traditional *embutido* (another Iberian pork product) competed with other meat products made from Jersey-Duroc pork, a much more profitable alternative.

Later still, cooked ham (referred to by consumers as *jamón de York* and commonly pronounced [*jamonyór*]) made its appearance. As was the case in France, the new variety did resemble York ham (cooked whole on the bone). Gradually, more accurate commercial monikers – coined not to damage the genuine, traditional York ham – imposed themselves in French and Spanish: *jambon de Paris*, *jambon blanc*, *jambon à la machine*, *jambon à los*, *jambon cuit au torchon*; *jamón cocido*, *paleta cocida*, *fiambre para sandwich*, etc. French even added supplementary quality indicators such as *label rouge*, *supérieur*, *tranché fin*, *choix*.

Only true connoisseurs continued to appreciate Iberian ham, while the rest enjoyed old-style dried ham made from Jersey-Duroc pigs. Consumers either belonged to the *Jamón* or the *Jamón de York* camp.

Following the increased interest in traditional products in general, Iberian pork products have become incredibly popular and now occupy a front row

position on national and international markets; hence also the need to distinguish between *Jamón serrano* and *jamón ibérico*. The term *jambon* /*jamón* has, in fact, been subdivided into two terms. A qualifier was added, which gave rise to two new pork meat terms that naturally also imply hugely different prices: today, *Jamón serrano* is a quality label in its own right. *Jamón ibérico*, and a number of variants (everything depends on what the animal was fed), has now also carved its own niche and a distinction is made between *Jamón ibérico de bellota*, *Jamón ibérico de recebo*, and plain, much cheaper *Jamón ibérico*.

Serrano ham is currently being diversified in yet another manner. Like cheeses, hams are kept and matured in special rooms for several months. The number of months hams – even simple *serrano* – are kept, earns them a supplementary age label. Waiting in a cave before being sold implies a certain immobilization, which in bookkeepers' terms translates into an added product cost. Mainly on promotional websites, marketing speak now also includes quality specifications such as '*Bodega, reserva, gran reserva*', indicating how long a specific ham has matured.

Consumers must be able to tell the difference, hence the need for terminological divulgation, partly by big supermarket chains.⁷

French consumers will have no problems adopting a term like *Jambon serrano* (certainly not its pronunciation). The same can not be said, however, of *Jamón ibérico*, and its variants. (Dubroca Galin, 2005)

Another ham variety, *jamón de cerdo cruzado*, made from cross-bred animals (hybrid ancestry of Iberian and white pig) will inevitably cause confusion across the Pyrenees, unless an easily understandable equivalent is found that is easy to place and spot within the existing terminology.

Iberian ham having become a landmark Spanish product, it has sprouted numerous subdivisions, each corresponding to certain production zones between which a fierce economic war is waged. The resulting terms include: *Jamón de Jabugo* for Iberian ham from Cadiz, *Jamón ibérico de Extramadura*, *Jamón ibérico de Teruel* (in Aragon) and *Jamón ibérico de Guijuelo* (Salamanca province). For each term, the equivalent in other languages calls for certain adaptations which, sadly, not all producing groups are prepared to make.

Similarly, provincial France went back to its ancestral roots and dug up a native pig, a so-called cousin of the Iberian pig that used to live in the Bigorre. It is called *porc noir*, like its Portuguese relative (*porco preto*). How well the meat will sell, remains to be seen. Will the media take to it, and what about consumers? These last two/three years, Iberian pigs in Spain, whose meat may be eaten fresh

7. Carriers of terminological divulgation in the food industry, gastronomic and cultural sections in the press, the cookery section in women's magazines, commercial manufacturers' websites; information on websites of large supermarket chains as well.

(not dried), have given rise to a series of terms referring to the way in which the animal's chine is cut up (this does not apply to Duroc-Jersey pork). Specific names for specific sections, like *el secreto*, *la pluma*, *la presa*, and *la sorpresa*, much appreciated by aficionados, are beginning to find their way to the broad public⁸. It will be interesting to see what solutions the French adopt.

In the case of pork meat, terminological divulgation is essential to help consumers and manufacturers/producers distinguish between various products. In an export context, the old terminology remains essential.

A detailed look at the information found on Iberian ham eaten in French-speaking countries shows that Jabugo ham (from Cadiz, in southern Spain) seems to be the exception to the rule, it being virtually the only one actually to be specifically named and get rave reviews in gastronomy newspaper supplements. Equally striking is the fact that those countries unanimously retain the term *pata negra* for *jamon ibérico*, despite the fact that this popular name is yet to be officially registered. Also worth mentioning, is the fact that in Spain, the phrase *pata negra* is used to indicate excellence, regardless of context.⁹

Things having changed quite drastically these last few decades, ham terminology must be placed in perspective. This evolution also proves that promotional terminology can be studied both synchronously (terminology as it used today versus that used 50 years ago) and diachronically (evolution of promotional terminology).

4. Conclusion

Promotional terminology is like any other jargon: it is codified and typified, especially when dealing with registered trademarks, whichever the kind. Promotional

8. More information on: <http://www.afuegolento.com/> (under: piezas de ibérico) 'Debemos saber qué piezas son la presa entraña, la pluma, la punta y la cabezada de lomo, el secreto o la cruceta, el falso secreto o engaño, el lagarto, la sorpresa, la careta o pestorejo.

Así, **la pluma** es la parte anterior del lomo y tiene forma triangular, que no tiene nada que ver con la punta de lomo que es el recorte caudal del lomo y se llama también 'el filete del carnicero'. En cuanto a **la cabezada de lomo** se localiza en la zona anterior del lomo. **La presa entraña** es una pieza en **abanico**, con mucha grasa intermuscular y se sitúa en el área cervical entre los músculos serratos. **El secreto** es otra pieza con mucha grasa intermuscular, y está integrada por latísimo del dorso. No tiene nada que ver con **el falso secreto** que se localiza en la musculatura cutánea del área del cuello. **El lagarto** es una pieza estrecha que está conformada por el músculo iliocostal. Y **la sorpresa** está formada por la musculatura temporal'.

9. The following rather strange example illustrates how terminology can infiltrate society: the older Spanish teachers, recruited through highly selective exams and very proud of their job, earned by winning a hard battle 20-odd years ago, actually refer to themselves as '*de pata negra*' (of superior quality...).

terms, however, are extremely varied and do not always match the norms established by producer organizations. In their eagerness to sell, some manufacturers have this knack of coming up with rather fanciful terms.

Such terminology hardly ever finds its way to the dictionary. It can, however, be found on most commercial websites and is earmarked in related pages as 'words to remember', 'mini lexicon' or more simply 'essential words', etc. This terminological information is geared to reassuring consumers, who will feel more confident and buy the product. The Internet offers consumers two uses for such terminology: as a tool for accessing information or to be assimilated and used to find out more about the difference between, say, *bloc de foie gras* and *galantine de foie gras*. In the latter case, consumers have indeed assimilated the terminology and terminological science has accomplished one of its missions, i.e. to protect all parties' rights.

Promotional terminology is extremely important, not just for producers, but also for buyers/consumers. Names and *appellations contrôlées* indeed filter out abusive intrusions from unscrupulous or barely scrupulous producers. Once guidelines become an actual frame of reference and have been imposed on the makers, a product can clearly be defined and explicit terms selected to guarantee limpid and loyal communication.

Such promotional terminology is mainly communicated in writing: it is what we find on price lists, labels, receipts, invoices, etc.; but it is also used in oral communication. Brutally manipulated, current words thus get thrashed about and automatically pushed towards new terminological solutions.

Promotional terminology keeps changing, for the simple reason that consumer products themselves also evolve. Those very changes give rise to specific names that in their turn resurface on various national and international markets via translations.

References

- A fuego lento www.afuegolento.com
- Actes du XXI Colloque d'Albi, (2001) Langages et signification. *Le stéréotype: usages, formes et stratégies*. Saint-Chamas: M.L.M.S. éditeur www.marges.linguistique.free.fr.
- Arnaud, A. (1989). *Les mots vendeurs*, La Maison du dictionnaire, Paris.
- Cabré, M.-T. (1998). *La terminologie. Théorie, méthode et applications*. PUO et Armand Colin (translated from the cat. 1st ed. 1992, p. 65).
- Cabré, M. T. (2002). El conocimiento especializado y sus unidades de representación: diversidad cognitiva. *Sendebarr*, 13, 143–153.
- Giraud, G. (1999). Les produits alimentaires régionaux ont-ils une place au sein de la globalisation? Une approche du marketing pour l'Europe. *Agroalimentaire*, N°8, Juin 1999, 29–35.
- Muchnik, J. (2006). Identidad territorial y calidad de los alimentos: procesos de calificación y competencias de los consumidores, *Agroalimentaria* N° 22, Enero-Junio 2006, 89–98.

- Muchnik, J. (2003). Nourrir le corps humain et le corps social, Colloque: *Le monde peut-il nourrir le monde? Sécuriser l'alimentation de la planète*, INRA, Palais de la découverte, 15 octobre, Paris.
- Dubroca Galin, D. (2005). II Congreso Internacional AIETI, Stéréotypes, clichés et tournures récurrentes: des ressources linguistiques au service de la mercatique? *Formación, investigación y profesión*, 739–749. AIETI (Asociación Ibérica de Estudios de Traducción e Interpretación), Comillas Madrid.
- González del Rey I. (2002). *La phraséologie du français*. Toulouse: Presses Universitaires du Mirail.
- GROSS, G. (1996). *Les expressions figées en français, noms composés et autres locutions*, Paris, Ophrys.
- Lerat P. (1991). Les langues spécialisées. Paris: PUF
- Lerat P. (1995). *Las lenguas especializadas*. Barcelona: Ariel.
- Maingueneau D. (1991). *L'Analyse du Discours*, Paris, Hachette Supérieur.
- Maniez, F. (2001). Extraction d'une phraséologie bilingue en langue de spécialité: corpus parallèles et corpus comparables, *Meta*, XLVI, 3.
- MAPA Spanish Agriculture, Fisheries and Food Ministry www.mapa.es
- Pescheux, M. (2003). Construction de sens et modèle argumentatif de la signification lexicale: une formulation de stéréotypes 'lexicaux'. *El texto como encrucijada*, Actas del Congreso de la APFUE, Universidad de la Rioja, 259–269.
- Schapira, Ch. (1999): *Les stéréotypes en français: proverbes et autres formules*, Paris, Ophrys.

Computer aided term bank creation and standardization

Building standardized term banks through automated term extraction and advanced editing tools

Jody Foo and Magnus Merkel

Using a standardized term bank in both authoring and translation processes can facilitate the use of consistent terminology, which in turn minimizes confusion and frustration from the readers. One of the problems of creating a standardized term bank, is the time and effort required. Recent developments in term extraction techniques based on word alignment can improve extraction of term candidates when parallel texts are available. The aligned units are processed automatically, but a large quantity of term candidates will still have to be processed by a terminologist to select which candidates should be promoted to standardized terms. To minimize the work needed to process the extracted term candidates, we propose a method based on using efficient editing tools, as well as ranking the extracted set of term candidates by quality. This sorted set of term candidates can then be edited, categorized and filtered in a more effective way. In this paper, the process and methods used to arrive at a standardized term bank are presented and discussed.

Keywords: terminology extraction, word alignment, quality assurance, revision, filtering, translation memories, term banks, multilingual terminology

1. Introduction

Quality assurance (QA) of products and services is standard procedure in most industrial areas today. In the area of document production and localization, quality assurance has been deemed to be both time-consuming and costly, as most of the linguistic quality assurance has to be done manually. When discussing QA in documents and their translations, both the original source text as well as the

translation memories used, need to be checked and kept consistent. This is a time-consuming process, and since translation memories also tend to grow very fast, many companies are facing problems with the quality of their translation memories. Some errors in translation and authoring can be detected and corrected by using controlled language checkers for Simplified English together with spell and grammar checkers. However, problems caused by inconsistent terminology, cannot be detected by solely using such tools. Inconsistent terminology is perhaps the most crucial in source language documentation (originals), since terminological mistakes multiply with every language that the original is translated into. Lombard (2006) illustrates this phenomenon by an example where an American software development company refers to a 'closing' or 'stopped application' by inconsistently using terms such as *cancel*, *quit*, *close*, *end* and *stop*, in the graphical user interface and in written documentation. This is not a problem for the developers/source text writers, as they know what those terms mean. Translators and localizers, however, will face problems, as they might be tempted to translate every distinct source term into a separate target term, even though the source terms represent the same concept, thereby reproducing inconsistency across languages and multiplying the total number of inconsistencies in the documentation. Furthermore, end users may also misunderstand the intended message in the documentation. It is clear that poor documentation quality can result not only in dissatisfied clients and end users, but also in substantially increased costs from revisions, re-translations and delays in delivery. Detecting mistakes in documentation before it reaches the end-users is the only way to avoid such extra costs and inconvenience.

One solution to improve the quality of documentation and localization of texts is to employ a standardized term bank. Creating a term bank, however, is incredibly time-consuming if done the traditional way, i.e. by manual analysis. This process usually involves several months, if not years, of corpus analysis to extract terms and decide which terms need to be standardized.

During the last decade, word alignment techniques have been used to create resources that are usable in practice for translation activities. Using such techniques can yield much quicker results than manual analysis. Sentence-aligned source and target texts (such as translation memories) are an excellent repository for extracting terminological solutions, both monolingual and multilingual. Given a proper methodological approach, the processing of translation memories containing source and target texts, will not only uncover an abundance of term candidates in the source and target texts independently, but also correspondence relationships, contexts and evidence for inconsistent usage. However, since word alignment can never produce 100 percent accurate term pairs, methods of how to filter out erroneous entries and efficiently revise the output from alignment systems need to be developed. Even if an alignment system is close to perfect, the data

itself (the source and target texts) will contain errors, omissions and additions that will result in terminological entries that are unwanted in a standardized term bank. The source-target alignment map can be used to reveal how consistent, or inconsistent, term usage is in reality. By using the proper method to find inconsistencies, the latter can be corrected and will thus not penetrate the term bank. If inconsistencies are left unchecked, they will remain in the final term bank. Having a standardized term bank, however, is not enough. Authors and translators must be able to easily access the term bank via the tools they themselves use to write and translate documents.

In this paper we describe a process that starts with a translation memory (or a pair of parallel source and target documents) and results in a standardized term bank.

2. Benefits of standardized term banks

As noted earlier, a standardized term bank is one of the most important resources for quality assurance processes in both source writing and translation activities. In a survey performed by LISA in 2005, terminological inconsistency was considered to be 'the biggest problem' in terminology management (LISA, 2005). A standardized term bank should be seen as a terminological resource containing information, ideally on all relevant concepts of the domains an organization requires, with translations in all active languages. The standardized term bank has a multitude of applications in an organization, some of the most obvious being:

- a. Support for in-house writing
- b. Support resource for suppliers (e.g. translators)
- c. Quality assurance and quality instrument (could regulate quality flaws by e.g. lower prices)
- d. Lower direct prices for translations when term banks are supplied.
- e. Lower costs by higher quality (no or less need for rework)
- f. Increased recycling rate in future versions
- g. Investments for the future (when machine translation systems improve)
- h. Necessary data for Simplified English or other controlled languages.

The advantages are significant, both for economic and quality reasons, but the road to a standardized term bank is not exactly smooth. Arriving there requires skilled terminologists and, usually, a lot of time.

3. Issues

In order to arrive at a standardized term bank within a reasonable cost and time period, it is practical to reuse the information contained in existing translations, such as translation memories. Parallel texts or translation memories, contain as Pierre Isabelle puts it, ‘more solutions to more translation problems than any other available resource’ (Isabelle et al., 1993: 205). Not only is the bilingual aspect of the translation memory essential in identifying specific source and target language information. The parallel aspect of the documentation also comes in extremely handy if the only aim is to create a monolingual term bank. Since the documentation is available in at least two languages in parallel it is possible, at least in theory, to take advantage of the choices made by writers and translators. For example, the information can be used to find sets of terms with semantic similarities, both across languages and within one language.

When designing a standardization process, a number of issues have to be resolved, e.g. extraction methods, term volume, roles and responsibilities, term distribution and access to terms, as well as the maintenance of the term bank. A method of collecting terminological entries must also be decided upon. There is also the issue of how the term bank should be structured, which can range from simple term lists that merely contain expressions, to rich databases with definitions, examples, grammatical information, usage, links to document files, domain membership, etc. In some cases, standardizing the most common and domain critical terms may suffice, in others tens of thousands or even more terms are needed. This choice depends heavily on the intended use of the term bank. In an undertaking as complicated as creating a term bank, it is imperative that roles are clearly defined. For example, who has the authority to take what decision on the status of a terminological entry, such as approving term usage and definitions as well as banning a certain terminological synonym? When the standardization process is complete, should the terminology be made available in the authoring and translation processes? The answers to those questions have an impact on how the term bank should be built.

4. Proposed semi-automatic term bank building process

There are at least two general approaches to extracting bilingual terms. The first approach is to first perform monolingual term identification, i.e. from source and target files separately, and then align the identified terms across the languages. This is an *extract-align* approach. The second approach is to first perform word alignment and then separate the aligned units into terms and non-terms. This is an

align-filter approach. The first *extract-align* approach is used in most commercial systems, such as Trados Term Extract and SDL's PhraseFinder. The second approach, *align-filter* (Ahrenberg, Merkel & Petterstedt, 2003), is the one used in our proposed building process.

Our proposal for arriving at a standardized term bank, as presented in this paper contains the following steps:

1. Grammatical analysis and interactive training of word alignment
2. Full text automatic bilingual alignment
3. Type aggregation and database storage
4. Term candidate detection, filtering, categorization and ranking
5. Revision and editing
6. Multi-user support and export to various formats (e.g. TBX, MultiTerm, Excel)

Our process uses full text alignment (2), meaning that alignment aims at aligning all elements of the parallel text (Merkel, Petterstedt & Ahrenberg, 2003). Before automatic full text alignment can be performed, the system must be trained, so it can adapt to the current parallel corpora. This is done through collaborative machine learning (1) using the ILink application. Training and automatic alignment is performed in cycles, the alignment results evaluated after each cycle, and the cycle repeated until acceptable results are achieved. The parallel texts are also grammatically analyzed and annotated with linguistic information (IFDG). When the automatic alignment (ITrix) is complete, the produced alignment map is studied and term candidates are identified and stored in a database along with statistical and grammatical information (3) using the TBM application. The alignment map contains both single-word units and multi-word units. The term candidates in the database are then filtered to remove bad candidates, categorized and ranked. The final step is to manually edit and revise the remaining term candidates using IView (5) before exporting the standardized term bank to one of the most commonly used file formats (6). The above process is also described in Figure 1.

As seen in Figure 1, translation memories can be used as the initial parallel texts. In fact, translation memories are preferable to working directly with parallel documents, as additional steps, such as document and segment alignment have to be performed in the latter case. Translation memories are already aligned on the translation segment level, which in most cases is equal to sentence level.

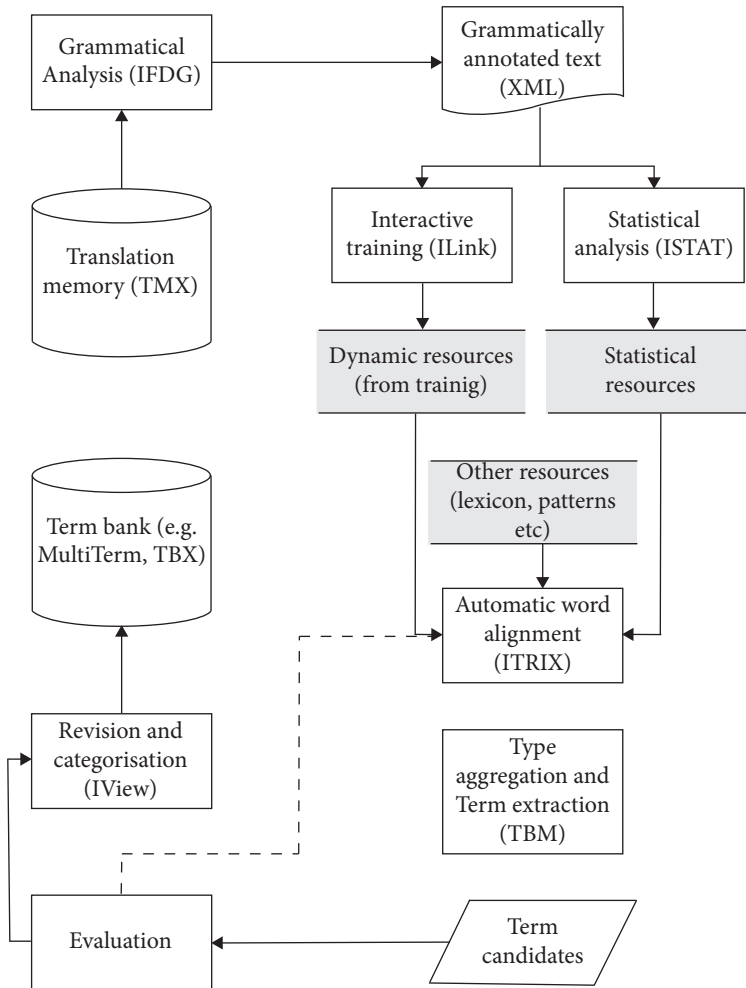


Figure 1. Process of going from translation memory to term bank via word alignment

4.1 The ITools suite

In this project, most of the software used, belongs to the ITools suite developed at Linköping University and at Fodina Language Technology. The most important parts of the ITools suite are the following 6 software components:

IFDG – a front-end to Connexor’s Machine Syntax syntactic parsers, formerly known as Functional Dependency Grammar parsers (Tapanainen & Järvinen, 1997). Currently, Machine Syntax supports 10 European languages and provides data of the following kinds for each word token: (1) Base form (lemma), (2) Part-of-speech and morphological features (3) Syntactic function,

(4) Dependency relation and head of dependency relation. The IFDG tool outputs XML files in a format very close to the XCES format (Eagles, 2002).

IStat – a statistical tool that creates bilingual lexical resources from the source and target text, based on co-occurrence measures.

ILink – an interactive word aligner. This tool is used for training the automatic word aligner. Bilingual resources are being built up incrementally each time words, phrases and terms are aligned in a graphical user interface. A user/annotator is in control of the alignment process and confirms, rejects or modifies the alignments proposed by ILink. The created resources are finally fed into the automatic word aligner.

ITrix – a fully automatic word aligner. The source and target texts are word aligned automatically, using the resources created by ILink and IStat, as well as static resources such as bilingual lexicons and pattern resources.

Termbase Manager (TBM) – a conversion utility that transforms the results from the automatic alignment into an SQL database, containing term candidates, inflectional variants, grammatical information, examples, etc.

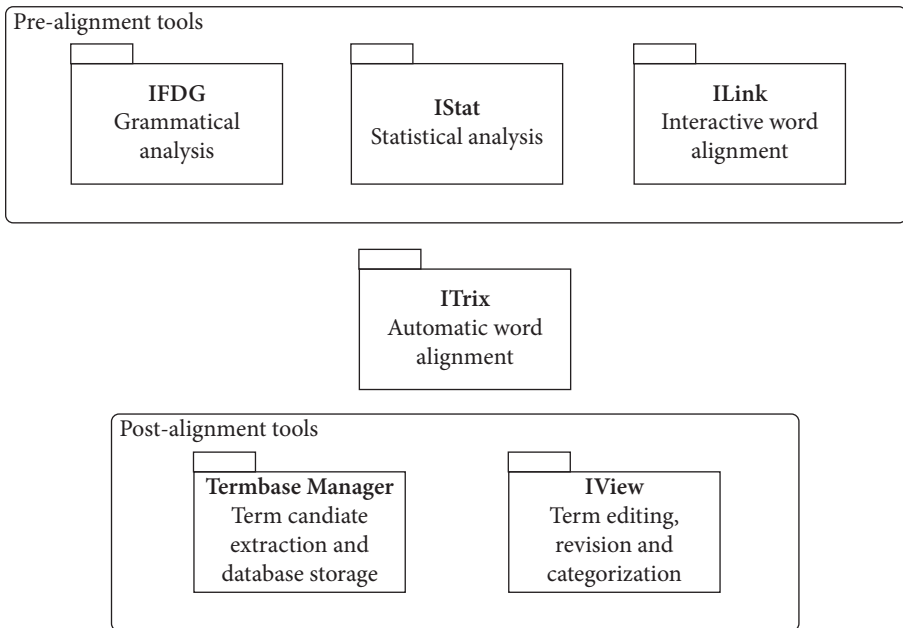


Figure 2. The ITools Suite

IView – a graphical interface to the SQL database. In *IView* the terminologist can filter, search, revise and categorize all term entries and finally export the term bank to an external file format, such as TBX, MultiTerm or OLIF.

The core of the *ITools Suite* is the automatic word aligner *ITrix*, which is preceded by the application of the pre-alignment tools *IFDG*, *IStat* and *ILink*, and succeeded with the post-alignment tools that aid the user in creating the term bank and standardizing it. The *ITools suite* is illustrated in Figure 2.

4.2 Grammatical analysis and interactive training (pre-alignment)

The source and target texts are grammatically analyzed with *IFDG*, resulting in a grammatically annotated text in XML format. After the grammatical analysis, two types of domain-specific resources are created. The first type of domain-specific resource is a set of bilingual dictionaries based on a statistical analysis of the source and target text (using *IStat*). The statistical dictionaries are created both for base forms (lemmas) and word forms (inflections).

The second type of resource is used by the alignment system to learn how to best align the current parallel corpora. Two sets of samples are isolated from the bilingual data. The first set is used as test data to measure the performance of the alignment system. The second set is used to train the alignment system, so that it can adapt to the current corpus. The machine-learning algorithm is based on collaborative training (explained below). The collaborative machine learning takes place in *ILink*, the interactive word aligner (Ahrenberg et al., 2003). The *ILink* application is shown in Figure 3. The annotator is presented with alignment proposals from the software (color coded). Each proposal is then judged manually by the annotator who can choose to accept, reject or revise the proposal by pressing buttons or changing the selections in the *Link* panel window. Every time the user makes a decision, the information contained in that particular alignment is stored as bilingual resources on the base form level (lemmas), word form level, parts-of-speech level and function level. An ‘Accept’ decision results in positive data, and a ‘Reject’ decision in negative data. Additions and deletions in the texts are also annotated in the same fashion and indicated in the interface. Depending on the size of the project, anything from fifty sentence pairs up to over a thousand sentence pairs could be trained in this manner. The quality of *ILink*’s proposals is improved with additional training, resulting in less time needed per interactively aligned sentence. In the beginning, many proposals are erroneous and need revising. Eventually, the annotator will in most cases only need to click the Accept button.

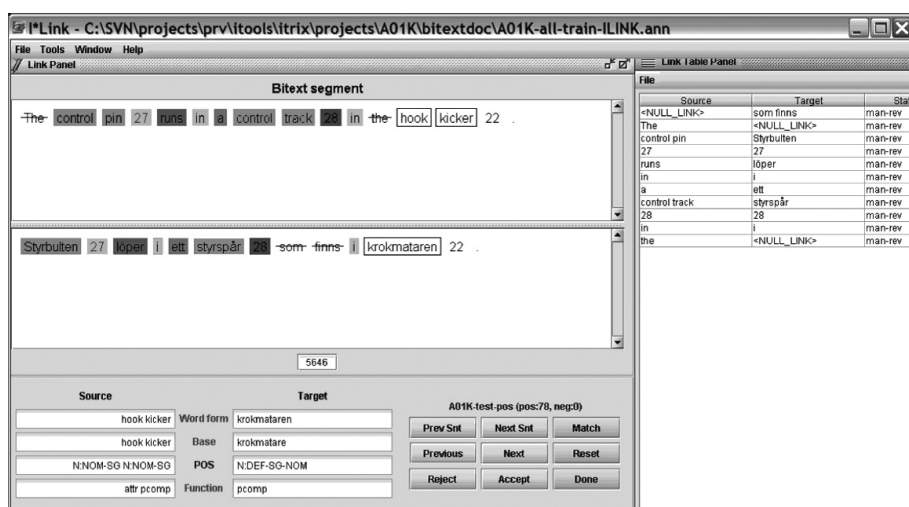


Figure 3. ILink, the interactive word aligner. Buttons enable the user to accept or reject alignment proposals. Alignments are colour coded and shown in a table to the right. Additions and deletions are visualized through strike-through lines

Apart from statistical resources and dynamic resources generated by training in ILink, other resources may be added to the alignment system, such as bilingual dictionaries and parts-of-speech correspondences.

4.3 Automatic word alignment

When resources have been compiled and sufficient training has been done, the automatic alignment system ITrix, is run. ITrix is fully automatic and does not require any intervention from the user, apart from configuring the source and target files and how the resources are to be applied. An example of how ITrix aligns a sentence pair can be seen in Figure 4.

The English-Swedish sentence pair alignment depicted in Figure 4 can also be seen in Table 1.

The example shown in Figure 4 and Table 1 is relatively easy to align, as the structure of the source and target sentences is almost identical. As can be observed, the four noun phrase source terms all consist of two words (*bearing arrangement*, *carriage structure*, *throttle box* and *throttle lever*) and these have all been aligned correctly to single word items in Swedish.

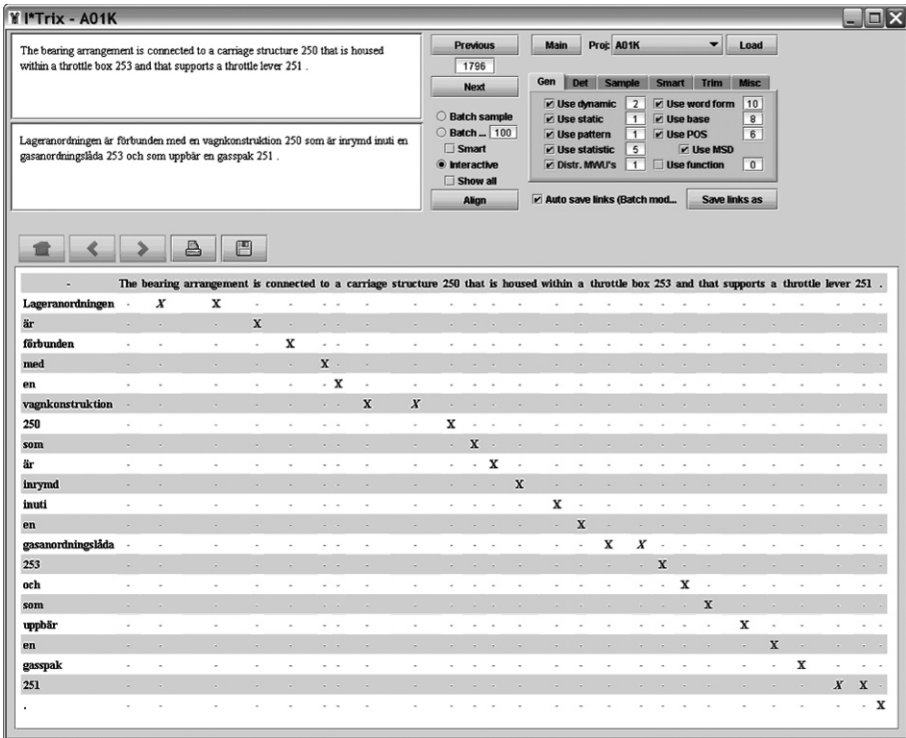


Figure 4. ITrix, the automatic word aligner. A screenshot of automatic word alignment of a sentence pair. The Xs indicate alignments. E.g., *throttle box* in the English sentence has been aligned with *gasanordningslåda* in the Swedish sentence

4.4 Type aggregation and database storage

The output from ITrix is a map (bitext map) of all the orthographic correspondences (tokens) found between the source and target elements in the translation memory. From a lexicographical and terminological viewpoint, however, type alignments are much more interesting than token alignments, i.e. it is more interesting to understand that *bearing arrangement* has been aligned with *lageranordning* rather than study all instances of that kind of alignment. Termbase Manager (TBM) uses the bitext map produced by ITrix, as well as the grammatically annotated documents to aggregate token alignments into type alignments and identify term candidates. During this process, other information such as part-of-speech and term candidate frequencies is also extracted and stored in the database. Since the input to the term candidate selection process is bilingual, term candidates are stored in pairs, i.e. source term candidate and target candidate are connected to

Table 1. Results from automatic word alignment

Source	Target
The	
bearing arrangement	<i>Lageranordningen</i>
is	<i>är</i>
connected	<i>förbunden</i>
to	<i>med</i>
a	<i>en</i>
carriage structure	<i>vagnkonstruktion</i>
250	<i>250</i>
that	<i>som</i>
is	<i>är</i>
housed	<i>inrymd</i>
within	<i>inuti</i>
a	<i>en</i>
throttle box	<i>gasanordningslåda</i>
253	<i>253</i>
and	<i>och</i>
supports	<i>uppbär</i>
a	<i>en</i>
throttle lever	<i>gasspak</i>
251	<i>251</i>
.	.

each other and cannot exist alone. Once the processing is done, the following information is available in the database for each term candidate pair:

- Source base (lemma)
- Target index term (lemma)
- Source term variants (inflectional variants of the source index term)
- Target term variants (inflectional variants of the target index term)
- Examples from the translation memory (term usage in context)
- Parts-of-speech label (for multiword terms the parts-of-speech of the head word is stored)
- Concept label (a preliminary label for a concept, based on the source index term)
- Statistical data, such as frequency of term pair, frequency of source term and target term
- Q-value (explained further on)
- Status, such as *Approved*, *Pending approval*, *Dismissed*, etc.

4.5 Term candidate detection, filtering, categorization and ranking

As mentioned earlier, the term extraction approach taken is to first align the full text and then filter the alignment map to remove non-terms. The first post-extraction step is to perform initial filtering to exclude function words and punctuation. In the alignment example in Table 1, this initial filtering would include the four noun phrases, and exclude the rest of the alignments. After the initial filtering, the database contains term candidates. Some terminologists may well like to keep the past participle verbs, like *connected* and *housed*, as domain-specific terms, as well as the simple present verb *support*. The goal of the secondary filtering strategies is to remove term candidates that are considered to belong to the *general language* domain, rather than specialized domain, i.e. the goal is to remove common words that frequently appear in non-domain-specific texts. Resources used during the alignment process, such as dictionaries, are re-examined in this filtering process. Term candidates that match data in these resources are categorized as general language (for example, core dictionary items or resources that have previously been judged as general language in IView). After the secondary filtering, the remaining term candidates are processed by humans.

This is a rather crude approach, but one that seems to work in practice. Any domain-specific resource (say a medical dictionary applied in a medical domain), can override a general language resource, e.g. allowing *hand* to be a medical term in a patient journal, whereas in most other domains it would be categorized as general language.

The filtering process can also be based on term pair frequency. In some applications only high frequency entries are of interest. Human effort used to process the remaining term candidates is guided by ranking the term candidates using a quality ranking measure, *Q-value*. The *Q-value* metric orders the term pairs in such a way that high-quality alignments are at one end of the scale and low-quality alignments at the other end (Merkel & Foo, 2007). In a previous study (Merkel & Foo, 2007), it was shown that the *Q-value* outperforms pure frequency and the Dice coefficient as a predictor of the quality of alignments. For example, in an evaluation of an alignment project for patent information texts in the Animal Care domain, it was shown that different *Q-value* thresholds would yield predictable precision and recall changes that would save time and effort. By ranking the term pairs according to the *Q-value* measure, the following volume and precision figures were acquired (see Table 2).

Table 2. Estimated precision and number of generated term pairs when ranked by quality estimate (Merkel & Foo, 2007)

Precision %	Q-value	Volume
95.8	0.53	10 400
91	0.50	20 040
~80	0.20	~30 000

The implication of the evaluation is that selecting a specific set of term candidates may be based on Q-value thresholds in order to arrive at a new set of term candidates with a desired quality (precision).

Note that the quality measure in a way only measures the accuracy of the alignments, not the importance of the term or term pair, i.e. whether it should be included in the term bank as a standardized entry. High alignment quality, however, is a prerequisite for a term candidate pair to be promoted to a validated term pair.

One important step when transforming a candidate term bank into a standardized term bank is to detect terminological inconsistencies and correct them. Inconsistent term usage is easy to detect using IView. One detection method is to sort the term candidate pairs by alphabetical source order. Doing so will reveal translation variants, as groups of similarly spelled source terms will be displayed next to their translations. In the above mentioned patent information text, the English term *pasture area* was found to have three different Swedish translations, *betesfält*, *betesfältsområde* and *betesfältsstycke*. Similarly, if the term candidate pairs are sorted by alphabetical target order, inconsistent source term usage can be detected. In the same domain, the Swedish term *sjukdom* was found to have four English translations: *disease*, *disorder*, *illness* and *sickness*. Using SQL queries, it is possible to collect all instances of translation inconsistency. This collection can then be processed in a focused effort by the terminologist.

4.6 Revision and editing

The term candidate pairs stored in the database can now be processed using IView, where the candidates are filtered, revised and categorized to form a standardized term bank. Figure 5 shows a screenshot where the term pair *engagement portion* – *ingreppsparti* is selected. In this view, the user can change the status of the term candidate pair, edit base forms as well as grammatical information. The term candidate pairs can be sorted by any of the available columns. When the user selects a term pair, information on inflectional variants is shown in the centre panel, and

the most common term contexts from the parallel text are displayed in the bottom panel of the application window. The selected terms are displayed in bold face.

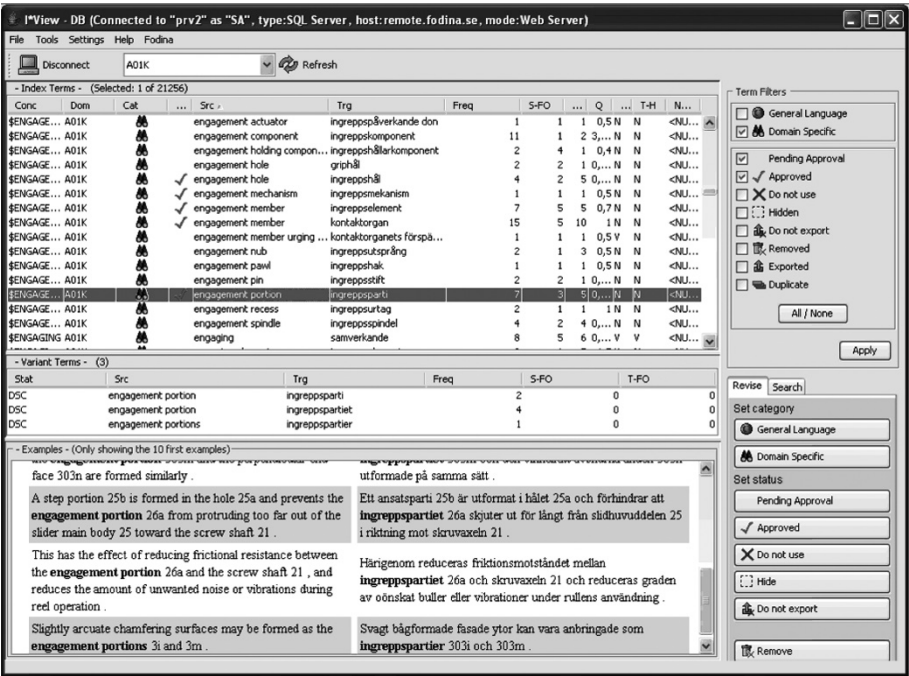


Figure. 5. The IView application. The top panel displays base forms of the source and target terms along with related information. Variant terms, can be viewed in the centre panel. The bottom panel contains examples from the source and target texts

All visible fields in the top panel – such as source/target term, term parts-of-speech, etc. – are editable directly in the main interface. Errors will always be found in the initial term candidates stored in the database, as they are the result of an automatic process. Systematic errors can be corrected automatically, while some exceptions can only be corrected by a human. Examples of machine errors resulting from the automatic process are e.g. incomplete, erroneous or omitted part-of-speech data and mistakes in lemmatization. Adding term definitions, notes and editing other data fields, is done by opening the *Edit dialogue*, see Figure 6.

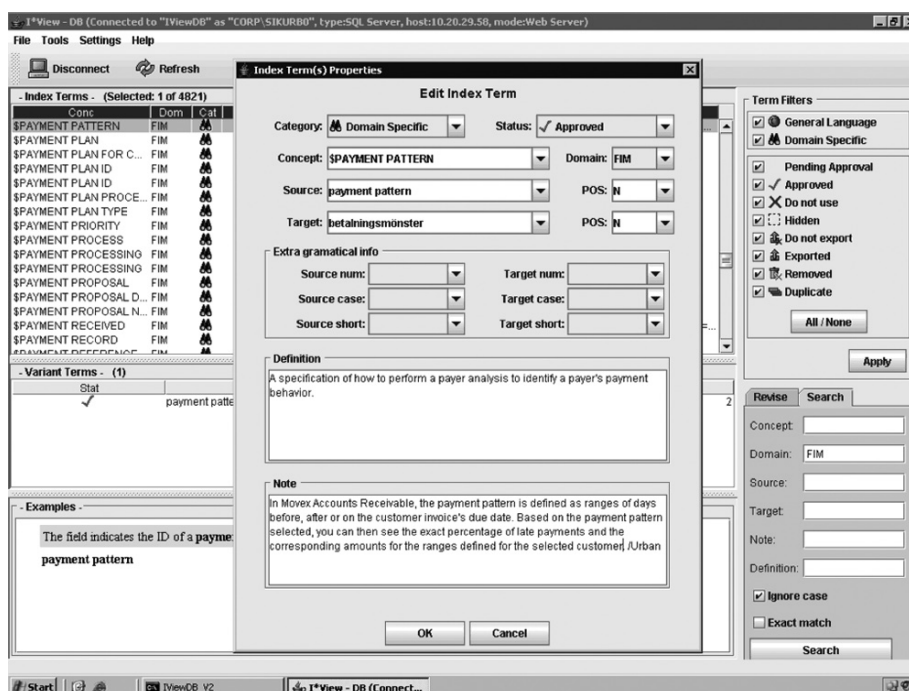


Figure 6. Edit Index Term dialogue box provides access to all data fields related to the selected term candidate pair such as notes and definition

5. Future work

The inconsistency detection could be improved by shifting the focus from terms to concepts, but then better ways of forming concept clusters would have to be applied, cf. Priss and Old (2005). By using a semantic mirroring strategy (Dyvik, 2002), sets of semantically related terms that form concept-like sets can be constructed, thereby making it possible to identify more inconsistencies. Semantic mirroring utilizes the fact that the lexicons in different languages are structured in different ways. Dyvik's technique has made it possible to cluster terms and words in concept-like groups and, to a certain degree, determine the semantic relation between concepts (such as synonymy, hyponymy). The semantic mirroring techniques can be applied to bilingual lexicons or term banks and are likely to be a promising method to detect inconsistencies. In preliminary experiments, where semantic mirroring is used on aligned term candidates using Q-value as a filter, we have found that terms belonging to the same concept can be grouped with promising results.

6. Conclusion

Standardizing term banks is a step towards better-quality source documents and translations. A standardized term bank can be used not only as a resource when creating texts and translations but also as a resource for quality assurance. One way of arriving at a standardized term bank is to apply word alignment techniques on parallel texts, and thereby detect the major obstacle to quality in the documentation world, namely terminological inconsistencies. In this paper, we have presented a process where a suite of alignment and alignment related tools are used, not only to detect terminological inconsistencies, but also to create a standardized term bank.

The proposed process has now successfully been applied (and is currently being applied) in several projects with both industry partners and government bodies. In one project together with the Swedish Patent Office (PRV) and the European Patent Office, the task was to process around 100 000 patent texts available in English and Swedish, and produce a term bank with more than 100 000 English-Swedish entries. This term bank is to be used for a machine translation application. During this project, term candidates are processed using the ITools suite and validated to determine if they belong to a specific patent class by domain experts, after which they are validated linguistically by language experts. The domain experts have the task of deciding whether a term candidate pair belongs to the domain in question and the language expert validates the associated linguistic information with the term entry (such as base form, parts of speech and inflection pattern). The overall experience is that the process works extremely well and that a domain expert can process between 4 000 and 6 000 term candidates per working day (8 hours). The linguistic validation is slightly more time consuming as there are more data to consider (a language expert can process between 2 000 and 3 000 term candidates per day).

Acknowledgements

This work has been partly made possible through co-operation with the Santa Anna IT Research Institute and Fodina Language Technology.

References

- Ahrenberg, L., Merkel, M., and Petterstedt, M. (2003). Interactive word alignment for language engineering. The 10th Conference of the European Chapter of the Association for Computational Linguistics April 12–17, 2003 Agro Hotel, Budapest, Hungary (EACL-2003). *Conference companion*, 49–52.
- Dyvik, H. (2002). Translations as semantic mirrors: from parallel corpora to wordnet. In Karin Aijmer and Bengt Altenberg (eds.), *Language and Computers, Advances in Corpus Linguistics. Papers from the 23rd International Conference on English Language Research on Computerized Corpora (ICAME 23)* (311–326). Göteborg 22–26 May 2002.
- Eagles (2002). *XCES Corpus Encoding Standard for XML*. Vassar College (<http://www.xces.org/>).
- Isabelle, P., Dymetman, M., Foster, G., Jutrac, J.-M., Macklovitch, E., Perrault, F., et al. (1993). Translation Analysis and Translation Automation. In *Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI'93)* (201–217). Kyoto.
- LISA. (2005). *LISA Terminology Survey for the Localization Industry*. <http://www.lisa.org/sigs/terminology/termsurvey2001preresults.html>. Localization Industry Standards Association.
- Lombard, R. (2006). A practical case for managing source-language terminology. In K. J. Dunne (ed.), *Perspectives on Localization* (155–171). Amsterdam: John Benjamins Publishing Company.
- Merkel, M., and Foo, J. (2007). Terminology extraction and term ranking for standardizing term banks. In Joakim Nivre, Heiki-Jaan Kaalep, Kadri Muischnek and Mare Koit (eds.), *Proceedings of the 16th Nordic Conference of Computational Linguistics NODALIDA-2007* (349–354). University of Tartu, Tartu, 2007. ISBN 978–9985–4–0513–0 (online) ISBN 978–9985–4–0514–7 (CD-ROM)
- Merkel, M., Petterstedt, M., and Ahrenberg, L. (2003). Interactive Word Alignment for Corpus Linguistics. In *Proceedings from the International Conference of Corpus Linguistics* (533–542). Lancaster.
- Priss, U., and Old, J. L. (2005). Conceptual Exploration of Semantic Mirrors. In *Lecture Notes in Computer Science* (21–32). Berlin/Heidelberg: Springer.
- Tapanainen, P., and Järvinen, T. (1997). A non-projective dependency parser. In *Proceedings of the 5th Conference on Applied Natural Language Processing (ANLP'97)* (64–71).

Competency-based job descriptions and termontography

The case of terminological variation

Koen Kerremans, Peter De Baer and Rita Temmerman

In this article we reflect on the problem of terminological variation in the context of electronic Human Resource Management (e-HRM). In particular, the article takes a closer look at central notions in the HRM policies of companies, such as the competency and job profile, and reflects on the problems that may arise when companies wish to automatically exchange competency-based job profiles. Those problems result from differences in the use and understanding of competency-related terminology as well as occupational terms. The article aims to illustrate with examples from the PoCeHRMOM project how we tried to overcome those problems by using our own terminology management practices and tools.

Keywords: competency management, e-HRM, semantic web, terminological variation, Termontography

1. Introduction

Competency management has become a crucial activity in the HRM departments of today's large companies. It allows them to maintain an overview of the individual competencies (i.e. the set of knowledge, skills and attitudes) of their employees and to evaluate the relevance of these competencies, when linked to the general aims, strategic choices and expertise of the companies.

According to Baisier (2002) competency management leads to better job planning, objective criteria for evaluating personnel, a higher interest in learning opportunities, an improvement of selection procedures, a better acknowledgment of acquired competencies, more insight in the structure of the organization, etc. Baisier (2002) therefore argues that competencies are the main vocabulary of companies. It helps them to communicate e.g. their expertise (knowledge and skills). The use of this vocabulary in combination with new, emerging technologies leads

to many innovative applications in the field of electronic Human Resource Management (e-HRM).

Identifying the individual competencies (i.e. knowledge, skills and attitudes of employees) and core competencies (i.e. the things a company is experienced in) in an organization and setting up competency lists or competency-based job profiles, requires a lot of work and effort. This is, among others, why small medium-sized enterprises (SMEs) are often discouraged to implement competency management and why they make less use of existing innovative e-HRM applications, such as competency-based matching between job offers and job candidates' CVs.

Financed by the Flemish government in the IWT-TETRA funding programme¹, the PoCeHRMOM project² aimed at developing a large knowledge resource of competency-based job profiles in three languages (Dutch, English and French). Companies can access the knowledge resource through a web interface (the 'Profile Compiler'), which allows them to view the occupation profiles, to customize the profiles according to their own needs and, finally, to export the customized versions to standardized formats (such as HR-XML³). The objective of the project was to stimulate SMEs in adopting competency management by helping them to set up competency-based occupation profiles so that they are eventually able to benefit from the existing and new, emerging applications in the field of e-HRM.

Apart from the general objective, this project also addressed a much more fundamental issue with respect to multilingual terminology management, in particular the question of how to deal with terminological variation when structuring competency-related and job-related terminology. For instance, a given occupation may be denoted by different terms while the same job title usually implies a different job content (e.g. different tasks), depending on the company where the job title is used.

Since the PoCeHRMOM project also examined how customized competency-based job profiles (cf. *supra*) could eventually be exchanged automatically in a semantic web environment, it was important to take into consideration the terminological variation associated with competencies and job titles.

In the first part of this article, we applied the 'Semantic Web' to the context of e-HRM. We will briefly explain what the Semantic Web is and in what way it is related to our research in the PoCeHRMOM project (cf. Section 1). The second part focuses on the problem of terminological variation. We summarize how terminological variation has been approached in previous terminology research, how it has been studied and defined and how it is perceived in the scope of the present study (cf. Section 2). In the third part, we will illustrate with concrete examples how we

1. <http://www.iwt.be/steun/steunpro/tetra/index.html>

2. <http://cvc.ehb.be/PoCeHRMOM/Frameset.htm>

3. <http://www.hr-xml.org/hr-xml/wms/hr-xml-1-org/index.php?language=2>

attempted to overcome the problems of terminological variation in the PoCeHRMOM project by using our own terminology management practices and tools. In particular, we will outline the Termontography methodology and describe the Multilingual Categorization Framework Editor (MCFE), a tool for creating and visualizing domain knowledge as well as for structuring terminology (cf. Section 3).

2. Semantic web technology in PoCeHRMOM

In Section 1.1, we will discuss in more detail the notion of ‘Semantic Web’ and the way ontologies are related to it. Next, in Section 1.2, we will reflect on a (semantic web) use case in the PoCeHRMOM project. By describing some possible scenarios for this use case, we will also address the problems mentioned in the previous section with regards to automatic information exchange.

2.1 The semantic web and ontologies

The Semantic Web was originally defined by Tim-Berners Lee (2001) as an evolving extension of the current web in which web content is given well-defined meaning so that it can be understood, interpreted and used by software agents. The benefit for different kinds of web applications is evident. Information retrieval (IR) systems, for instance, will have it easier deciding which search results are relevant and which are not. Whether a user was, for instance, looking for ‘books about Jane Austen’ and not ‘books written by Jane Austen’, is something that an IR system will understand better if the search query ‘Jane Austin’ is marked in certain web pages as for instance ‘Author’, while in others as for instance ‘Topic’.

Realizing a web of semantics is only possible if web developers use a language far more expressive than HTML that allows them to describe the content of web pages. In the context of the Semantic Web, where software agents not only use the semantic information but also share it with one another, the standardization of semantic (meta-)data or tags (such as ‘Author’ or ‘Topic’) is believed to be an advantage. From this perspective, the Semantic Web is therefore also perceived as a common reference framework that allows semantic (meta-)data to be shared and (re)used.

The meaning of semantic metadata can be specified in ontologies. In the domain of computer science, an ontology is defined as a formal and shareable knowledge repository in which individuals (instances), classes (concepts), attributes and relations are made explicit for computer processing (Daconta et al., 2003). The need to establish a common understanding and reference framework of competencies in the PoCeHRMOM project, is met through the development of an ontology.

2.2 Automatic exchange of competency-related information

Tim-Berners Lee's idea of a Semantic Web in which content is annotated in such a way that it can be understood and communicated by information systems may be further exemplified by two case scenarios from the PoCeHRMOM project. Whenever a company plans to hire new personnel, it will probably first write out a job offer. A job offer typically consists of a description of the tasks associated with the given job position inside the company. Apart from that, it usually also contains a list of competencies which a person should have in order to qualify for the job.

In the first case scenario, writing out a job offer is done in collaboration with an intermediary, such as a recruitment office. A recruitment office will first try to identify the competencies considered relevant for a particular job in the given company. From this analysis, it will derive a job profile which is used to evaluate job candidates during interviews. The job profiles may to some extent be 'pre-defined' in the sense that e.g. all programmers need to have knowledge of programming languages. In a later phase, these stereotypical job profiles obviously need to be further customized according to the needs of each company. The first phase in this process, i.e. the matching between a job title used in a company and a stereotypical job profile set up by a recruitment office, may be automated if the matching application 'knows' that the two more or less address the same concept.

The second case scenario is that, instead of working with a recruitment office, a company may also decide to directly publish a job offer on a job site. In this case, the company will have to make certain decisions of its own, such as e.g. trying to figure out into which category of the job classification the job offer belongs. This activity may also be automated if a system 'knows' to which category the job offer corresponds.

In the two cases illustrated above, automatic matching would benefit from first establishing a common understanding about jobs and competencies. This is best realized if everyone (companies, recruitment offices, job sites, job candidates, etc.) agrees to use the same vocabulary and concepts. In reality, however, this will probably never be the case. A closer look at job offers on a job site, allows us to observe that a wide variety of terms are used to denote similar occupations (e.g. programmer, software engineer, software developer or application developer). This variation may be explained by the fact that job titles have become less important in determining rank and pay than skills and experience (Evans, 2004). Another reason is that companies may prefer to use their own terminology to denote specific occupations.

Despite the problems of variation, the PoCeHRMOM project tries to provide information systems with a common understanding of occupations and competencies by developing an ontology and terminological resource. Instead of

approaching the problem from a prescriptive point of view, we have worked out a method by which we try to establish a common reference framework while at the same time taking into account the many terminological variants denoting jobs and competencies. This will be further explained in Section 3. In the next section, we will summarize how terminological variation has been approached in previous terminology research, how it has been studied and defined and how it is perceived within the scope of the present study.

3. Terminological variation

Variation being an inherent property of every natural language, occurring at all levels of linguistic structure (e.g. phonology, morphology, syntax, semantics), it could be considered an obvious linguistic phenomenon to be studied in descriptive linguistics. Despite the fact that domain-specific languages are natural languages, this notion of variation was for a long time obscured in research of domain-specific languages (Lerat, 1995). The general neglect by traditional terminographers concerning the study of variation is explained by the fact that studies in domain-specific languages and terminology did not originate from descriptive linguistics but grew out of an urge for normalization in which a term was defined as the normalized, non-variable denomination of a concept. For a more elaborate discussion, see e.g. Cabré 1999, Temmerman 2000 or Desmet 2005. Since one of the main functions of terms is to facilitate specialized communication and knowledge transfer, traditional terminology theory stated that terms are fixed items that should not be prone to variation (Bowker & Hawkins, 2006).

In more recent years, this principle of univocity (only one term should be assigned to a concept and vice versa) has been questioned, among others, in socioterminological (Gaudin, 1993), sociocognitive (Temmerman, 1997, 2000) and communicative (Cabré, 1999) approaches in terminography. The critique has led to a radical shift, both theoretically and methodologically, in the study of terminology (Bertels, 2005). As a result, the notion of variation has taken up a more central position in descriptive terminology research and is now studied at different linguistic levels of specialized language. In Section 2.1, we refer to previous related work pertaining to terminological variation. Next, in Section 2.2, we discuss how terminological variation is perceived in the present study.

3.1 Previous work on terminological variation

Dubuc (1997), for instance, focused on morphological variation (e.g. *terminology unit* vs. *terminological unit*), orthographic variation (e.g. *focussed interview* vs.

focused interview), elliptic forms (e.g. *pre-paid telephone card* vs. *phone card*) and abbreviations (e.g. *compact disc* vs. *CD*). Daille et al. (1996) have studied the phenomenon of permutation (e.g. *blood flow* vs. *flow of blood*), while Bowker and Hawkins (2006) studied combining forms (e.g. *abdominothoracic* vs. *thoracoabdominal*) in a corpus of medical texts.

A number of previous studies have approached terminological variation from a monolingual perspective (e.g. Lauriston, 1996; Freixa, 2002; Bertels, 2005), thereby studying synonymy and semantic ambiguity of terms within a given language.

Studies of terminological variation have also been concerned with the possible reasons for variation. Bowker and Hawkins (2006), for instance, have revealed conceptual (e.g. patterns of cause and effect), linguistic (e.g. collocations, shortened forms) and social motivations (e.g. social conventions) behind term choice in the medical domain. This is further elaborated in Freixa (2006), who identified six main categories according to which causes of variation can be classified: (a) Preliminary (caused by the characteristics and behaviour of language); (b) Dialectical (caused by different origins of authors); (c) Functional (caused by different communicative registers); (d) Discursive (caused by different stylistic and expressive needs of authors); (e) Inter-linguistic (caused by contact between languages); and (f) Cognitive (caused by different conceptualizations and motivations).

Social motivations, as outlined in Bowker and Hawkins (2006), seem to arise because a concept can be seen from different perspectives. In terminology literature, this is referred to as multidimensionality (Cabr , 1999; Bowker, 1997; Rogers, 1997). Bowker and Hawkins (2006) note that in order to study those factors, one needs a much more detailed and rigorously designed corpus or other approaches to data gathering (e.g. interviewing subject experts).

3.2 How terminological variation is defined in the present study

It should be noted that in the scope of this article, terminological variation pertains on the one hand to lexical variation (i.e. the fact that different terms refer to a similar category or unit of understanding). On the other hand, it also pertains to semantic variation, both intra-linguistically (i.e. meaning differences between terms and synonyms) as well as inter-linguistically (i.e. meaning differences between terms and translation equivalents), meaning differences between terms, synonyms and translation equivalents will need to be represented in (multilingual) terminological resources.

4. Dealing with terminological variation in PoCeHRMOM

As mentioned in Section 1.2, automatic information sharing of HRM-related knowledge will be best achieved if everyone uses the same terminology. Obviously, even in monolingual settings, this is never the case.

In the Termontography approach, we try to establish a common reference framework while at the same time taking into account the many existing terminological variants denoting jobs and competencies. In Section 3.1, we will describe our current method and software tool for setting up terminological resources in which terms are linked to a categorization framework. In Section 3.2, we will show by means of some examples how terminological variation, as defined in Section 2.2, is dealt with in this project.

4.1 Termontography: Method and tool

The *Centrum voor Vaktaal en Communicatie* (CVC) has worked out a method, called Termontography, for developing (multilingual) terminological databases. Termontography combines theories and methods of the sociocognitive terminological analysis (Temmerman, 2000) with methods in ontology engineering (Sure & 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 & 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 inter-categorical relations are therefore structured in a knowledge base or 'categorization framework'. Three types of inter-categorical relations can be discerned: generic-specific, whole-to-part and associative. The generic-specific and whole-to-part relations may be used to group related information and to identify and retrieve this information by means of an abstract group name. Associative relations specify a relation between two categories other than generic-specific or whole-to-part.

In the Termontography approach, the categorization framework supports the information gathering phase during which a corpus is developed (Kerremans et al., 2003). Moreover, 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 pre-defined knowledge also affects the terminographers' working method as well as the software tools that will be used to support that working method (Aussenac-Gilles et al., 2002).

The categorization framework in the PoCeHRMOM project is set up by means of the Multilingual Categorization Framework Editor (MCFE) and will eventually be translated into a formalized representation of domain knowledge (i.e. a domain ontology).

Figure 1 shows a screenshot of part of the PoCeHRMOM termontological resource in the MCFE software. This figure gives a general idea of the information types a user can add in order to create a categorization framework. Left on the screenshot, several meta-categories can be identified, such as 'Competency', 'Hierarchy', 'Occupation' and 'Resource'. As the meta-category 'Occupation' has been selected, the pane in the middle shows all categories that have been classified according to this meta-category. By selecting one of those categories, for instance 'travel agent', users can start specifying relations to other categories in the framework, shown on the right. The example shows, for instance, that through an associative relation 'has domain competency', 'travel agent' is linked to the 'competency 'geography''. The definition of this category, taken from O*NET⁴, has been added in the 'term description' window on the right.

Apart from adding relations, the MCFE tool also allows users to link terms in several languages to the categorization framework. Below the specification field of 'Category relations' (right hand side of the figure) two terms have been specified in the 'Category terms' field: i.e. the Dutch term '*reisagent*' and the English term '*travel agent*'.

In the next section, we will show how terminological variants are dealt with in our approach by focussing on specific parts of the MCFE software.

4. O*NET is a large American resource of occupational information. The full resource can be consulted online at <http://www.onetcenter.org/>.

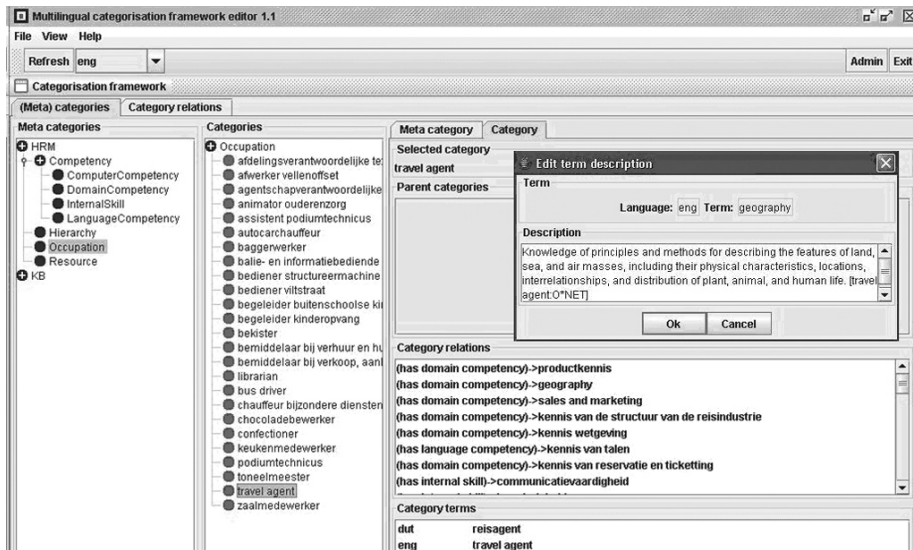


Figure 1. The MCFE software

4.2 Examples from PoCeHRMOM

Figure 1 shows how lexical variation is dealt with in the MCFE software. Translation equivalents such as '*reisagent*' and '*travel agent*' are linked to the same category '*travel agent*' in the framework. Theoretically speaking, any element in the categorization framework (a meta-category, a category or a relation) can be denoted by more than one term in multiple languages. Terms, synonyms and translation equivalents are indirectly linked to one another by means of the elements in the framework.

Although terms linked to the same category share the same meaning, some meaning differences may still need to be made explicit in some way or another. For instance, despite the fact that, for instance, the Dutch terms '*kennis van rekenen*' and '*rekenkunde*' are linked to the same category '*arithmetic*', the meaning of those terms is dependent on the occupation to which they are related.

In the current version of the MCFE software, the meaning of each term is further specified in the 'Term description' window. An example is shown in Figure 2.

This figure shows three windows. The bottom window, left of the figure, lists three Dutch terms linked to the category '*arithmetic*': '*kennis van rekenen*', '*rekenennis*' and '*rekenkunde*'. All three terms can be linked to a general definition: 'the part of mathematics that involves the adding and multiplying, etc. of numbers'⁵.

5. This definition was taken from the Cambridge Advanced Learner's Dictionary: <http://dictionary.cambridge.org/define.asp?key=3952&dict=CALD>

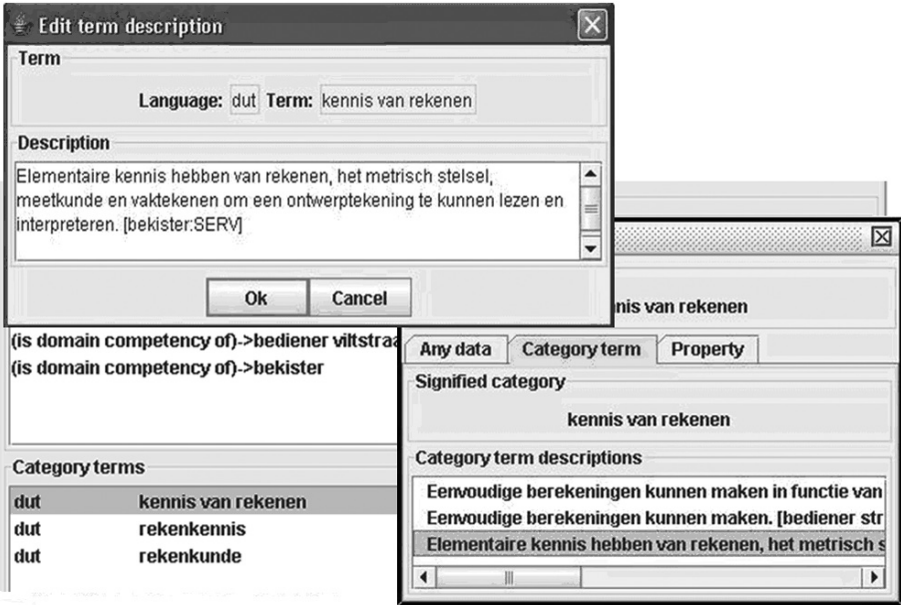


Figure 2. Description of the term ‘*kennis van rekenen*’

However, as those terms happen to be related to different occupations (e.g. chocolate confectioner, carpenter or bus driver), the description of each term is dependent on the occupation to which it applies. In the case of a chocolate confectioner, the context-dependent description or definition of the term ‘*kennis van rekenen*’ applies to purchase prices (of ingredients), calculation of profits and costs, etc. For a bus driver, the context-dependent description is more related to fuel prices, speed, travel distances, fuel consumption, etc. Finally, in the case of a carpenter, the context-dependent description of the term mentions issues such as geometry, interpretation of technical maps, etc.

By clicking on the term ‘*kennis van rekenen*’, a new window pops up on the right, showing the three context-dependent definitions in which this term is used. As is shown in Figure 2, a definition may be read or edited simply by selecting it in the ‘Edit term description’ window.

The reason for adding context-dependent definitions to the termontological resource is that it is easier for companies to judge on the basis of such definitions whether or not a competency is relevant and should be added to a specific job profile. Moreover, companies will find it easier to see how a competency is related to a specific job. This would be more difficult on the basis of a more general description, such as the one for ‘*arithmetic*’ mentioned in the previous paragraph.

Nevertheless, general descriptions are necessary as they contribute to a better understanding of what terms should be linked to a similar category.

Summarizing this section, we conclude that the MCFE tool allows us to structure terms, synonyms and translation equivalents according to a categorization framework. Despite the fact that e.g. terms are linked to the same category, users of the tool are able to represent possible meaning differences in natural language. Those meaning differences are a result of the different contexts in which terms are used.

5. Conclusion

In this article, we have presented results of a CVC project that revolves around developing a termontological resource of competency-based occupations. The resource is to be linked to a web application that will allow companies to create their own competency-based job profiles. The profiles are linked to a common reference framework so that it becomes possible for different types of e-HRM applications to share HR-related information between one another in a semantic web environment.

As companies need to understand how competencies may apply to the occupation for which a profile is being developed, we have decided to make a clear distinction between a general description and a context-dependent description. A general description is linked to a category, meta-category or relation in the categorization framework and helps users to organize terms, synonyms and translation equivalents. The context-dependent definitions will help companies to better understand how certain competencies may apply to a given occupation.

Based on a number of examples, we have shown how terminological variation (as defined in this article) is dealt with and represented in the MCFE tool, used for developing categorization frameworks and structuring terminology. In the current version of this tool, however, the way to specify possible meaning differences between related terms is still rather limited. Further research is required to examine how those meaning differences can be represented in a much more structured and systematic way, as opposed to text fields.

Acknowledgements

This research is financed by the Flemish government within the framework of IWT-TETRA (<http://www.iwt.be/tetra>).

References

- Aussenac-Gilles, N., A. Condamines and S. Szulman (2002). Prise en compte de l'application dans la constitution de produits terminologiques. In *Actes des 2e Assises Nationales du GDR I3*. 289–302. Toulouse: Cepaduès Editions.
- Baisier, L. (2002). *Competentiebeheer als instrument van personeelsbeleid. Een verkenning in de industrie*. Brussel: SERV.
- Berners-Lee, T., Hendler, J. and Lassila, O. (2001). The Semantic Web – Computers navigating tomorrow's Web will understand more of what's going on--making it more likely that you'll get what you really want. *Scientific American* 284 (5), 34–43.
- Bertels, A. (2005). Les Spécificités en contexte: comment étudier la polysémie dans un corpus technique? In D. Blampain, P. Thoiron and M. Van Campenhoudt (eds.), *Mots, Termes et Contextes. Actes des septièmes Journées scientifiques du réseau de chercheurs Lexicologie Terminologie Traduction* (371–380). Paris: Éditions des archives contemporaines.
- Bowker, L. and Hawkins, S. (2006). Variation in the organization of medical terms. Exploring some motivations for term choice. *Terminology* 12 (1), 79–110.
- Cabré, M. (1999). *Terminology: Theory, methods and applications*. Amsterdam & Philadelphia: John Benjamins.
- Daconta, M.C., Obrst, L.J., Smith, K.T. (2003). *The Semantic Web. A Guide to the Future of XML, Web Services, and Knowledge Management*. Indiana: Wiley Publishing.
- Daille, B., Habert, B., Jacquemin, C. and Royauté, J. (1996). Empirical observation of term variations and principles for their description. *Terminology* 3 (2), 197–258.
- Desmet, I. (2005). Variabilité et variation en terminologie et langues spécialisées: discours, textes et contextes. In D. Blampain, P. Thoiron and M. Van Campenhoudt (eds.), *Mots, Termes et Contextes. Actes des septièmes Journées scientifiques du réseau de chercheurs Lexicologie Terminologie Traduction* (235–247). Paris: Éditions des archives contemporaines.
- Dubuc, R. (1997). *Terminology: A Practical Approach*. Quebec: Linguattech.
- Evans, N. (2004). The Need for an Analysis Body of Knowledge (ABOK) – Will the Real Analyst Please Stand Up? *Issues in Informing Science & Information Technology*, 1, 313–330.
- Freixa, J. (2002). *La variació terminològica. Barcelona, Institut Universitari de Lingüística Aplicada*. Barcelona: Universitat Pompeu Fabra.
- Freixa, J. (2006). Causes of denominative variation in terminology. A typology proposal. *Terminology* 12 (1), 51–77.
- Gaudin, F. (1993). *Pour une socioterminologie. Des problèmes sémantiques aux pratiques institutionnelles*. Rouen: Publications de l'Université de Rouen.
- Kerremans, K., Temmerman, R. and Tummers, J. (2003). Representing multilingual and culture-specific knowledge in a VAT regulatory ontology: support from the termontography approach. *Lecture Notes in Computer Science* 2889, 662–674.
- Lauriston, A. (1996). *Automatic term recognition: performance of linguistic and statistical learning techniques*. Manchester: UMIST.
- Lerat, P. (1995). *Les Langues spécialisés*. Paris: PUF.
- Meyer, I., D. Skuce, J. Kavanagh and L. Davidson (1997). Integrating Linguistic and Conceptual Analysis in a WWW-based Tool for Terminography. In *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.

- Moreno, A. and Pérez, C. (2001). From Text to Ontology: Extraction and Representation of Conceptual Information. In Vandoeuvre (ed.), *Terminologie et intelligence artificielle. Rencontres No4* (233–242). Nancy: INIST-CNRS.
- Rogers, M. (1997). Synonymy and equivalence in special-language texts. A Case Study in German and English Texts on Genetic Engineering. In A. Trosborg (ed.), *Text Typology and Translation* (217–245). Amsterdam: John Benjamins.
- Sager, J. C. (1990). *A practical course in terminology processing*. Amsterdam: John Benjamins.
- Sure, Y. and Studer, R. (2003). A methodology for Ontology-based Knowledge Management. In Davies, J., Fensel, D. and Van Hamelen, F. (eds.), *Towards the Semantic Web. Ontology-Driven Knowledge Management* (33–46). New York: John Wiley and Sons.
- Temmerman, R. (1997). Questioning the Univocity Ideal. The difference between socio-cognitive Terminology and traditional Terminology. *Hermes* 18, 51–90.
- Temmerman, R. (2000). *Towards New Ways of Terminology Description. The sociocognitive approach*. Amsterdam: John Benjamins.

Proposals to standardize remote sensing terminology in Spanish

M. Lara Sanz Vicente and Joaquín García Palacios

New and rapid advances in remote sensing and the use of English as *lingua franca* in scientific and technological research bring about new concepts and terms in English that are then introduced into other languages through borrowing and translation. This results in an increasing number of English loan-words and stresses the need to standardize the terminology specific to this field. This paper addresses remote sensing terminology in Spanish and provides a methodology based on bilingual comparable corpora, to further terminological research in this subject area and face its standardization in Spanish. The final aim is to improve professional communication in the field and help translators and interpreters in two complementary tasks: text comprehension and text production.

Keywords: comparable corpora, cross-linguistic comparison, loan-word, neologism, standardization, translation

1. Introduction

Technical and scientific vocabulary is witnessing a considerable upsurge of new terminologies brought about by rapid advances in scientific and interdisciplinary research, and high-technology development. One example among those recent and emerging terminologies is that of a domain devoted to Earth observation from space: remote sensing.

English has always played a leading role in remote sensing communications. The new and rapid advances in this field, headed by the U.S., and the use of English as an international language in scientific and technological research bring about new concepts and terms in English that are then introduced into other languages through borrowing and translation.

Because of the predominance of English in scientific publications, many non-English-speaking experts on remote sensing base their work on English references

and tend to write their scientific papers in this language to reach a wider audience. When those experts write in their own languages, such as Spanish, they need to import new concepts which require a designation, and thus, they often also import their original designation that can be more or less integrated into the target language system. This results in an ever-increasing number of English loan-words to expand the lexicon of this scientific language in Spanish.

Besides, research into remote sensing terminology in Spanish is almost non-existent and English-Spanish vocabularies and dictionaries are scarce and do not derive from interdisciplinary studies. They are simply carried out by domain experts and, though not lacking in scientific accuracy, they do not necessarily help improve professional communication and, above all, assist translation.

The influence of English, the scarce number of bilingual lexical resources in Spanish and, particularly, the lack of terminological studies in this domain stress the need to organize and standardize this terminology.

The present study fits inside the above outlined context. It addresses the standardization of this terminology in Spanish by proposing a procedural method based on bilingual comparable corpora. This goal brings along a number of specific objectives such as: describe, systematize and classify remote sensing terms in English and Spanish, carry out cross-linguistic comparisons and suggest some guidelines for dealing with remote sensing neologisms in Spanish. The results will be collected in a bilingual terminological database to improve professional communication in this field and help translators and interpreters in two complementary and key tasks: text comprehension and text production (Fuentes & García Palacios, 2002: 122).

2. An overview on remote sensing terminology standardization

The International Society for Photogrammetry and Remote Sensing (ISPRS), and particularly its Commission on Education and Outreach, has been the driving force behind remote sensing standardization and the appearance of many bilingual and multilingual lexical resources. At the XIII ISPRS Congress held in Helsinki in 1976, this commission, under the encouragement of two French representatives, Henri Bonneval and Serge Paul, fostered the compilation of the *Multilingual Dictionary of Remote Sensing and Photogrammetry* (Rabchevsky, 1984). This dictionary, published by the American section of the ISPRS (ASPRS), included equivalent terms in French, German, Italian and Portuguese. The ISPRS commission has also fostered the creation of working groups among all its national members to add new languages to its multilingual dictionary and revise the equivalent terms. But not all countries have reacted in the same way and efforts on standardization also differ (Sanz Vicente, 2007: 25–28).

France has led the research into this terminology together with Canada. France established the *Commission ministérielle de terminologie de la télédétection aérospatiale* (COMITAS) in 1978, whose results were recorded in several regulations¹ and encouraged many parallel works (Paul et al., 1982, 1991, 1997). Canada has jointly worked with France through the *Office québécois de la langue française* (OQLF) and the *Comité canado-québécois de terminologie de la télédétection* since 1978, and the federal government has published some glossaries with the Canadian Space Agency (*Comité d'uniformisation*, 1994) and the Canada Centre for Remote Sensing (Canada Centre, 2005). In Germany, also, remote sensing terminology has long been a matter of concern (Albertz, 1977) which has been addressed by standardization² and the publication of a dictionary by the *Bundesamt für Kartographie und Geodäsie* (Lindig, 1993). Italy has also shown an interest in remote sensing terminology through the *Italian Association of Remote Sensing* (Brivio & Zani, 1995) and the *Consiglio Nazionale delle Ricerche* (Grignetti et al., 2003, 2005). Meanwhile, in other languages, such as Spanish, joint and integrated actions have hardly been promoted (Sanz Vicente, 2007: 25).

In Spanish only one standardization attempt has been made, namely the dictionary compiled by the Association of Latin American Experts in Remote Sensing (SELPER), with English and Portuguese equivalents (SELPER, 1989). The absence of terminological studies in Spanish makes this dictionary of due reference, even if it is based on a questionable methodology: the translation of the terms included in the ISPRS dictionary. The SELPER dictionary has been recently updated, including new terms and equivalents in French but following the same methodology. This up to date version is freely available on the Internet and merely offers language equivalents (Raed et al., 2008). In Spain there are no specific studies on remote sensing terminology. There is only a cartography dictionary (Alcalá et al., 1995) that includes remote sensing terms derived from the SELPER dictionary and from standardization works in France. There are no specific publications on this terminology in Spain, and vocabularies come down to those included in two academic books on remote sensing (Chuvienco, 2002: 569–573; Pinilla, 1995: 289–297). Accordingly, no clear commitment has been made towards terms in context, term variation and unnecessary neologisms and loan-words. In short, we lack a

1. Cf. regulations of the Journal Officiel de la République Française (JO) about aerospace remote sensing vocabulary: JO 28-11-1980, JO 14-3-1982, JO 20-10-1984, JO 17-01-1986, JO 17-04-1987, JO 09-09-1988, JO 26-09-1990 and JO 14-02-95.

2. Cf. regulations of the Deutsches Institut für Normung (DIN) about photogrammetry and remote sensing: 1995: DIN-18716-1, 1996: DIN 18716-2, 1997: DIN 18716-3, 2001: DIN 18740-1, 2003: DIN 18740-3.

thorough and systematic examination of this field from a terminological point of view to describe and organize its specific vocabulary in Spanish.

Unfortunately, the above circumstances surrounding remote sensing in Spanish are not exclusive to this domain. They keep recurring because, in spite of continuous references made by scientists and linguists to the need to intervene in these new vocabularies, no serious standardization attempts have been made. Standardization tends to be developed within language policies of speech communities threatened by a dominant language, such as French in Quebec or Catalan in Catalonia. As for majority and multinational languages, the closer example is that of French. This language has undergone a formal standardization process, which particularly arose from France. And, in view of the English expansion, this standardization has been mostly defensive, as revealed in the prescriptive nature of its standardization procedures. Attitude towards loan-words, however, has not always been the same in France and Quebec. In the beginning, both of them adopted parallel attitudes, but Quebec has evolved towards a less defensive policy. This Canadian province has recently adopted a more realistic strategy, from a linguistic point of view, *'qui écarte les prises de position exclusivement défensives à l'égard de l'emprunt ou, à l'opposé, celles qui lui seraient exagérément favorables'* (Office québécois, 2007: 4).

All standardization examples given here for other countries and speech communities reinforce what we have observed and highlighted as a priority in Spanish: the need to standardize remote sensing terminology. This standardization should be immediately addressed with sustained actions, but remembering how general language and languages for specific purposes (LSP) work. It should have a non-prescriptive purpose and care about language use, terms in context and not see the dominant language, English, as an enemy.

3. An approach to remote sensing

The term 'remote sensing' is a relatively new addition to the technical lexicon. It was first used in the U.S. in the late 1950s by Evelyn Pruitt, a geographer and oceanographer of the U.S. Office of Naval Research, to refer to the new and more *remote* medium for recording images of the Earth's surface from space (Short, 2006), in contrast to the traditional one of recording images from aeroplanes, aerial photography. The Spanish translation of the English term, *'teledetección'*, derives from the French *'télédétection'*, a translation coined in 1967 (Sobrino, 2000: 19). There are other Spanish translations, such as *'percepción remota'* and *'sensores remotos'*, mainly used in Latin American countries. But the most frequently used in Spanish-speaking countries

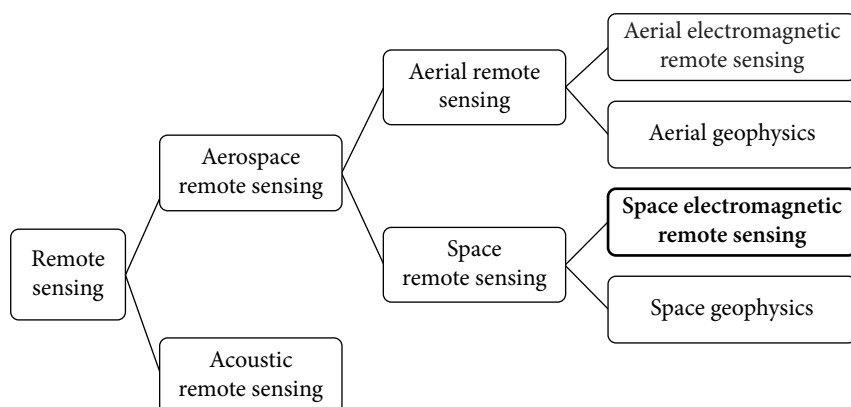


Figure 1. Remote sensing fields and subfields, based on Paul et al. 1991: 64

is *teledetección*, perhaps due to the Greek combining form *tele-*, meaning ‘distant’ or ‘far off’, commonly used in the formation of compound words in Spanish.

Remote sensing generally refers to the technique of obtaining information about an object, surface or phenomenon through the analysis of data recorded by a device that is not in physical contact with it. In this study we approach the domain from an environmental point of view – that of sciences such as geography, geology, forestry, agriculture, meteorology, etc. –, which see remote sensing as the technique of acquiring, processing and interpreting images of the Earth’s surface from space for environmental purposes. This also means we consider aerospace remote sensing, i.e., satellite and spacecraft remote sensing and, particularly, space electromagnetic remote sensing (see Figure 1). This remote sensing subfield is confined to space remote sensing, different from acoustic remote sensing – based on sonar systems –, aerial and space geophysics – based on gravity and magnetic techniques –, and aerial remote sensing – the traditional technology of aerial photography.

3.1 Space electromagnetic remote sensing

First of all, and because of its final aim – Earth observation and study from space –, space electromagnetic remote sensing is an interdisciplinary and transdisciplinary field. It is inherently interdisciplinary because many different sciences and experts are involved in the process of acquiring images of the Earth’s surface from sensors on board space platforms: physicians, who study the radiation principles and their interaction with the Earth’s surface; engineers, who design sensors and satellite systems and put them in orbit; mathematicians and statisticians, who process data; and computer experts, who develop and implement software. But it is also transdisciplinary because these images can be interpreted for different environmental applications (geology, biology, meteorology, oceanography, cartography...).

Secondly, space electromagnetic remote sensing is a new but emerging field. It dates from the 1960s and has experienced a rapid rise in the last 20 years thanks to scientific and technological development. The trend towards higher resolution sensors has steadily and exponentially increased the volume of data acquired, and there has been a sustained improvement in image processing systems. It has now become possible to cover and analyse more and more of the Earth's surface, more frequently and using wider sections of the electromagnetic spectrum.

Finally, the U.S. is the leading country in terms of number of users, organizations, companies, scientific research, university courses and publications, thanks to high and sustained public investment in space programmes. In contrast, remote sensing has experienced a late, limited and slow development in Spain, speeded up in the 1990s when remote sensing courses were introduced at university. Meanwhile, Latin American countries show a slower and later growth, expected to pick up in the coming years.

These three aspects – inter- and transdisciplinarity, recent origins but rapid growth and highest development in the U.S. – manifest themselves in the terminology of this domain. They explain the existence of many terms from other disciplines, the presence of some term variation and the large number of English loan-words in languages such as Spanish.

3.2 Remote sensing terms in Spanish and English

The most significant features regarding these terms can be briefly summarized as follows. Some terms come from sciences such as photography (filter), physics (spectral band) and computing (pixel), and show less variation as they mainly stem from a long-established tradition. But most are newly coined terms in the specific domain of remote sensing, which have not yet become established, and show more variants (i.e., synonyms) and non-uniform patterns. For example 'digital number', 'grey level', 'pixel value' and 'brightness value' express the same concept. Some terms have a Greco-Latin origin (bistatic, geostationary, hyperspectral, infrared...), and some are usually formed by complex syntactic structures whose base often corresponds to one of the essential remote sensing concepts (angles, images, instruments), such as, for example, 'angle' (illumination angle, pitch angle, detector analysis angle, detector angle of view, angle of field). Finally, many of them are abbreviations and acronyms that sometimes result in a lexicalized form as 'lidar' (Light Intensity Detection and Ranging).

The introduction of those English terms into other languages results in an increasing number of English loan-words. English is now the most scientifically active language, and thus the one that loans more terms. Many new concepts are directly created and named in English and then imported to other languages through

borrowing and translation. As a result, the development of this terminology in Spanish has been mainly determined by the dominance of the English language. English terms are often used in Spanish texts, usually along with a translation equivalent quite close to the English term. Therefore, there is a close correspondence between the Spanish neologism and its English counterpart, which facilitates its gradual acceptance among Spanish-speaking experts and the ongoing replacement of the foreign term. This process is common to all scientific languages, and causes many problems in the target language: calquing of English term formation patterns, misleading variation, false friends, etc. This transfer of terms from English to Spanish results in: first of all, direct loans (*raster*), loan translations (*falso color* > false colour), half translations (*transformación tasseled cap* > tasseled cap transformation), semantic loan-words (*mosaico* > mosaic) and the use of English abbreviations and acronyms but the translation of their full expressions (IFOV, *campo de visión instantáneo*); and, secondly, the transfer of the term variation existing in English (*nivel digital*, *nivel de gris*, *valor de píxel*, *luminancia*).

4. Proposals to standardize remote sensing terminology in Spanish

Bearing in mind the context set out in previous sections, a systematic analysis from a cross-linguistic perspective seems absolutely necessary in Spanish. The scarce number of bilingual lexical resources and the lack of standardization procedures demand bilingual studies, including more than language equivalents, to really help improve professional communication and meet the translators' needs.

The aim is to address this terminology from a systematic and cross-linguistic perspective to give a deep and detailed description of its usage. That has never been done before, because existing studies in Spanish do not reach complete conceptual subfields (such as space electromagnetic remote sensing) and terminology descriptions do not derive from real texts. As outlined in Section 2, a full description of this terminology and its usage is yet to be carried out in Spanish, unlike in other speech communities (French or German, for example). The following proposals represent a small but necessary step towards meeting the difficult challenge of organising and standardising this terminology. These proposals and their theoretical and methodological context are discussed in the following sections.

4.1 Theoretical and methodological context

From a theoretical point of view, these proposals are in line with the recent communicative approaches to two applied linguistic areas that are undoubtedly related: terminology and translation. And from a methodological point of view they

are particularly founded on a specific terminological application: corpus-based terminography,³ that results from the closeness of ties between descriptive-communicative terminography and corpus-based lexicography when they agree in the need for studying real data and base their methodology on the use of corpora (Pérez, 2002: 4.1.4). In modern terminology practice, the emphasis is consequently on usage and real texts as primary source of data. Thus, corpus-based terminography relies on context. It is its starting point in that it places terms, illustrates their real use and their language functioning, helps decide their definition and translation and shows their most frequent syntactic structures and their most significant collocations. Consequently, corpora work both as general contextual frameworks and as a group of smaller contexts that helps us describe a conceptual domain and its terms, as De Bessé suggests:

Le contexte est le point de départ de tout travail terminographique. C'est en effet le dépouillement du corpus, qui est à la fois un macrocontexte et une collection de microcontextes, qui permet de décrire un ensemble conceptuel et qui fournit les informations et les matériaux nécessaires à cette description. (1991: 115–116)

As a result, text corpora are essential in terminography studies today because they represent its documentary basis.

This theoretical and methodological framework results from our ultimate goal: the compilation of a remote sensing bilingual terminological database to improve professional communication in this field and help translators and interpreters. To meet this goal, it is essential, first, to include in the database and for each language, equivalent terms and information about remote sensing concepts and their conceptual relationships and, secondly, to compile information about each term and how to use it in context. It is very important to give a phraseological dimension to explain neologisms and provide a full description of the terms:

From a terminological perspective, meeting the requirements of effective communications means incorporating the social-interaction aspect into neology work methods as outlined earlier, and giving a phraseological dimension to terminology research by describing the actual functioning of terms in LSP discourse.

(Pavel, 1993: 28)

Context plays a key role in term description, and using a bilingual corpus-based method we can easily fulfil the above requirements. Corpus analysis is an effective way to analyse terms in their full extent, i.e., as units of knowledge, language and communication (Cabré, 2003: 183). Contexts are relevant for the identification of usage, because from them we can extract the conceptual, linguistic and contextual information needed to draw up bilingual terminological resources.

3. Cf. Meyer and Mackintosh (1996), Bowker (1996) and Pearson (1998).

In short, this paper proposes, on the one hand, descriptive terminography as a pre-requisite for terminology organization and standardization and, on the other, bilingual corpus analysis as a basis for recognizing and extracting terms and for retrieving their contextual information. We base our work on real texts from the remote sensing domain and on the close co-operation between terminologists, linguists and experts on remote sensing (Sanz Vicente, 2008: 265).

4.2 Method: bilingual corpus analysis

Taking into account the context set out in Section 3, and in the light of these new communicative approaches, we propose a methodology based on bilingual corpus analysis to carry out the standardization of this terminology.

4.2.1 *Corpus design and compilation*

The desirable corpora would be composite bilingual ones, comparable and parallel, but due to the lack of parallel texts (English-Spanish translations) and the urge to study real texts, the proposed methodology is only based on the use of bilingual comparable corpora. This means that the corpus will only include original texts in English and Spanish and not translations. We will rely on comparable corpora, i.e., on the compilation of original texts in each language, similar in terms of subset matter, size, text type, communicative setting and publication date, in order to be able to carry out cross-linguistic comparisons. We are aware of the significant imbalance between the two languages concerned. There is a disproportionately higher number of publications in English, recent advances in remote sensing are generally presented in this language and, as a result, Spanish texts are largely dependent on English ones. Even the most important references to translate and write documents in Spanish have been written in English or derived from English translations.

We focus on text selection criteria to guarantee comparability between corpora. For this purpose, we follow Corazzari and Picchi (1994) who suggest a combination of external criteria (field, publication date, text type...) and internal criteria (a list of key words taken from remote sensing vocabularies) to ensure comparability between comparable corpora. These internal and external criteria should be shared by both corpora (language A corpus and language B corpus). The preliminary list of key words helps to achieve comparability and provides a list of term candidates in both languages that will be revised and extended during corpus processing. This will imply revising text selection criteria and extending the corpora to meet the new criteria. To see if the corpus meets the intended objectives of our project, we will proceed cyclically as suggested in Biber (1993: 256). By doing this we will prove if it is well-balanced from a conceptual point of view and if it is representative of the text types produced in this field.

The corpus will particularly contain texts from a specific domain, electromagnetic remote sensing and from different environmental applications: meteorology, vegetation assessment and agriculture, natural disasters, geology and soil mapping, coastal processes, urban planning, and paying more attention to those with a strong tradition in Spanish, such as agriculture, land change detection, climatology and forest fires. It will comprise the following text types: papers from learned journals and congress proceedings, research reports, Ph.D. thesis, academic books and subject-specific textbooks published between 1990 and 2007. The number of publications in Spanish is relatively low compared to English. For example, there are just two subject-specific journals, *Revista SELPER* and *Revista de Teledetección* – the latter edited by the Spanish Association of Remote Sensing (AET) –, against more than ten in English, and about four others in Spanish which deal with remote sensing issues, against fifteen. Corpora size will thus depend on the number of publications available in Spanish but it should be as large as possible to give adequate results. All those publications correspond to two communicative settings, expert-expert communication and expert to initiates,⁴ which show a high density of terms. Finally, the list of key words to ensure comparability between corpora will result from the analysis of different monolingual and bilingual glossaries and dictionaries, paying special attention to the SELPER dictionary (SELPER, 1989; Raed et al., 2008).

4.2.2 *Procedural steps and corpus processing*

The overall methodology follows the process proposed by the Communicative Theory of Terminology (Cabr , 1999: 143–146), which comprises six main stages: (1) acquiring cognitive competence in the field, (2) specifying the field and describing the work, (3) preparing and organizing the work, (4) collecting terminology, (5) selecting and processing difficult cases and (6) laying out the results. But it introduces changes in the third stage to adapt to comparable corpus analysis.

The first stage is devoted to acquiring sufficient knowledge in the field by resorting to all types of information and to remote sensing experts. This means analysing the subject context and its professional matters, and becoming familiar with the basic subject-specific concepts. After that, we will be able to draw up a rough conceptual structure of the domain showing the most important conceptual relationships. This conceptual structure will be open to changes during the next stages and will invariably be supervised by remote sensing experts. The second stage entails specifying the field and reflecting on the final product users, as well as describing the objectives and purposes of the study.

4. We follow Pearson (1998: 35–39) and distinguish between expert to expert communication, expert to initiates, relative expert to the uninitiated and teacher-pupil communication.

These two first stages are essential, but the proposed method is mostly developed within the third stage, focused on preparing and organizing the work and on designing and compiling the corpora. This stage has been adapted to comparable corpus analysis with the introduction of some specific features, some of them already outlined in Section 4.2.1 (see Figure 2). When the corpus has been compiled, according to the external and internal criteria described above, we will start extracting term candidates in each language and then finding language equivalents between term candidates.

The first step is to identify the terms contained in each sub-corpus (English and Spanish). In order to do this, we use frequency lists retrieved automatically with the WordList toolkit provided by WordSmith Tools.⁵ The WordList toolkit produces lists of single words as well as two-word and three-word clusters from which we can identify single-word and multi-word term candidates. Comparing the frequency lists and the key word lists drawn from glossaries and dictionaries, we will obtain a first list of term candidates. Once a preliminary list of term candidates has been drawn up for each language, the next step is to produce concordances for them, first in the English sub-corpus and then in the Spanish one, using the Concord tool provided by WordSmith Tools. The analysis of term concordances is mainly aimed at giving a phraseological dimension, but also at comparing terms across Spanish and English in order to assess the extent to which they are functionally complete units of meaning – that is, units of language (single words and multi-word units) – that are comparable across languages in denotation, connotation and pragmatic meaning. Establishing to what extent terms are comparable across the two languages teaches us about their meaning and function as well as the contexts in which they can be accurately used.

The gaps that will result in the list of terms from this cross-linguistic text mining will be filled using dictionaries, glossaries and additional resources and asking domain experts. Afterwards, the corpora will probably need to be extended again with more texts, so once more we will be proceeding on a cyclical way checking the initial corpora. After this cross-linguistic comparison we will be ready to redesign and complete the conceptual structure roughly described on the first stage. This conceptual structure will help us build up a more detailed list of terms and will guide us to find conceptual and terminological gaps between languages, in order to compile the final list of terms for each language.

5. A corpus processing tool designed by Michael Scott at the University of Liverpool, UK. WordSmith Tools 5.0 [Online]. Available: <http://www.lexically.net/wordsmith/>

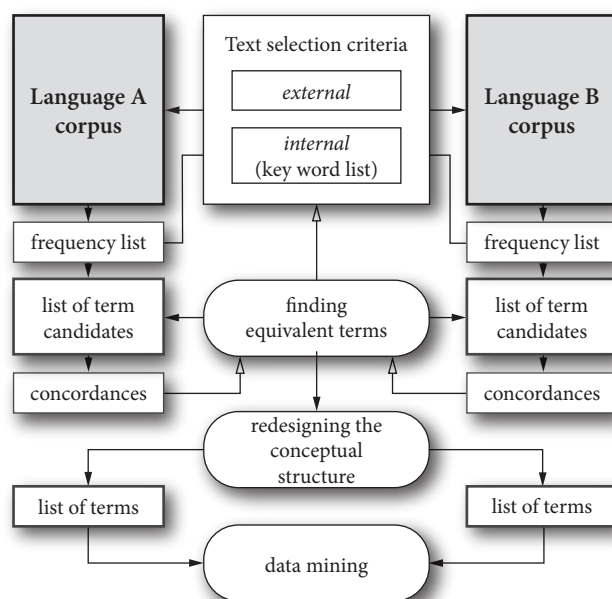


Figure 2. Proposed method based on comparable corpus analysis

Having clearly defined the list of terms, we will move on to the fourth stage. We will use concordance and frequency lists to analyse each corpus and search for information about terms, their meanings, their contexts of use, their most frequent collocations, their related terms, etc. The results obtained in each language will be included in the database, which will in fact consist of two databases, one for English and another for Spanish. Each entry in the database will include at least the following fields: headword, reference, grammar label, subject field and sub-field, definition, one or several contexts, related terms including cross-references and tips on usage. Then, both databases will be merged in a bilingual one with the same fields stated for each language database (see Figure 3).

The last two stages deal with selecting and processing the most difficult cases and laying out the results. Ultimately, all stages have to fulfil the same criterion, to adapt to the objectives and purposes established in the second stage. This criterion constitutes the cornerstone of the process proposed by the Communicative Theory of Terminology, and it also reinforces the need to proceed cyclically to see if the corpus meets the aims and requirements of the project in every stage.

Entry Number	80
Field	REMOTE SENSING
Subfield	space electromagnetic remote sensing
English	
tasseled cap transformation <i>G.L.: n. Morph: noun phrase (adj + n) + n</i> <i>Spec.: very high Internet: > 3,000</i>	
Definition	Linear transformation performed on each pixel of an image, or of a specified region of an image, to enhance different vegetation and soil properties. It consists of four orthogonal indices for Landsat MSS bands 4 to 7 (brightness, greenness, yellowness and non-such), and three for Landsat TM bands 1 to 5 and 7 (brightness, greenness and wetness). <i>Ref.: CCRS/LSV</i>
Context	Two spectral enhancements were performed on the raw six band Landsat TM band stack. Both an NDVI and a tasseled cap transformation were performed on the spectral data. The NDVI and the first three bands of the tasseled cap (representing greenness, wetness, and brightness respectively) were used as variables in the model. <i>Ref.: JAS 2004</i>
Related terms	TCT acronym Kauth-Thomas transform synonym
Variant	tasseled-cap transformation
ES equivalent	transformación tasseled cap
Español	
transformación tasseled cap <i>G.L.: n. Morph: noun phrase (adj + n) + n</i> <i>Spec.: very high Internet: < 1,000</i>	
Definición	Transformación lineal que se aplica a cada píxel de una imagen o región de una imagen para resaltar diferentes propiedades de la vegetación y los suelos. Consta de cuatro índices ortogonales para las bandas 4 a 7 de Landsat MSS (brillo, verdor, marchitez y uno sin significado aparente), y con tres para las bandas 1 a 5 y 7 de Landsat TM (brillo, verdor y humedad). <i>Ref.: CCRS/LSV</i>
Contexto	Al igual que con el Análisis de Componentes Principales, el objetivo de la transformación Tasseled Cap (T.T.C.) es obtener un nuevo conjunto de bandas mediante la combinación lineal de las originales a fin de realzar algunos rasgos de interés en la escena. <i>Ref.: BRA 1998</i>
Remisión	TTC sigla
Variantes	transformación "tasseled cap" transformación <i>tasseled cap</i>
EN equivalente	tasseled cap transformation

Figure 3. Sample term record of the bilingual database

5. Conclusion

To conclude, we can give a brief description of our first results in the application of the proposed methodology. The first stage has been completed with the creation of a remote sensing documentary database and the design of a rough conceptual structure of the domain to guide the terminological analysis. The second stage has also been completed with the definition and specification of the particular subject field analysed and the description of the theoretical and methodological considerations presented here. We are currently working on the third stage, which has been partially completed with the proposal of a specific methodological process and the

compilation of texts – identification of potential material, OCR scanning, when necessary, and conversion to .txt format – to compile the electronic corpora.

In short, the proposed methodology underlines the need to work with real texts, stresses the importance of creating a bilingual terminological database to improve communication in this field, highlights the need for interdisciplinary co-operation between terminologists, linguists and experts on remote sensing, and represents an initial modest step towards the organization and standardization of remote sensing terminology in Spanish. This paper ratifies the need to take action over the emergence of new terminologies, such as remote sensing, brought about by rapid advances in scientific research and high-technology development.

Acknowledgements

The research described in this paper has been jointly funded by the Spanish 'Consejería de Educación de la Junta de Castilla y León' and the European Social Fund.

References

- Albertz, J. (1977). Vorschläge für eine einheitliche Terminologie in der Fernerkundung. *Bildmessung und Luftbildwesen* 4, 119–124.
- Alcalá, A. R. et al. (1995). *Diccionario de cartografía: topografía, fotogrametría, teledetección, GPS, GIS, MDT*. Madrid: Ediciones de las Ciencias Sociales.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing* 8 (4), 243–257.
- Bowker, L. (1996). Towards a corpus-based approach to terminography. *Terminology* 3 (1), 27–52.
- Brivio P. A. and Zani, G. (1995). *Glossario Trilingue di Telerilevamento*. Milano: Associazione Italiana di Telerilevamento (AIT). <http://milano.irea.cnr.it/3gloss/glossario.htm>
- Cabré, M. T. (1999). *La terminología: representación y comunicación: elementos para una teoría de base comunicativa y otros artículos*. Barcelona: IULA, Universitat Pompeu Fabra.
- Cabré, M. T. (2003). Theories of terminology. Their description, prescription and explanation. *Terminology* 9 (2), 163–200.
- Canada Centre for Remote Sensing (2005). *Glossary of Remote Sensing Terms = Glossaire des termes de télédétection*. Ottawa: Natural Resources Canada (NRC). http://www.ccrs.nrcan.gc.ca/ccrs/learn/terms/glossary/glossary_e.html
- Chuvieco, E. (2002). *Teledetección ambiental. La observación de la Tierra desde el espacio*. Barcelona: Ariel.
- Corazzari, O. and Picchi, E. (1994). A proposal for the construction of comparable multilingual corpora. *Multext Project Report*. Pisa: Istituto di Linguistica Computazionale, CNR.
- Comité d'uniformisation de la terminologie spatiale (1994). *Vocabulaire de RADARSAT et de la télédétection hyperfréquence = RADARSAT and Microwave Remote Sensing Vocabulary. Terminology bulletin 229*. Ottawa: Groupe Communication Canada.

- De Bessé, B. (1991). Le contexte terminographique. *Meta*, 36 (1), 111–120.
- Fuentes, M. T. and García Palacios, J. (2002). Los diccionarios de especialidad y el traductor. In G. Guerrero and M. F. Pérez (eds.), *Panorama actual de la terminología* (117–136). Madrid: Comares.
- Grignetti, A. et al. (2003). Terminologia e interoperabilità semantica per il Telerilevamento e il GIS. In *Proceedings of the 8ª Conferenza Nazionale ASITA: Geomatica, standardizzazione, interoperabilità e nuove tecnologie*, (1255–1259). Milano: Federazione delle Associazioni Scientifiche per le Informazioni Territoriali e Ambientali (ASITA). http://www.t-reks.cnr.it/docs/termgis_asita2004.pdf
- Grignetti, A. et al. (2005). A Thesaurus for Remote Sensing and GIS: preliminary version and future plans. In, M. Flörke et al. (eds.), *Proceedings of the 19th International Conference Informatics for Environmental Protection, Enviroinfo 2005* (783–787). Brno: Masaryk University. http://www.t-reks.cnr.it/docs/Brno_48.pdf
- Lindig, G. (1993). *Deutsches Fachwörterbuch Photogrammetrie und Fernerkundung: deutschsprachige Benennungen und Definitionen mit vorläufigen englischen und französischen Äquivalenten sowie deutschen, englischen und französischen Stichwortlisten*. Frankfurt a.M.: Verlag des Instituts für Angewandte Geodäsie.
- Meyer, I. and Mackintosh, K. (1996). The corpus from a terminographer's viewpoint. *International Journal of Corpus Linguistics*, 1 (2), 257–285.
- Office québécois de la langue française (2007). *Politique de l'emprunt linguistique*. Québec: Office québécois de la langue française (OQLF).
- Paul, S. et al. (1982). *Dictionnaire de télédétection aérospatiale = Airborne and spaceborne remote sensing dictionary*. Paris: Masson.
- Paul, S. et al. (1991). *Introduction à l'étude de la télédétection aérospatiale et de son vocabulaire*. Paris: La Documentation Française.
- Paul, S. et al. (1997). *Manuel terminologique didactique de télédétection et photogrammétrie: français-anglais*. Paris: CILF.
- Pavel, S. (1993). Neology and Phraseology as Terminology-in-the-making. In H. B. Sonneveld and K. L. Loening (eds.), *Terminology: applications in interdisciplinary communication* (21–34). Amsterdam & Philadelphia: John Benjamins.
- Pearson, J. (1998). *Terms in Context*. Amsterdam & Philadelphia: John Benjamins.
- Pérez, C. (2002). *Explotación de los corpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento*. PhD diss., Universidad de Málaga, Spain.
- Pinilla, C. (1995). *Elementos de teledetección*. Madrid: Ra-Ma.
- Rabchevsky, G. A. (1984). *Multilingual Dictionary of Remote Sensing and Photogrammetry: English Glossary and Dictionary, Equivalent Terms in French, German, Italian, Portuguese*. Falls Church: ASPRS.
- Raed, M. A. et al. (2008). *Diccionario de términos de percepción remota: equivalencias de español, inglés, portugués, y francés*. Bogotá: SELPER. <http://www.selper.org/2007/diccionario08/COMIENZO X.htm>
- Sanz Vicente, M. L. (2007). La terminología de la teledetección: experiencias de adquisición léxica bilingüe. In I. Ahumada (ed.), *Lenguas de especialidad y lenguajes documentales* (19–34). Madrid: Asociación Española de Terminología (AETER).
- Sanz Vicente, M. L. (2008). Propuesta metodológica basada en el uso de corpus bilingües para la elaboración de un diccionario de teledetección inglés-español. In D. Azorín et al. (eds.), *Proceedings of the II Congreso Internacional de Lexicografía Hispánica: El diccionario como*

- punte entre las lenguas y culturas del mundo* (264–270). [CD-ROM] Alicante: Universidad de Alicante – Fundación Biblioteca Virtual Miguel de Cervantes.
- SELPER. (1989). *Diccionario SELPER: Percepción remota: inglés-español-portugués*. Lima: SELPER.
- Short, N. M. (2006). *The Remote Sensing Tutorial*. Maryland: Applied Information Sciences Branch, NASA's Goddard Space Flight Center (GSFC). <http://rst.gsfc.nasa.gov>
- Sobrino, J. A. (2000). *Teledetección*. Valencia: Servicio de publicaciones, Universitat de València.

SECTION IV

Terminology in a medical setting

The PERTOMed project

Exploiting and validating terminological resources of comparable Russian-French-English corpora within pharmacovigilance

Cedric Bousquet and Maria Zimina

The PERTOMed project is a pluri-disciplinary research initiative undertaken by several institutions in France. Applications considered within the part of the project described in this article concern pharmacovigilance and adverse drug reactions. We had multiple objectives: to create a specialized Russian Internet corpus; to test new tools and methods for term extraction from comparable multilingual texts and to build terminological resources including Russian. A trilingual Russian-French-English lexicon resulting from this work is freely available from the PERTOMed server.

Keywords: adverse drug reactions, alignment, comparable corpora, corpus linguistics, medical texts, natural language processing, parallel corpora, pharmacovigilance, Russian, term extraction.

1. Introduction: The PERTOMed project

Our study presents the results of part of a collective research project on the production and evaluation of terminological and ontological resources in the medical field – PERTOMed (in French: Production et Evaluation de Ressources Terminologiques et Ontologiques dans le domaine de la MEDecine) (Charlet et al., 2006).

Development of terminological resources in medicine is a major issue towards collecting data and browsing knowledge databases. In this respect, one of the project's main objectives (officially finished in December 2005) was to explore *Natural Language Processing* (NLP) tools and methodologies for compiling terminological resources from parallel and comparable medical texts in several languages. Terminology extraction from texts was then considered to design better

tools to help specialists code acts and diagnoses with an ontology-based software, Baneyx et al. (2007).

Potential applications considered within PERTOMed covered several fields:

- Pharmacovigilance
- Pneumology
- Drug-drug interactions
- Multilingual terminology management.

The project resulted in cross-field collaboration, critical thinking and exchanges among linguists, specialists of natural language processing and knowledge engineers, information scientists and medical practitioners. Several research institutions contributed to PERTOMed:

- INSERM UMR_S 872, Eq 20, Faculté de Médecine – Paris 5 (France).
- ERSS: Equipe de Recherche en Syntaxe et Sémantique, UMR 5610 CNRS and Toulouse le Mirail University (France).
- CRIM: Centre de Recherche en Ingénierie Multilingue, INaLCO (Paris, France).

Overall project management and co-ordination were performed by the INSERM research team under the direction of M.-Ch. Jaulent and J. Charlet.

As part of our contribution to the project, we conducted a series of experiments in order to develop, explore and evaluate terminological resources of comparable Russian-French-English medical text corpora sharing common semantic and pragmatic characteristics within pharmacovigilance. According to the World Health Organization (WHO), pharmacovigilance is *'the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other drug-related problems'*, WHO (2002).

We pursued three objectives for multilingual terminological study using Russian:

- Creation of a Russian Internet corpus within Pharmacovigilance;
- Study of methods for building terminologies from comparable corpora;
- Elaboration of French-English-Russian adverse reaction terminological resources.

The trilingual Russian-French-English lexicon resulting from this work is available on the Web from the PERTOMed server.¹

Multilingual terminology management within PERTOMed turned out to be quite a challenge from the point of view of research methodology, tools and software. In this article, we focus on some essential results of this work, methodological findings and suggestions for further research, keeping in mind the fact that many new projects faced with disparities across multilingual medical texts are most likely to come up.

1. The PERTOMed web server: <http://www.pertomed.info>

2. Multilingual terminology management within pharmacovigilance

2.1 Pharmacovigilance

In a most general sense, practical pharmacovigilance involves collecting, monitoring, researching, assessing and evaluating information from healthcare providers and patients on the adverse effects of medications. This pharmacological science is particularly concerned with Adverse Drug Reactions (ADRs). According to the World Health Organization, ADR means a response to a drug which is noxious and unintended, and which occurs at doses normally used for the prophylaxis, diagnosis or therapy of disease, or for the modification of physiological function, WHO (1972).

Pharmacovigilance plays an increasingly important role, since the number of drug recalls is growing rapidly. Clinical trials usually involve limited study populations and might be insufficient to detect possible side effects and ADRs at the time a drug enters the market. For this reason, research tools and activities within Pharmacovigilance (including data mining and investigation of case reports) are essential to identify possible relationships between drugs and ADRs.

2.2 International terminologies within pharmacovigilance

Since reports on side-effects and ADRs contain a wide variety of domain specific terms, maintaining international terminological standards for pharmacovigilance is of great relevance. However, due to the existence of well-known historic and cultural boundaries between countries, building terminological resources worldwide is not just a trivial issue. For the moment, two international terminological resources are widely acknowledged in this field: *World Health Organization – Adverse Reaction Terminology* (WHO-ART) and *Medical Dictionary for Drug regulatory Activities* (MedDRA).

WHO-ART has been in use more than thirty years for coding adverse reaction terms in relation to drug therapy. It is widely used by drug regulatory agencies and pharmaceutical manufacturers in many countries. WHO-ART is initially developed in English with more or less complete translations into French, German, Spanish, Portuguese and Italian. The system is maintained by the WHO Collaborating Centre for International Drug Monitoring, Uppsala Monitoring Centre (UMC).

MedDRA defines fully equivalent medical terms in different languages, including English, French, German, Japanese and Spanish. This international terminology applies to almost all stages of drug development, as well as health effects and malfunction of devices. MedDRA is owned by the International Federation of Pharmaceutical Manufacturers and Associations (IFPMA).

2.3 Pharmacy-related issues in Russia

The Russian pharmaceutical industry and distribution network underwent radical transformations after the break-up of the USSR. Rapid political changes and transition to a market economy, resulted in massive disruption to pharmaceutical production and the distortion of co-operative links between former partners. Today, many efforts are still required to establish countrywide activities, as pharmaceutical distribution in Russia remains fragmented, with thousands of pharmaceutical wholesalers. Only few of them offer nationwide coverage.

In Russia, all drugs and biological products must be registered with the Ministry of Health at the federal level. The federal and regional governments are entitled to develop and use lists of essential drugs, and give recommendations for supply and use in the public health system. Enterprises and insurance companies develop their own versions. Thus, many lists of essential drugs have been developed in parallel, based on several criteria, such as efficacy, safety, price (as well as origin and production characteristics).

In 1997, the Russian Federation established a federal centre for monitoring adverse reactions, which has joined the adverse drug reaction monitoring programme within the WHO. WHO-ART is not currently being translated into Russian. Terminological resources in Russian are available within the *International Classification of Diseases* (ICD), currently being adopted in Russia to classify diseases and health problems on many types of medical records.

Progressive integration of the Russian Federation and language into the World Health system is an active process. At the international level, there exist medical terminologies including Russian, such as a Russian translation of *Medical Subject Headings* (MeSH) in at least two versions developed by different institutions.² On a national level, translation of terminological resources is in progress for health-care disciplines and medical sciences in general.³

2. MeSH offers controlled vocabulary and thesaurus features used to index, catalogue and retrieve the world's medical literature. It has been translated into Russian by the US *National Library of Medicine* (NLM), Russian National Public Library for Science and Technology (CYRMESH integrated within automated library system IRBIS).

3. English translation provided on the following Russian web sites: *Antibiotics and Antimicrobial Therapy*: <http://antibiotic.ru/index.php?newlang=eng> and *Interregional Association for Clinical Microbiology and Antimicrobial Chemotherapy*: <http://www.iacmac.ru/iacmac/en/>

3. Parallel vs. comparable text processing: State of the art

In the past few years, many efficient tools and methods have been developed for bilingual lexicon extraction from parallel corpora (source texts accompanied by their translations in one or several languages), Véronis (2000). It is commonly known that parallel corpora are not always available for specific domains/language pairs. Nowadays, the objective is to exploit the richness of non-parallel yet comparable corpora existing in almost any field of knowledge.

According to Fung and McKeown (1997), a rather problematic task of bilingual lexicon extraction from non-parallel corpora can be sorted out by the statistical study of context word similarities between candidate terms, representing mutual translation pairs. In this vein, most of the work in this field is done comparing the distributional contexts of source and target words, testing several weighting factors and similarity measures, as in Chiao and Zweigenbaum (2002). This approach relies upon existing bilingual resources ('base lexicon' or 'pivot translations') used to calculate translation similarities between source and target words' context vectors, as in Sadat et al. (2002) and Morin and Daille (2004).

Developments within comparable text processing, motivate new experiments and research initiatives aiming at terminology extraction from multilingual texts from several cultural and linguistic sources.

4. The trilingual PERTOMed corpus on adverse drug reactions

For the PERTOMed project, parallel English/French translation texts of *Summaries of Product Characteristics* (SPC) formed a good starting point for ADR terminological study. For Russian, only comparable texts are currently available in the field of Pharmacovigilance. Moreover, Russian pharmaceutical documentation is of a heterogeneous nature. Lack of visibility and decentralization create difficulties in terms of localization of high-quality authentic medical data in Russia (comparable to SPCs).

4.1 Parallel English/French sub-corpus

The English/French part of the PERTOMed corpus was composed of 156 SPC downloaded in PDF format from the EMEA website by the INSERM team in February 2004.

An SPC is a special type of medical text describing a certain medicinal product's properties and the conditions attached to its use. This document is intended

for product certification performed by the European Medicines Agency (EMA)⁴ and for medical professionals. The SPCs are provided in all EU languages (without official Russian translation).

Figure 1 shows an extract of the English/French sub-corpus from the PERTOMed project. Only two SPC sections (*Section 4.5* and *Section 4.8*) containing information on drug-drug interactions and undesirable effects, were used for the corpus.

4.2 Comparable Russian sub-corpus

Localization and data retrieval for Russian were carried out by the CRIM-INaLCO team: Ivanova and Nuk (2005). First of all, a list of available pharmaceutical resources on the Russian web had to be developed, as well as a set of criteria for corpus construction.

4.2.1 *Sampling frame*

In corpus linguistics studies, a comparable multilingual corpus is usually defined as a corpus containing components that are collected using the same sampling frame, similar balance and representativeness. The components representing the languages involved, must match in terms of proportion, genre, domain and sampling period, as in McEnery and Xiao (2005).

In order to collect comparable medical data for the PERTOMed corpus, special attention was paid to the following characteristics of Russian pharmaceutical texts:

- Degree of specialization
- Recognition by domain experts in Russia
- Style (summarization)
- Clarity and precision
- Information granularity
- Availability of active component and product name indexation.⁵

Following preliminary research described in Ivanova and Nuk (2005), three Russian websites were selected for the project:

- RECIPE: <http://www.recipe.ru>
- PJIC: <http://www.rlsnet.ru>
- Russian Vidal: <http://www.vidal.ru>

4. *European Medicines Agency* (EMA) is a decentralized EU body with headquarters in London. The EMA issues certificates of a medicinal product in conformity with the arrangements laid down by the World Health Organization.

5. Following these principles, the choice of Russian pharmaceutical texts was made by selecting active components identical to those mentioned in English/French SPC sub-corpus.

LAMIVUDINE	
English	French
Lamivudine may inhibit the intracellular phosphorylation of zalcitabine when the two medicinal products are used concurrently.	La Lamivudine peut inhiber la phosphorylation intracellulaire de la zalcitabine lorsque ces deux produits sont administrés de manière concomitante.
Zeffix is therefore not recommended to be used in combination with zalcitabine.	Par conséquent, il n'est pas recommandé d'utiliser Zeffix en association avec la zalcitabine.
In clinical studies of patients with chronic hepatitis B, Lamivudine was well tolerated.	La Lamivudine a été bien tolérée au cours des essais cliniques réalisés chez des patients atteints d'hépatite B chronique.
The most common adverse events reported were malaise and fatigue, respiratory tract infections, throat and tonsil discomfort, headache, abdominal discomfort and pain, nausea, vomiting and diarrhoea/.../	Les effets indésirables le plus souvent rapportés étaient: malaise et fatigue, infections respiratoires, gêne au niveau de la gorge et des amygdales, céphalées, douleur ou gêne abdominale, nausées, vomissements et diarrhée./.../

Figure 1. Parallel English/French sub-corpus: extract

The *RECIPE* site is an online information catalogue devoted to legal pharmacological documentation of the Russian Federation. It provides a Medline user manual, detailed index of Russian bio-medical websites and offers possibilities to search for medical products using several criteria (including ICD-10).

The *P/JC* site (Russian acronym of *Регистр Лекарственных Средств России*, or *Register of Medical Substances of Russia*) is an online encyclopaedia of medical products with a precise product description.

The *Russian Vidal* website is maintained and regularly updated by the private company AstraPharmService in accordance with the Industrial Standard of the Russian Federation.

The three websites provide standard medical data acknowledged by legal authorities of the Russian Federation and referenced by international pharmaceutical partners present on the Russian market. In order to compile a balanced Russian sub-corpus of a size similar to that of the English/French one, text-to-text alignment was used to achieve best results. For each drug product listed within the English/French SPC part, localization of corresponding texts in Russian was performed through the following steps:

- Cross-check if the product is commercialized in Russia.
- Search *RECIPE*, *P/JC* and *Russian Vidal* using product name or active component.
- Localize product description pages.

<Recipe LAMIVUDINE> Побочные действия : Лейкопения, тромбоцитопения, анемия, недомогание, утомляемость, головная боль, периферическая нейропатия, парестезии, гастралгия, тошнота, рвота, диарея, панкреатит, повышение активности «печеночных» трансаминаз, гиперкреатинемия, панкреатит, повышение активности сывороточной амилазы, лихорадка, развитие вторичной инфекции. Передозировка: усиление проявлений описанных побочных действий. Лечение: промывание желудка, активированный уголь, симптоматическая терапия, непрерывный гемодиализ. <PLC LAMIVUDINE> Побочные действия: Головная боль, головокружение, слабость, нарушение сна, бессонница, гипотония, нейропатия, кашель, гриппоподобный синдром, анорексия, тошнота, диарея, рвота, боль в эпигастриальной области, некротический панкреатит (возможен летальный исход), миалгия, артралгия, лейкопения, анемия, лихорадка, потливость, аллергические реакции. <Russian Vidal LAMIVUDINE> Побочное действие : Со стороны пищеварительной системы: возможны боли и дискомфорт в эпигастриальной области, тошнота, рвота, диарея, снижение аппетита, повышение активности печеночных трансаминаз. Со стороны ЦНС: возможны повышенная утомляемость, головная боль. Со стороны дыхательной системы: возможны инфекции дыхательных путей. Прочие: возможно общее недомогание.

Figure 2. Comparable Russian sub-corpus: extract

- Extract textual data present within adverse reaction section only.
- Keep all variants in case of several possible descriptions for the same product.

Figure 2 shows the extract of the Russian sub-corpus resulting from this study described in Ivanova and Nuk (2005).

4.2.2 Evaluation by Correspondence Analysis

We used correspondence analysis (CA) available within *Lexico3*⁶ textometric toolbox to evaluate whether Russian ADR descriptions collected on three different websites, using a common sampling frame, could form a homogeneous corpus for a terminological study.

CA is a multidimensional descriptive data analysis method used for describing contingency tables (or cross-tabulations), Lebart et al. (1997). In our case, correspondence analysis was used to describe a lexical table, cross-tabulating word forms and medical texts on same drug products collected on the *RECIPE*, *PIC* and *Russian Vidal* websites. In other words, our table had as many rows as there were

6. *Lexico3* textometric toolbox presents a wide range of functions (segmentation, concordances, measurements and counts based on graphical forms, computation of characteristic elements and correspondence analyses of forms and repeated segments): <http://www.cavi.univ-paris3.fr/Ilpga/ilpga/tal/lexicoWWW/>

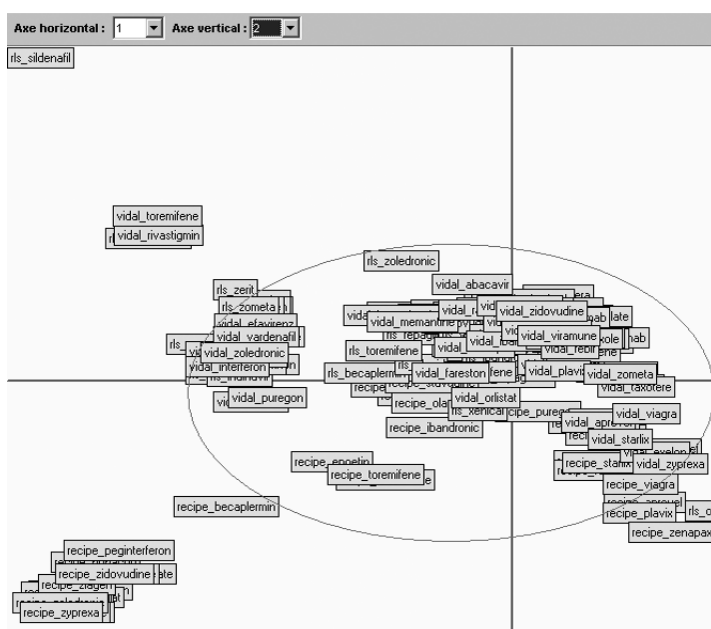


Figure 3. Comparable Russian sub-corpus: first plane of the Correspondence Analysis

word forms in collected Russian texts and as many columns as there were texts describing adverse drug reactions of each drug product. To analyse the information contained in this table, the row-profile and column-profile tables were calculated and the distances among the words on the one hand and drug product adverse reaction descriptions on the other displayed (see Figure 3).

Figure 3 shows the *first plane* of the correspondence analysis (that is, the plane of the first two principal axes). It showed a high density of groupings of columns-points (texts) in the origin of the axes, displaying important lexical affinities between texts. Only some incomplete or shortened texts departed from the *average profile*, showing common lexical characteristics of the Russian sub-corpus. This dominance of core vocabulary confirmed global terminological homogeneity of texts coming from three different sources (*Recipe*, *PJIC*, *Vidal*). The result was positive, regarding the initial objectives of our terminological study.

4.3 Main lexicometric characteristics of the trilingual PERTOMed corpus

Figure 4 shows the main lexicometric characteristics of the PERTOMed corpus. Each sub-corpus was subdivided into graphical forms using the *Lexico3* textometric toolbox. The results of automatic segmentation are represented in separate

			<i>Nb occ</i>	<i>Nb forms</i>	<i>F max</i>	<i>Hapax</i>
<i>comparable</i>	<i>parallel</i>	Russian	15 465	3 034	461	1 483
		English	133 984	5 957	4 936 (of)	1 836
		French	161 995	7 280	8 022 (de)	2 389

Delimiting characters: .,:;!/?/_\"'()[]{}\$\$

Figure 4. The trilingual PERTOMed corpus: structure and main lexicometric characteristics

columns, showing respectively the total number of occurrences of word forms (tokens) within each corpus part, the number of different word forms (types) used, most frequent word form (maximum frequency) and the total number of hapaxes (*hapax legomenon*, i.e. one-off occurrences of word forms) in a given sub-corpus.

5. Term alignment in parallel English/French texts using SYNTAX

Term alignment in parallel English/French SPC texts was performed by the ERSS team. These bilingual texts were first aligned at the sentence level using *JAPA*⁷ tool. Each word in two parts of the corpus was assigned a lemma and a grammatical tag through Part-Of-Speech (POS) tagging performed by *TreeTagger*⁸.

Both parts were then analysed with a dependency parser *SYNTAX*: Bourigault and Fabre (2000). Identification of syntactic dependencies (for instance subjects, direct and indirect objects of verbs) was performed independently for each language.

Bilingual term extraction was performed using both statistical word similarity measures (such as *Jaccard's Coefficient*) and *alignment by syntactic propagation*: Ozdowska (2004), Ozdowska et al. (2005). The description of the Bilingual French-English lexicon resulting from this work is available on the PERTOMed website. This hierarchically structured (type head-expansion) lexicon contains 1 278 validated bilingual terminological units (French term ~ suggested equivalent in English).

7. *JAPA* is a programme that aligns parallel texts at the sentence level: <http://rali.iro.umon-treal.ca/japa/>

8. *TreeTagger* is a tool for annotating text with part-of-speech and lemma information: <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger>

6. Mapping terms in comparable contexts

6.1 Corpus-based approach adopted within the PERTOMed project

Available tools and methods for processing parallel English/French sub-corpora could not be used to process comparable Russian texts. Lack of specific term-extraction software and resources for Russian⁹, made us consider quantitative corpus-based tools and methods (such as computation of word *n-gram* repetitions and *multiple co-occurrences*)¹⁰. This fine-grained context-based approach provided flexibility, in order to work simultaneously on different language combinations in trilingual contexts.

For testing purposes, term extraction for Russian was first done semi-automatically, using fine-grained segmentation into terminological units performed by the CRIM-INaLCO team: Ivanova and Nuk (2005). Information chunks on ADRs collected within a Russian Internet corpus were automatically split into distinct textual units using common punctuation marks: *comma* [,], *colon* [:], *semi-colon* [;], *dot* [.]. This type of segmentation could be applied thanks to the nature of adverse reaction texts. Those concise summary style texts are chiefly composed of enumerations describing particular groups of undesirable effects (cf. Figure 2).

Generated lists of term-candidates were then filtered to keep frequent relevant patterns, such as [Ajective + Noun] [*депрессивное состояние*, in English: *depressed state*] or [Noun_{Nominative} + Noun_{Genitive}] (*потеря аппетита*, in English: *lost appetite*). Those terminological units were put into correspondence with validated SYNTAX term candidates for French through manual context-based alignment. The results of those experiments are available within the Russian-French lexicon (485 main terms) on the PERTOMed website.

The experiments of English/French and Russian/French term alignment gave the impetus to start working on a trilingual terminological resource of adverse drug reactions. In order to avoid disjoint terminologies with cross-language boundaries, we decided to attempt to create a unified *Russian-French-English* adverse reaction terminology.

9. Localization of tools and resources for Russian NLP was unsuccessful during the project. Since then, new tools and resources have been made available. For instance, *TreeTagger* has been adapted for Russian.

10. *N-gram* is a sub-sequence of *n* items (letters, word forms, etc.) from a given text sequence. *Co-occurrence* is a simultaneous, but not necessarily contiguous presence of occurrences of two given word forms in one fragment of text.

6.2 Textometric browsing

Trilingual terminology creation was managed through textometric browsing in comparable contexts. The concept of textometric browsing, enables the researcher to move among the results produced by different quantitative methods and the original corpus, as in Zimina (2005). This interactive method of corpus exploration helps produce an automatic selection of contexts in one of the parts of the multilingual corpus where any textual unit under study (*word form, n-gram, collocation, etc.*) is found.

6.2.1 Anchor points

As a rule, initial anchor points (corresponding textual units) for exploring comparable corpora are set using several criteria. Most of the time, cognates (etymologically related words having a common origin in one or more languages, e.g. English: *syndrome*, French: *syndrome*, Russian: *синдром*), general language translation correspondences (e.g. English: *pain*, French: *douleur*, Russian: *боль*) as well as existing translation or dictionary resources contribute to anchor point identification.

In our case, *English-French* and *Russian-French* lexicons (sharing a common French part) helped identify simple translation correspondences to start browsing through trilingual texts. Moreover, in order to prospect for lexical anchors, we relied upon frequency counts produced on corpus *vocabulary* (the entire set of word forms in each corpus part) and on larger units consisting of several word forms (*segments*), as in Lebart et al. (1997).

After setting aside words with a purely grammatical role as well as the term *patient* for English and French, the most frequent word forms in the PERTOMed corpus were: *disorders* [Freq=835] in English, *troubles* [Freq=922] in French and *стороны* [Freq=216] in Russian (Genitive of *сторона*, meaning *side* in English). Those lexical items were selected as anchors to automatically recognize corresponding contexts in three languages. We used *repeated segments* and *multiple co-occurrences networks* to build context vectors.

6.2.2 Repeated segments extraction with Lexico3

A *repeated segment* (RS) is a series of consecutive word forms whose frequency is greater than or equal to 2 in the corpus: Lebart et al. (1997), Lamalle et al. (2005). Figure 5 shows an excerpt from the repeated segments inventories in Russian, French and English around the pivotal terms *стороны/troubles/disorders*. We noted that even though those terms are never suggested as translation correspondences in bilingual French/Russian or English/Russian dictionaries, they are positioned preferentially in equivalent contexts (see Figure 5).

	<i>Repeated segments</i>	<i>Freq</i>
RU	со <i>сторон</i>	216
	со <i>сторон</i> сердечно сосудистой системы	28
	со <i>сторон</i> органов	27
	со <i>сторон</i> пищеварительной системы	19
	со <i>сторон</i> органов жкт	18
	побочные действия со <i>сторон</i>	17
	со <i>сторон</i> цнс	16
	со <i>сторон</i> нервной системы	15
	со <i>сторон</i> системы	14
	побочное действие со <i>сторон</i>	13
	со <i>сторон</i> нервной системы и органов чувств	12
	со <i>сторон</i> системы кроветворения	12
	со <i>сторон</i> сердечно сосудистой системы и крови	11
	со <i>сторон</i> кожных покровов	10
	со <i>сторон</i> опорно двигательного аппарата	10
	/.../	
FR	<i>troubles</i> du	192
	<i>troubles</i> du système	130
	<i>troubles</i> de	116
	<i>troubles</i> de la	81
	<i>troubles</i> du système nerveux	79
	<i>troubles</i> gastro intestinaux	76
	<i>troubles</i> généraux	76
	<i>troubles</i> cutanés	47
	<i>troubles</i> psychiatriques	46
	<i>troubles</i> du métabolisme et de	42
	<i>troubles</i> généraux et	42
	<i>troubles</i> du métabolisme et de la nutrition	40
	<i>troubles</i> respiratoires	38
	<i>troubles</i> cutanés et	36
	<i>troubles</i> vasculaires	36
	/.../	
EN	system <i>disorders</i>	178
	<i>disorders</i> common	160
	<i>disorders</i> very	96
	general <i>disorders</i>	84
	nervous system <i>disorders</i>	77
	gastrointestinal <i>disorders</i>	69
	<i>disorders</i> uncommon	68
	<i>disorders</i> very rare	52
	<i>disorders</i> and	50
	tissue <i>disorders</i>	49
	general <i>disorders</i> and	46
	<i>disorders</i> rare	45
	general <i>disorders</i> and administration	45
	<i>disorders</i> very common	44
	subcutaneous tissue <i>disorders</i>	44
	/.../	

Figure 5. 15 most frequent repeated segments around the pivotal terms *сторон*/*troubles*/*disorders*

6.2.3 Multiple co-occurrences networks computation with COOCS

Procedures that select repeated segments in a corpus cannot yet track repetitions that are slightly altered by minor lexical modifications of one of the components. Given a pivotal word form, several methods can be used to select the set of word forms that tend to often appear in the neighbourhood of this word. In order to select these words, a unit of context or neighbourhood must be chosen, within which two words are considered to be co-occurring. For example, this unit can be similar to a sentence, as in Lebart et al. (1997).

Pivot	Context vectors
RU <i>синдром</i> Freq=43	→ гриппоподобный → озноб → лихорадка
	→ джонсона → стивенса
	→ токсический
	→ болевой
	→ острый → респираторный → дистресс
	→ лайелла
	→ пилула
FR <i>syndrome</i> Freq=145	→ grippal → pseudo
	→ fièvre → frissons
	→ johnson → stevens → lyell
	→ johnson → stevens → érythème → multiforme
	→ respiratoire → détresse
EN <i>syndrome</i> Freq=117	→ johnson → stevens → multiforme → erythema → epidermal
	→ necrolysis → toxic
	→ flu → chills
	→ flu → like
	→ respiratory → distress
	→ lupus
	→ fever

Figure 6. Lexical networks around the pivot *синдром/syndrome/syndrome* showing corresponding context vectors (extract).

Martinez (2005) proposed an original method of defining the lexical universe of a given pivotal word based on iterative calculation of lexical attractions: *multiple co-occurrences* networks. In our study, we used the tool COOCS resulting from his work, in an effort, for each pivotal word (anchor point), to find a network of words that are positioned preferentially in the same sentences. A detailed presentation of this research method for the identification of translation correspondences can be found in Martinez and Zimina (2002).

For comparable texts, the statistical study of the intensity of lexical relations through collocation allowed to build context vectors and map terminological equivalents within the neighbourhood of corresponding lexical anchors (see Figure 6).

7. Choosing terms and domains:
Collaboration domain expert/corpus linguist

Two types of human knowledge were necessary to succeed in creating our terminological resource: methodological knowledge on text processing from corpus linguists and domain-specific knowledge on adverse drug reactions from domain experts. The details of this collaboration for terminology management process are presented in Figure 7.

<i>Task</i>	<i>Tools, Methods and Resources</i>
Automatic segmentation into textual units (wordforms, repeated segments)	Lexico 3 [Lamalle <i>et al.</i> , 2004]
Identification of trilingual lexical anchors (starting points)	Frequency counts, cognates, general language correspondences, existing bilingual lexicons
Computation of trilingual collocation networks	COOCS [Martinez, 2005]
Identification of similar context vectors	Domain expert / corpus linguist
Semi-automatic segmentation into terminological units	Corpus linguist
Cross-language check	Corpus linguist
Validation of trilingual terminological records	Domain expert
Attribution of domains (organ classes)	Domain expert / corpus linguist
Final resource validation	Domain expert

Figure 7. Management of trilingual ADR terminology creation

Corpus linguists provide methodological knowledge on tools and methods for text exploration as well as quantitative results on corpora. The role of the domain expert was to choose and validate relevant terms in case of several variants attested in texts (see Figure 8). We also used WHO-ART terminology to check English-French term correspondences.

<i>Terminological variants in the PERTOMed corpus</i>		
RU	гриппоподобный синдром (13 occurrences)	☒
	гриппоподобные симптомы (2 occurrences)	☐
	симптомы гриппоподобного синдрома (1 occurrence)	☐
	гриппоподобная симптоматика (1 occurrence)	☐
FR	syndrome pseudo-grippal (23 occurrences)	☒
	syndrome pseudogrippal (2 occurrences)	☐
	symptômes pseudo-grippaux (7 occurrences)	☐
EN	influenza-like symptoms (9 occurrences)	☒
	influenza-like illness (6 occurrences)	☐
	flu-like symptoms (8 occurrences)	☐
	flu-like symptom (4 occurrences)	☒
	flu-like syndrome (6 occurrences)	☐
	flu-like illness (5 occurrences)	☐

Figure 8. Choosing terms through collaboration domain expert/corpus linguist

Simple term:	Simple term:
<i>диабет</i> <i>diabète</i> <i>diabetes</i>	<i>кома</i> <i>coma</i> <i>coma</i>
Context:	Context:
<i>диабетическая кома</i> <i>coma diabétique</i> <i>diabetic coma</i>	<i>диабетическая кома</i> <i>coma diabétique</i> <i>diabetic coma</i>
Domain:	Domain:
<i>Обмен веществ</i> <i>Troubles du métabolisme</i> <i>Metabolism disorders</i>	<i>Обмен веществ</i> <i>Troubles du métabolisme</i> <i>Metabolism disorders</i>
Medical product:	Medical product:
PJIC: <i>OLANZAPINE</i>	PJIC: <i>OLANZAPINE</i>

Figure 9. Trilingual terminological entries with complex term co-occurrence.

The role of the domain expert was essential to attribute a particular organ class (domain) for each terminological record (see Figure 9).

8. Results: Russian-French-English lexicon of adverse reaction terms

8.1 Qualities

A Russian-French-English ADR terminological resource resulting from the project is freely available on the PERTOMed website: <http://www.pertomed.info/>

This base lexicon comprises 430 validated trilingual terminological entries in XML format. The entries are structured by 2002 simple terminological records (single word terms) and 1006 complex terminological records (multi-word terms). Accordingly, approximately 50% of the records are complex terms. Each trilingual entry comprises the following fields (see Figure 9):

- Simple term (with possible variants).
- Abbreviation (if applicable).
- Related composed term(s).
- Domain(s).
- Drug product(s) concerned.

As shown in Figure 9, the structure of the lexicon preserves co-occurrence relations between terms. For example, the terminological record *diabetic coma* is listed under two simple main terms *coma* and *diabetes*.

8.2 Limits

For the moment, the trilingual lexicon is a structured list of Russian-French-English terms. It lacks visual aids for navigation and text look-up facilities. This work remains to be done, following experiments described in Zimina (2004).

We faced serious difficulties concerning the evaluation of Russian-French-English terminology. The choice of criteria was not straightforward. On the one hand, we had to keep in mind the existence of well-established international terminological standards for English and French, on the other, it was important to preserve the specificity of original Russian terms coming from authentic medical texts. Further steps should be taken in this direction in order to find a balanced approach of this complex methodological issue.

9. Conclusion

Creating terminological resources from comparable corpora is directly linked to the intrinsic heterogeneity of texts. Following our experiments within the PERTOMed project, we are convinced that the challenge of exploring texts from different cultural and linguistic sources should be taken into account in the terminology project feasibility study.

We hope that our exploratory work on creating a Russian Internet corpus in the field of pharmacovigilance will be followed by new research initiatives in order to collect and explore authentic terminological resources in Russian.

The use of textometric browsing for comparable text processing and term extraction gives encouraging results. Suggested methods for Russian corpus exploration should be improved, taking into account the availability of new resources for processing Russian medical texts.

References

- Antibiotics and Antimicrobial Therapy: <http://antibiotic.ru/index.php?newlang=eng>
- Baneyx, A., Charlet, J. and Jaulent M.-C. (2007). Building an ontology of pulmonary diseases with natural language processing tools using textual corpora. *International Journal of Medical Informatics* 76, 208–215.
- Bourigault, D. and Fabre, C. (2000). Approche linguistique pour l'analyse syntaxique de corpus. *Cahiers de Grammaire* 25, 131–151. Toulouse: Université Toulouse le Mirail. <http://w3.erss.univ-tlse2.fr/textes/framesetpublications.html>

- Charlet, J., Jaulent, M.-C., Slodzian, M., Bourigault, D., Baneyx, A., Bousquet, C., Mille, F., Ozdowska, S. and Zimina, M. (2006). Pertomed: Production et évaluation de ressources terminologiques et ontologiques dans le domaine de la médecine. *Rapport final*. INSERM U729.
- Chiao, Y.-Ch. and Zweigenbaum, P. (2002). Looking for candidate translational equivalents in specialized, comparable corpora. In *Proceedings of COLING'02*: 1208–1212. Taipei. <http://www-test.biomath.jussieu.fr/~pz/FTPapiers/Chiao:COLING2002.pdf>
- European Medicines Agency: <http://www.emea.europa.eu>
- Fung, P. and McKeown, K. (1997). Finding terminology translations from non-parallel corpora. In *Proceedings of the 5th Annual Workshop on Very Large Corpora*: 192–202. Hong Kong. <http://www.ece.ust.hk/~pascal/Publications/conference/1997/WVLC-1997.pdf>
- Interregional Association for Clinical Microbiology and Antimicrobial Chemotherapy: <http://www.iacmac.ru/iacmac/en>
- Ivanova, J. and Nuk, I. (2005). Création d'une terminologie français/russe dans le domaine de la pharmacovigilance. *Mémoire de DESS*. CRIM-INaLCO, Paris.
- Lamalle, C., Martinez, W., Fleury, S., Salem, A., Fracchiolla, B., Kuncova, A., Lande, B., Maisondieu, A. and Poirot-Zimina, M. (2004). *Lexico3 Textometric toolbox User's manual*. SYLED-CLA2T, Université de la Sorbonne nouvelle – Paris 3. <http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/manuelsL3/L3-usermanual.pdf>
- Lebart, L., Salem, A. and Berry, L. (1997). *Exploring Textual Data*. Boston: Kluwer Academic Publishers.
- Martinez, W. (2005). COOCS – Outils lexicométriques pour l'analyse des cooccurrences. *Manuel d'utilisation*. SYLED-CLA2T, Université de la Sorbonne nouvelle – Paris 3. <http://www.cavi.univ-paris3.fr/ilpga/individus/martinez/>
- Martinez, W. and Zimina, M. (2002). Utilisation de la méthode des cooccurrences pour l'alignement des mots de textes bilingues. In *Actes des JADT'02*, 495–506. Saint-Malo. http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2002/PDF-2002/martinez_zimina.pdf
- McEnery, A. and Xiao, R. Z. (2005). Parallel and comparable corpora: What are they up to? In *Incorporating Corpora: Translation and the Linguist*. *Translating Europe*, G. James and G. Anderman (eds.), Clevedon: Multilingual Matters. http://eprints.lancs.ac.uk/59/01/corpora_and_translation.pdf
- Morin, E. and Daille, B. (2004). Extraction de terminologies bilingues à partir de corpus comparables d'un domaine spécialisé. *Traitement Automatique des Langues* 45(3), 103–122.
- Ozdowska, S. (2004). Identifying correspondences between words: an approach based on a bilingual analysis of French/English parallel corpora. In *Proceedings of the COLING'04 Workshop on Multilingual Linguistic Resources*: 55–62. Geneva. <http://www.computing.dcu.ie/~sozdowska/papers/ozdowska-coling04.pdf>
- Ozdowska, S. (2005). Using bilingual dependencies to align words in English/French parallel corpora. In *Proceedings of the ACL Student Research Workshop*: 127–132. Ann Arbor (USA). <http://acl.ldc.upenn.edu/P/P05/P05-2002.pdf>
- PERTOMed Project <http://www.pertomed.info/>
- PJIC <http://www.rlsnet.ru>
- RECIPE <http://www.recipe.ru>
- Russian Vidal <http://www.vidal.ru>
- Sadat, F., Déjean, H. and Gaussier, É. (2002). A Combination of Models for Bilingual Lexicon Extraction from Comparable Corpora. *Papillon 2002 Seminar*. Tokyo. <http://www.papillon-dictionary.org>
- Uppsala Monitoring Centre <http://www.who-umc.org/>

- Véronis, J. (ed.). (2000). *Parallel Text Processing: Alignment and use of translation corpora*. Dordrecht: Kluwer Academic Publishers.
- WHO (1972). International drug monitoring: The Role of National Centres. *Technical Report No 498*. Geneva: World Health Organization. <http://www.who-umc.org/graphics/9277.pdf>
- WHO (2002) The importance of pharmacovigilance. *Safety monitoring of medicinal products*. Geneva: World Health Organization. <http://whqlibdoc.who.int/hq/2002/a75646.pdf>
- World Health Organization Regional Office for Europe <http://www.euro.who.int/>
- Zimina, M. (2005) Bi-text Topography and Quantitative Approaches of Parallel Text Processing. In *Proceedings from the Corpus Linguistics Conference Series 1(1)*. Birmingham. <http://www.corpus.bham.ac.uk/PCLC/>

Instrumentality in cognitive concept modelling

Paul Sambre and Cornelia Wermuth

The present investigation addresses instrumentality in multidimensional terminological definitions with a double aim. The first aim is to elaborate a cognitive framework of instrumentality in LSP. The second aim is to outline a usage-based linguistic typology of instrumentality. The study is based on a mixed methodology in which instrumentality is analyzed both on the conceptual level and the predication level using corpus data. So far, instrumentality has merely been investigated in an indirect, not corpus-based manner so that this semantic role, more specifically its multilayered character, is underspecified. In order to provide a definition of instrumentality which incorporates instrumental subtyping and its salience, thus allowing for an extended view of instrumentality and so-called associative relations, we investigated a German-French multidisciplinary corpus composed of abstracts in the field of microsurgery and cardiosurgery. Our results have led to some recommendations for existing ontological knowledge management tools like i-Term and i-Model.

Keywords: medical ontology, abstracts, medicine, instrumental role, associative relations, cognitive grammar, causality, corpus linguistics, French-German, knowledge management

1. From concept modelling to instrumentality

Our contribution combines two layers of analysis: the domain-specific ontological concept modelling used in expert communication and the linguistic or textual resources conveying terminological knowledge in different languages (Rothkegel, 2005: 84). More specifically, our objective is to provide a usage-based cognitive framework and typology of instrumentality in a French and German medical corpus.

A recent tool for ontology design like the Danish terminology and knowledge management tool i-Term, contains a state-of-the-art module, called i-Model, for the graphic representation of concept systems or ontologies (<http://www.i-term.dk/>, 02 February 2007). This tool stresses the traditional generic terminological

relations, as a series of linked superordinate and subordinate concepts on the one hand, and part-whole relations (also called meronymic or partitive relations) on the other. In i-Model, the non-generic and non-partitive relations are clustered in one category called 'other' relations. In this paper we will show a possible contribution of cognitive linguistics in building a more fine-grained typology of instrumentality as one of those 'other' relations.

2. A theoretical cognitive account of instrumentality

Cognitive linguistics (CL) is a linguistic theory which analyzes language in its relation to general cognitive abilities. CL claims that grammar, as a continuum reaching from syntax, over morphology and the lexicon, is equal to conceptualization (Ungerer & Schmid, 2006²). Language(s) cannot therefore be considered (an) autonomous module(s) of the mind: linguistic structure is clearly linked to other mental faculties as categorization and (abstract) knowledge (Croft & Cruse, 2004). A gestalt-like grammar reflects linguistic traces of non-linguistic perception (Dirven, 2005). CL offers a key to linking the linguistic structure and the ontological (medical) expert knowledge level in instrumentality. Since instrumentality is poorly defined in terminological principles (2.1), we will introduce it in an existing causal model (2.2).

2.1 Terminology and instrumentality

Instrumentality is a poorly analyzed topic both in linguistics and in terminology. Linguistic research has focused on the instrumental case (in Slavic languages, particularly in Polish, see Tabakowska 2002), which linguistically marks the referential or nominal means leading to the accomplishment of the action expressed by a clause. In languages without instrumental case, instrumentality is neither linked to a single grammatical category nor to a syntactical position.

ISO norm 704 states that the equipment or tool – what we call the instrumental role – is an associative relation linked to causality:

Some associative relations exist when dependence is established between concepts with respect to their proximity in space or time. These relations may involve raw material – product, action – equipment/tool, quantity – unit, material – property, material – state, matter/substance – property, concrete item – material, concrete item – shape, action – target, action – place/location, action – actor, etc. Some relations involve events in time such as a process dependent on time or sequence; others relate cause and effect. (ISO, 2000: 16)

Work in lexicology, terminology and standardization has focused mainly on hierarchic and meronymic relations, the latter being considered a hierarchical relation by some authors (for a criticism see Sambre, 2005). In addition, meronymy is not always recognized as a lexical relation by lexicologists. On the other hand, less attention has been given to the role of associative relations, a theoretical point reflected in concept modelling tools like the previously mentioned i-Model. Despite Picht's (1997) insight that meaning-form patterns are to be included in terminological descriptions, little or no attention has been paid to those patterns regarding associative relations in general and instrumentality in particular. Terminology is still influenced by Wüster's neo-positivist goal of monosemy with monosymy (Rogers, 2005: 1852).

Since it is the spreading of instrumentality as a semantic role over different linguistic, viz. syntactical, grammatical or lexical mechanisms which is at stake here, the following section explores how this heterogeneity can be accounted for by CL as part of a causal chain.

2.2 Instrumentality in CL

We propose an abstract conceptual scheme for instrumentality (2.3.1). This conceptual scheme is then linked to different types of linguistic predications (2.3.2).

2.2.1 *Instrumentality and the causal chain*

We combine two major CL strands in order to capture the role of instrumentality: Cognitive Grammar (Langacker, 1987, 1991) and force dynamics (Talmy, 2000). According to Talmy (2000: 421–424), the semantic structure of a causal chain is typically composed of a causing and a caused event, with varying degrees of explicitness and agency expressed in their linguistic manifestation:

All of the interrelated factors in any force-dynamic pattern are necessarily co-present wherever that pattern is involved. But a sentence expressing that pattern can pick out different subsets of the factors for explicit reference – leaving the remainder unmentioned – and to these factors it can assign different syntactic roles with alternative constructions. (Talmy, 2000: 422)

This idea of alternative foregrounding (and corresponding backgrounding) of semantic information in linguistic predications is a key notion to understand different form-meaning pairings linked to instrumentality. The Figure in the causing event is the Instrument in the encompassing causal situation. This typically happens to the nominal in the with-phrase of an agentive sentence: *John slices the salami with a knife*, with a corresponding non-agentive predication *The knife sliced the salami*, in which the underlying causative scheme could be paraphrased in the

following way: *the salami falling into slices* (caused event) *resulted from* (causal relation between two events) *John's cutting the salami with a knife* (causing event).

In Cognitive Grammar, events are structured as action chains, which trace the flow of energy from a source to an energy sink along a mental path (Langacker, 1991: 292) with (human) participants and moveable physical objects, construed in a particular scene (Langacker, 1991: 296). Clauses code (parts of) this (causal) scene by means of syntactical configurations and (sub)lexical data (Langacker, 1991: 405). Traditional case-marking on a noun

evokes in schematic terms the conception of a relationship in which the nominal referent participates; that relationship is subsequently rendered specific in a higher-order construction taking the case-marked nominal as one component structure (Langacker, 1991: 404).

This happens in the example *John cuts the salami with a knife*. The predication profiles the instrumental role rendered by the instrumental nominal with a preposition *with* (schematic instrumentality) and the associated noun *knife* (elaborated instrumentality) on the one hand, and an agentive process on the other: the agent *John*, the verb *cut* and the patient *salami*. Like Langacker, who focuses on the fact that not all instrumental cases are linguistically explicit in the predication (as in *John cut the salami in slices*), we note that instrumentality needn't be expressed by a schematic preposition. Apart from this discussion, which we broaden in the descriptive section, the combination of Langacker and Talmy leads us to an abstract conceptual scheme (Figure 1), in which instrumentality is embedded in a causal chain of medical therapy.

In this temporal process (t standing for time) a medical patient (MP) is diagnosed with a disease. This initial state (St1) should lead to a final state (St2) of recovery after a therapeutic treatment. This treatment we call, in line with Talmy, the causal chain of therapy, with a causing and caused process. In the causing process, the medical doctor (MD) intentionally uses an instrument on the MP. As a result,

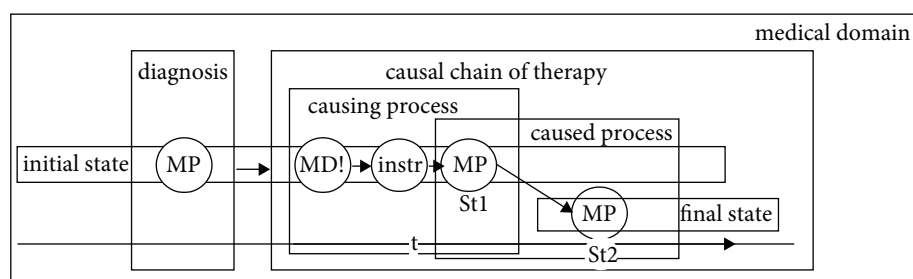


Figure 1. Instrumentality in the causal chain of therapy

the MP moves to a final, different state. This conceptual scheme makes abstraction of the specific linguistic predication used for the (partial) expression of the causal chain. In the following section, we link this conceptual layer to the linguistic one.

2.2.2 *Conceptual entities*

Instruments are typically nouns filling slots with frame elements around support verbs (Fillmore et al., 2001, 2004). We extend this nominal layer of instrumentality to other word classes. Langacker (1987: 249) proposes a typology of three basic classes of predications to which correspond different word classes. Conceptual entities are typically things (nouns), atemporal relations (adjectives and prepositions) or processes (verbs). Processes display atemporal relations with a positive temporal profile. Our basic hypothesis is that instrumentality needn't be expressed exclusively by atemporal prepositions.

3. Methodology

On a conceptual level, we theoretically have three types of entities at our disposal. We combine this top-down approach with a bottom-up corpus approach, in which entities are analyzed with respect to the level of linguistic predication: nouns, adjectives and adverbs, prepositions and verbs. This particular combination of corpus and conceptual analysis we call cognitive corpus linguistics. The following briefly sketches the layout of our corpus (3.1) and our analytical methodology (3.2).

3.1 Corpus

Our dataset is based on a text library of 80 titles and abstracts of scientific medical articles. For this paper, we concentrate on two contrasting languages and two medical disciplines: French and German, cardiology and microsurgery. The abstracts have been broken down into 947 sentences, in which the explicit linguistic instrumentality was annotated. In this paper, we focus on the highest syntactical level of instrumentality, which we label central instrumentality. We illustrate this by means of the following example.

- (1) A **catheter** treatment was used successfully.

The example contains different instrumental predications, on different levels of syntactical analysis. Most centrally, the passive verb *was used* triggers the subject *catheter treatment* as the instrumental noun. But this nominal compound itself is again composed of a head *treatment* and modifier *catheter*. The latter is again a

nominal, since the catheter is the instrument the treatment is performed with. Hierarchically speaking, *use* is more central to the sentence than the modifier of the nominal compound. We analyzed 40 abstracts (10 for every discipline in both languages), and decomposed all of their sentences in 844 instrumental predications, of which 487 are central. These 487 central instrumental predications constitute the corpus of the current paper. The text library consists of: French *Annales de Cardiologie et d'Angéiologie* 53 (2004), 54 (2005), 55 (2006), 55; *Réanimation*, 12 (2003), 13 (2004) 14 (2005), 15 (2006); *Chirurgie de la main* 24 (2005), 25 (2006); German *Zeitschrift für Kardiologie* 94 (2005); *Handchirurgie – Mikrochirurgie – Plastische Chirurgie* 34 (2002), 35 (2003), 36 (2004), 37 (2005), 38 (2006).

3.2 Method

Text library, sub-disciplines, languages, sentences and predications are integrated into a relational Access-database. This database not only contains the corpus data, but also the cognitive parameters used for the conceptual coding of the instrumental predications. The relation between top-down conceptual breakdown and bottom-up corpus feedback in cognitive corpus analysis is shown in Figure 2, the abstract labels are in black (I for instrumental) and ATEMP meaning atemporal.

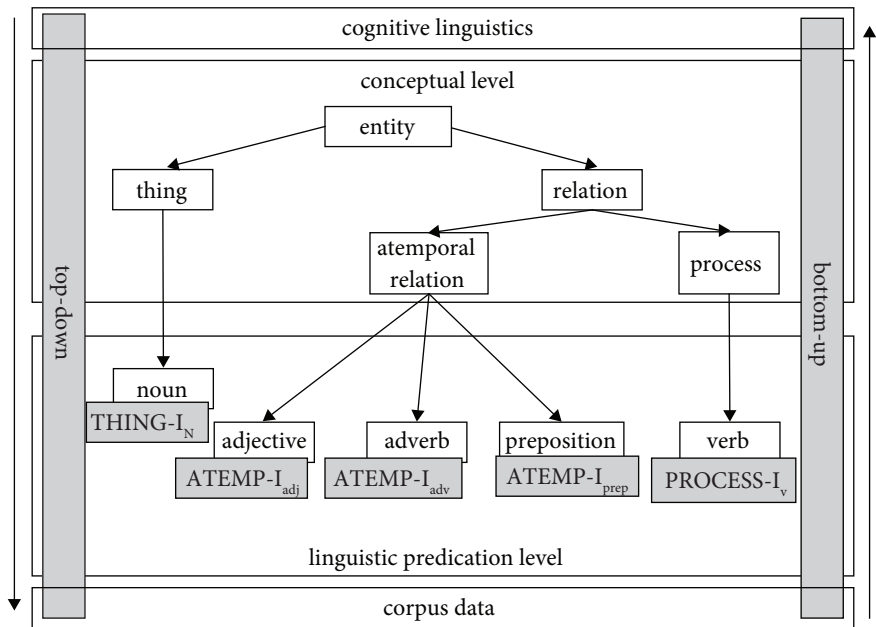


Figure 2. Cognitive corpus linguistics for instrumentality

As opposed to traditional collocational analysis, frequently used in corpus linguistics (Baker, 2006) on reference corpora, and to large-scale corpus-driven research, the method of cognitive corpus linguistics concentrates on detailed conceptual description of small-scale corpora (Sambre, 2005) in LSP.

The corpus and methodological breakdown used here are relatively new in terminology and have five advantages. First, this methodology offers a more authentic view of terminological data in natural language; cognitive linguistics is usage-based (Barlow and Kemmer, 2001). Second, it shows particular terminological features in particular textual, discursive (Gerzymisch-Arbogast, 1996) or domain-specific (Temmerman, 2000) genres. Third, it allows a contrastive multilingual approach to form-meaning patterns (for a case study on the sublanguage of cardiosurgery see Wermuth, 2005). Fourth, it brings in other perspectives (*Merkmale* in Picht, 1997; ways-of-seeing in Croft & Cruse, 2004) than traditional generic and partitive constellations. Fifth, it allows quantification: text genres display preferential salience representations of instrumental language patterns.

4. Descriptive results

The results of our research are centered on two basic elements: an instrumental typology (4.1) and the salience of each instrumental subtype (4.2).

4.1 Instrumental typology

In accordance with Langacker's classification (cf. 2.2.2), the linguistic predications under investigation can be grouped into three conceptual categories, namely things (4.1.1), atemporal relations (adjectives and prepositions) (4.1.2) and processes (4.1.3). In the following sections, we describe and illustrate the subtyping of these main predication groups.

4.1.1 *Things*

First, instrumentality may be expressed as a thing, more specifically as a nominal predication (without support verb). According to Langacker (1987: 183) a 'thing' is to be defined in an abstract way, because it refers to *cognitive* events (and not to physical objects). In our corpus, a language-independent feature of this nominal type of instrumentality is that of depersonalization: the intentional actor (i.e. the physician who performs the procedure) is left implicit so that the action and its related aspects like instrumentality are in the foreground (Busch-Lauer, 2003: 192). Within this predication type, we distinguish two patterns with different degrees

of complexity. The basic pattern typically provides the name of a treatment or surgical technique, as in the following corpus samples.

French:

- (2) déterminer le niveau de prise en charge pour chacune des prescriptions suivantes: **statines**, **aspirine** ou **autres antiplaquettaires/antithrombotiques**, **bêtabloquants** et **inhibiteurs de l'enzyme de conversion (IEC)**
defining the coverage of the following prescriptions: statines, aspirin or other antiplatelet/antithrombotic beta-blockers and IEC inhibitors
- (3) *la coronarographie est la **technique de référence***
coronarography is the preferred technique
- (4) *le **scanner spiralé multibarrettes***
multi-array spiral scanner
- (5) *les items **techniques du geste de revascularisation***
technical revascularization items
- (6) *la **resynchronisation cardiaque***
cardiac synchronization

German:

- (7) *eine erneute **Intervention***
a second intervention
- (8) ***Nachuntersuchung***
follow-up examination
- (9) ***Erstimplantation***
first implantation
- (10) ***Bosentan** (Tracleer®) zur **Behandlung** der Hypertension*
Bosentan (Tracleer®) for the treatment of hypertension
- (11) ***Thorakotomie***
thoracotomy

Basic nominal predications frequently display nouns representing therapeutical substances. For example, the above mentioned French expression (2) *Déterminer le niveau de prise en charge* is completed by a set of nouns referring to substances which imply instrumentality, for example the words in bold in (2). The same holds for (10) where the substance Bosentan is used as a 'means' in order to reduce hypertension. In cognitive terms, the full nominal expressions are composite semantic structures: their meaning includes the relation of the separate components to the composite whole, in which specific components are profiled (cf. Langacker, 1989: 292) as instrumentals. Still in (10), the profile focuses on the substance by means of which the causative activity *Behandlung* is performed. Semantically

spoken, the noun *Bosentan* is also part of the causing action in which an implicit agent uses the substance in order to cause an effect on a given medical condition (namely hypertension). From a syntactical viewpoint, the instrumental noun frequently takes subject position which profiles its instrumental meaning relatively to the other components of the depersonalized expression. This subject position is, however, not mandatory, as in the French Example (12) where the instrument is realized as prepositional object.

- (12) *le but de cet article est de faire une mise au point sur l'état actuel des connaissances **sur ces techniques non invasives***
the objective of this article is to clarify the actual knowledge of these non-invasive techniques

Next to the fact that one and the same component may display different potential semantic 'roles', we notice that instrumentality is both explicitly and implicitly realized. This is e.g. the case in Examples (3) and (10). Here, in the two nouns *coronarographie/ technique de référence* and *Bosentan/Behandlung* the second noun expresses instrumentality in general, whereas the first one is more explicit than the second. The German compound

- (13) *Kathetertherapie*
 catheter therapy,

which means 'a therapy which is conducted by means of a catheter', the instrumentality is rendered by the noun modifier *Katheter* which, though implicitly, refers to the instrument used to perform the therapy.

Within the subcategory of double instrumentality, an interesting subtype displays two different kinds of instrumentality, as in the French predication

- (14) *les techniques d'extractions percutanées de stimulateurs cardiaques*
 percutaneous extraction techniques of cardiac stimulators.

In (14), the noun *les techniques d'extractions percutanées* has, albeit indirectly, an instrumental function referring to the action performed in order to remove implanted cardiac stimulation tools. Those tools, in their turn, have an instrumental role in the predication as they are used as (appositive) direct instruments or artefacts by means of which the heart rate is increased.

The occurrence of 'negative', i.e. unwished instrumentality associated with the instrumental substance has also been attested, like in the French Example (15).

- (15) *le **lithium** est connu pour être responsable de nombreux **effets indésirables** sur le système cardiovasculaire*
 lithium is well-known for its undesired effects on the cardiac system

With respect to the linguistic realization of the predication thing-type, some language-specific patterns can be identified. In German, the prototypical nominal predication consists of a noun-noun-compound like

- (16) *Katheterinversion*
catheter inversion,

a postmodified deverbal noun like

- (17) *die intravenöse Gabe von insgesamt 2×190 mg Digitalis-Antidot*
the intravenous administration of a total dose 2×190 mg digitalis antidote,

or a prepositional phrase like

- (18) *Behandlung mit Bosentan*
treatment with Bosentan.

As one would expect, French makes frequent use of analytical prepositional phrases, as in

- (19) *l'imagerie par résonance magnétique*
Magnetic Resonance Imaging

and

- (20) *les arthroplasties d'interposition par anchois en Dacron® dans le traitement des rhizarthroses essentielles*
interpositional anchovy Dacron arthroplasty in the treatment of essential rhizarthrosis.

In those occasionally complex nominal predications, the instrumentality expressed by the technique is most frequently specified by noun modifiers. An illustrative example is the French predication

- (21) *deux tentatives de sclérothérapie, l'une à l'alcool pur, l'autre à l'Éthibloc®*
two attempts at sclerotherapy with absolute alcohol and Ethibloc®,

where the modifiers *alcohol* and *Ethibloc* are therapeutical substances specifying the instrumental technique *sclérothérapie*. In the German corpus, we also find prepositional instrumentality in nominal predications. An example is the PP *mit Sirolimus-Eluting-Stents* in the German predication

- (22) *Behandlung einer In-Stent-Restenose mit Sirolimus-Eluting-Stents – eine 6-monatige klinische und angiographische Nachbehandlung*
treatment of an in-stent-restenosis with sirolimus-eluting stents – a six month clinical and angiographical follow-up.

This recurrent construction type prototypically consists of an artefact (e.g. *Sirolimus-Eluting-Stents*) in combination with a nominalized instrumental (e.g. *Behandlung*), the instrumentality of the artefact being determined by the deverbal noun.

4.1.2 Atemporal relations

According to Langacker (1987: 214f.), linguistic predications are either nominal or relational. Nominal predications describe a thing (i.e. a reified event), relational predications describe either a process or an atemporal relation. In our corpus, both types of relational predications are used in order to express instrumentality. We first focus on the predications with atemporal relations: prepositions (A.) versus adjectives and adverbs (B.).

A. Prepositions

In our corpus, instrumentality is frequently realized by means of a prepositional phrase. Obviously, in both languages under investigation, the preposition is followed by a noun, but the instrumentality is triggered by the meaning of the preposition in the first place, which transfers the instrumental role to the noun in the phrase. We identified three different prepositional subtypes, namely ‘mittels’/ ‘par’ (*by means of*), ‘avec’/ ‘mit’ (*with*), and ‘par’/ ‘anhand’ (*by*, which can be considered a specification of the first one).

German:

- (23) ***mittels** kardialer Magnet-Resonanz-Tomographie*
by means of magnetic resonance tomography
- (24) *der Stenosegrad wurde **mittels** QCA bestimmt*
stenosis severity was determined with QCA
- (25) *Therapieversuche **mit** Phenytoin und Isoprenalin-Infusion*
therapeutic trials of phenytoin and isoprenalin infusion
- (26) *etwas **anhand** der derzeitigen Datenlage klinischer Studien beschreiben*
to describe something by means of actual clinical trial data

French:

- (27) *un visage **avec** agrafe*
screw connection with clip
- (28) ***par** scie circulaire du majeur*
with a circular saw of the middle finger
- (29) *les allongements digitaux **par** distraction progressive*
finger lengthening by progressive distraction
- (30) *traitement antiagrégant **par** aspirine et clopidogrel*
antiaggregant treatment with aspirin and clopidogrel

As those examples show, the preposition may be followed by nouns representing various conceptual categories like techniques (*Magnet-Resonanz-Tomographie, agrafe, distraction*), artefacts (*scie*), abstract entities (*Datenlage klinischer Studien*) or therapeutical substances (*Phenytoin, aspirine, clopidogrel*).

It is worth mentioning that the prepositional type of instrumentality in neither language displays complex patterns. At least in French, this instrumental subtype may combine a positive instrumentality with a negative one, expressed by ‘*sans*’ (without) like in the French predication

- (31) *le brochage a été réalisé sous contrôle de l'écran d'ordinateur, sans fluoroscopie*
the drilling was performed under computer guidance without fluoroscopy

B. Adjectives and adverbs

In a cognitively inspired approach, adjectival and adverbial expressions represent the same category: both word classes link instruments to medical processes such as clinical investigations, treatments, scientific research or the like. Those processes may be expressed both as a processual predication expressed by a verb (cf. Section 3) or an atemporal non-relational process or thing (cf. Langacker, 1987: 247) that corresponds to adjectives and adverbs. The examples we identified concern relational adjectives derived from nouns on the one hand or deverbal adjectives on the other. Note that those adjectival and adverbial constructions remain quite close to patterns expressed by verbs and nouns. Other types are to be further examined. Two patterns can be distinguished, a basic one and a complex one. In the basic pattern (32)–(33),

- (32) (Ge) *medikamentöse* Kardioversion
drug cardioversion
(33) (Fr) *le traitement bêtabloquant*
beta-blocking treatment

the instrumental adjective selects a limited list of treatment nominals (*Kardioversion, traitement*). The adjective may refer to a technique like *medikamentös* in (32) or to a substance (*bêtabloquant* in (33)).

The following two German examples display more complex patterns: the adjective combines nominal instrumentality with a non-instrumental adjective in a compound structure.

- (34) *katheterbasierte* Septumablation
catheter-based septal ablation
(35) *katheterinterventioneller* Verschluss
catheter interventional occlusion

In those examples, the instrumental noun *catheter* is triggered by the derivational element ...-*basiert* (*based on*) and ...-*interventionell* (*interventional*).

Adverbial instrumentality only displays basic patterns. In the following examples, the adverb is a morphological derivation based on an adjective, but this is, however, not necessarily the case.

German:

- (36) **angiographisch** *untersuchen*
to investigate angiographically
- (37) **histologisch** *nachweisen*
to prove histologically

French:

- (38) *fractures traitées* **orthopédiquement**
orthopedically-treated fractures

As the corpus analysis showed, adverbs only occur in combination with full verbs and not with support verbs. The latter would include the instrumental in the adjectival part of the full nominal. The instrumental adverbs in the examples may refer to different aspects like the discipline involved (*orthopédiquement*), a surgical technique (*angiographisch*) or the type of investigation (*histologisch*).

4.1.3 Predicational type 'processes':

Instrumental processes express instrumentality by means of a verbal predication. According to Langacker (1987: 143), a process is a continuous sequence of similar or distinct states, typically expressed by a verb. A verbal predication incorporates one or more distinct states within this process. With respect to our corpus, we identified two basic types of verbal predications. The first type expresses instrumentality by means of a unique instrumental verb. The second type uses support verbs with nominal instruments. The following samples show pure, instrumental verbs:

German:

- (39) *Digitalis-Antidot ist* **anzuwenden**
digitalis antidote has to be used
- (40) *Calciumantagonisten sollten* **eingesetzt** *werden*
calcium antagonists should be used
- (41) *wurde erfolgreich ein Sirolimus Eluting Stent* **implantiert**
a sirolimus-eluting stent has been successfully implanted

French:

- (42) *des outils d'éducation thérapeutique ont été créés*
therapeutic training tools have been created
- (43) *le 5-fluorouracil, agent antimétabolique de base fluoropyrimidique, **est** largement **utilisé***
5-fluorouracil, an antimetabolic fluoropyrimidine-based agent is widely used
- (44) *le scanner **permet** donc de détecter des lésions coronaires significatives*
the scanner allows to detect significant coronary lesions

In the German examples, the instrumental thing triggered by the (passive voice) verb (*anwenden, einsetzen, implantieren*) is linked to the subject (*Digitalis-Antidot, Calciumantagonisten, Sirolimus Eluting Stent*). Hence, the instrument is in focus, and not the intentional actor, a semantic operation we already described above in the context of nominal predications. Instrumental subjects can also be found in the French corpus: *des outils d'éducation thérapeutique, le 5-fluorouracil, scanner*.

The second type of verbal predication is that of the so-called Support Verb Constructions (for an extensive discussion of those verbs, which are also labeled as light verbs or function verbs, see Pustejovsky 1988 and 1995). Those constructions contain semantically underspecified support verbs (*durchführen, réaliser, pratiquer*) and a predicative noun, either as a direct or prepositional complement. It is the instrumental noun which essentially provides the major part of the (instrumental) semantics¹. The main function of the verbs is to transform the (atemporal) nominal into a verbal predication (with a temporal profile), as the following examples show.

German:

- (45) *eine Herzkatheteruntersuchung **durchführen***
to perform a heart catheterization
- (46) *wurde eine aortokoronare Bypass-Operation mit Aneurysmektomie und Entfernung des Verschluss-Schirmchens **durchgeführt***
an aortocoronary bypass operation with aneurysmectomy and removal of the closing umbrella was performed

1. In 4.1.1, we described similar instrumental things in a nominal predication without support verb: their instrumentality is then included in a non-instrumental verbal predication. The main difference in the current section is that the support verb integrates the instrumental noun in a more global instrumental verbal predication.

French:

- (47) *un remplacement valvulaire aortique et un pontage aortocoronaire ont été réalisés*
the replacement of the aortic valve and an aortocoronary bypass have been performed
- (48) *un peu plus de 180 centres pratiquent aujourd'hui l'angioplastie en France*
more than 180 centres nowadays performe angioplasties in France

Constructions with support verbs can express different event structures with a large number of state- and process-relations (Vendler, 1967). In the given examples, the reading of the nouns as happenings (*eine Herzkatheteruntersuchung durchführen, un peu plus de 180 centres pratiquent aujourd'hui l'angioplastie en France*) and states (*wurde eine aortokoronare Bypass-Operation mit Aneurysmektomie und Entfernung des Verschluss-Schirmchens durchgeführt, un remplacement valvulaire aortique et un pontage aortocoronaire ont été réalisés*) is triggered by the tense and aspect of the support verb. Both types of construction yield full instrumental readings, the instrumentality being triggered by the nouns. The conceptual status of processes with and without support verbs is similar, but the structural compositionality they display in the imagery is different: pure instrumental verbs directly operate on direct or prepositional patient roles, whereas support verbs take an intermediate step, in which a thing is given a temporal profile.

4.2 Typological salience

The five types of instrumentality show different degrees of relative frequency or salience. We provide some absolute quantitative data based on the conceptual typology in Section 4.1. (N of central instrumental predication = 487). After an overall picture of instrumental subtypes in our corpus, we will add some reflections on two variational aspects of instrumentality. The general idea is that lexical onomasiological choices for instrumentality, and, consequently, their underlying conceptual entities are amenable to quantitative analysis (Grondelaers and Geeraerts, 1998: 372) and consequently display a clustering of grammatical devices around a central prototype (Langacker, 1991: 284).

The instrumental prototype encompasses all five subtypes. First of all, in line with our initial hypothesis, we observe that things are globally the most prominent instrumental subtypes (Figure 3).

The expected predominance of prepositions is not attested by our data. Processes are less dominant, but still an important subtype. Adjectives and adverbs, finally, are less important, but non-neglectable instrumental subtypes (Figure 3).

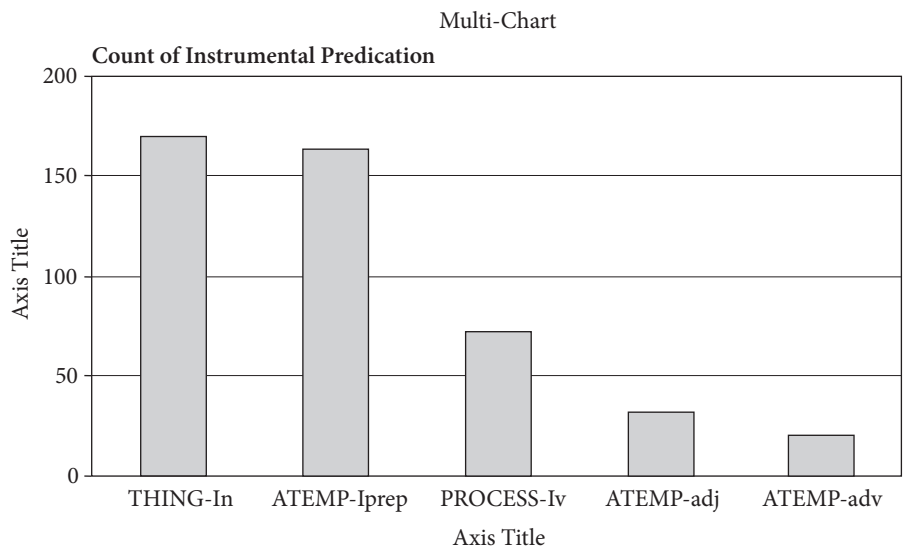


Figure 3. Salience of instrumental subtyping – Overall picture

As to the salience of instrumental subtypes according to the discipline, different salience patterns in cardiology vs. microsurgery can be identified (Figure 4). In the field of cardiology, things and atemporal prepositions are the most salient means to express instrumentality, followed by processes, adjectives and adverbs. In the field of microsurgery, however, prepositions are the most dominant, followed by things, processes, adjectives and adverbs (which both have a comparable salience). Things and atemporal prepositions are clearly the most frequent entities used for instrumentality, but they differ in relative weight in the two disciplines scrutinized.

Differences in salience in instrumental types have a linguistic bias. The prepositional instrumental type is much more dominant in German than in French (Figure 5). In French, the thing-instrumentality and process-instrumentality are most salient, whereas the cognitive literature (Langacker, 1991; Talmy, 2000) has emphasized the impact of instrumental prepositions, as a logical consequence of Fillmore’s case grammar.

Count of Instrumental Predication

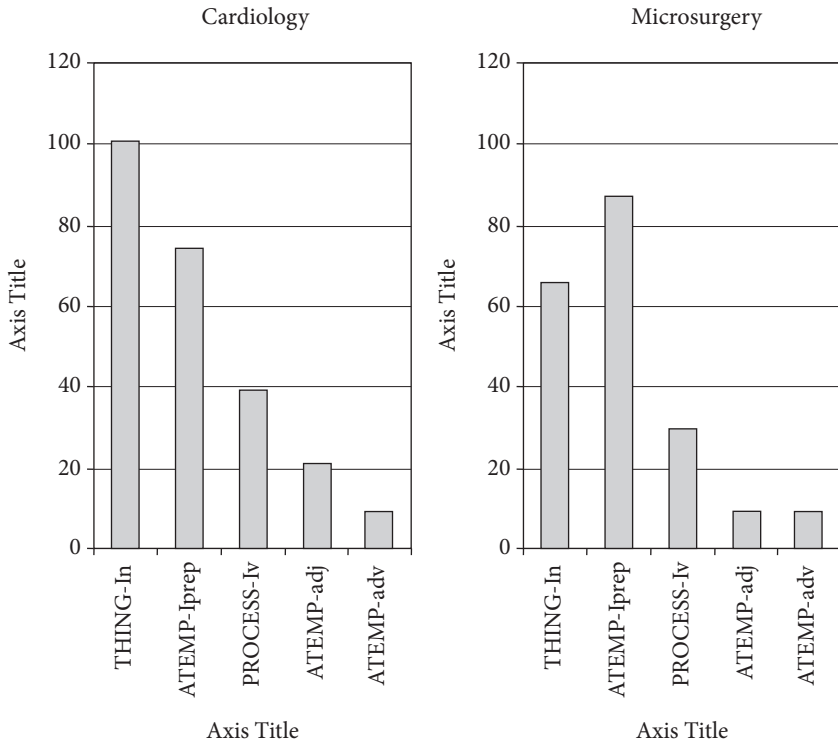


Figure 4. Salience of instrumental subtyping – Discipline variation

We assume that the focus on prepositions in the literature might be caused by English, viz. by the Germanic perspective in linguistic description of instrumentality.

Count of Instrumental Predication

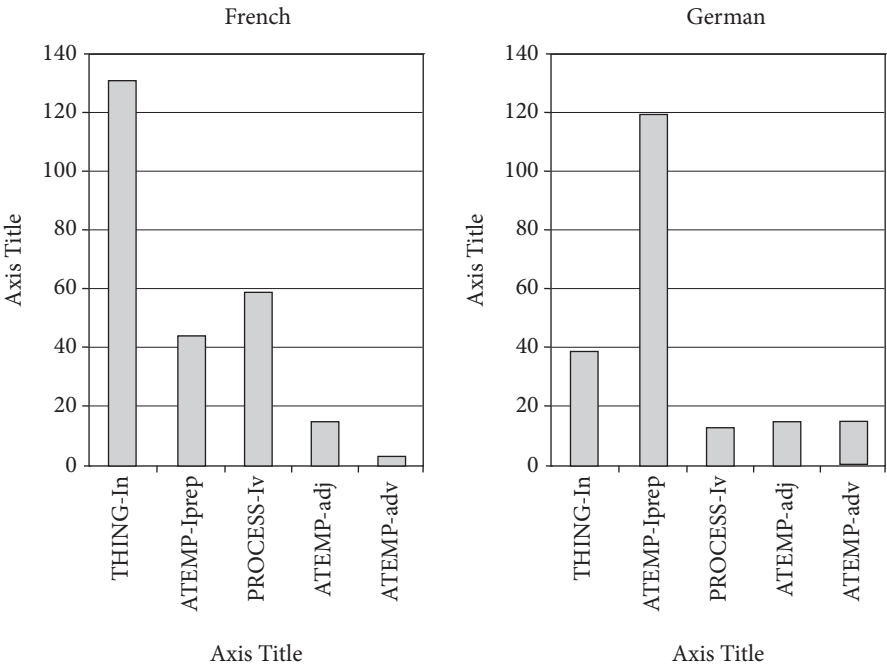


Figure 5. Salience of instrumental subtyping – Linguistic variation

5. Conclusion

In this contribution, we described the relation between the ontological and linguistic knowledge in multilingual medical texts (1.). We adopted a theoretical approach inspired by cognitive linguistics (2.). Based on this framework we adopted a conceptual corpus methodology (3.) which leads to a typology and salience patterns of instrumentality in a French and German sample (4.).

Summing up, we draw the following conclusions:

1. We have elaborated a usage-based framework for and an exploratory description of instrumentality which combines cognitive insights of Talmy and Langacker. The analysis relies on an authentic French and German corpus.
2. Medical abstracts display heterogeneous instrumental expressions. In this genre, instrumentality is a key terminological associative relation. More often than not, medical abstracts show what can be done with an instrument rather than its place in a medical nomenclature of techniques, artefacts and substances.

3. The syntactical, grammatical and lexical heterogeneity of instrumentality is not adequately accounted for by the linguistic and particularly by the terminological literature. In the terminological definition of semantic relations, instrumentality is part of the so-called associative relations and not described in greater detail. The focus on some particular (prepositional or case-bound) entities in the linguistic literature does not reflect the language-specific salience of instrumentality in languages as German and French.
4. In line with cognitive grammar, the description of instrumentality calls for a double layer of analysis: a conceptual versus a predicational one. Instrumentality is coded in terms of all Langackerian conceptual entities, and leads to different subtypes: things, atemporal adjectives, adverbs and prepositions, and, finally, processes.
5. Quantitatively, the linguistic structure of the conceptual coding of instrumentality displays patterns of salience with a language-specific and domain-specific bias.

In line with those conclusions, five theoretical and applied recommendations emerge with respect to the further development of terminology and (tools for) ontology design.

1. The latter two fields should strengthen their theoretical foundations. A cognitive theory as sketched out here for instrumentality provides a compatible and yet more sound semantic basis for those fields.
2. Terminological work should rely on the analysis of more and more authentic (medical) language data. Since terminology and ontology increasingly deal with (automatic retrieval in) large text corpora and data mining, this usage-based approach provides a more realistic account of what terminological relations, i.e. instrumentality, look like in everyday discursive practice.
3. Instead of *a priori* conceptual entities and relations, ontological and terminological descriptions of instrumentality should focus more on the singular and contrastive predicational linguistic forms conceptual entities display, particularly in multilingual (or translated) discourse. The terminological focus on prepositional instrumentality and noun is clearly rejected by our empirical results.
4. A further methodological step is the domain-specific instrumental subtyping. More generally, the domain-specific research on the salience of associative relations in their interaction with generic and partitive ones is an essential stage in the further development of terminology and ontology design. From the viewpoint of applied science, taking into account linguistic patterns and their probabilistic status in a large text database could facilitate (automatic) tagging of instrumental concepts in the context of a future (medical) Semantic Web.
5. Finally, in order to comply with the heterogeneity both of instrumentality and of terminological definitions in general, we recommend three measures for

terminological software tools like i-Term: (a) i-Term should integrate instrumentality as a central component in its concept modelling and (b) strengthen its associative module (the so-called ‘other relations’); (c) Terminology should allow for the co-presence of multiple, multidimensional definitions. Next to the traditional unidimensional hyperonymic definition, multidimensionality includes instrumental and other associative relations and provides a more realistic image of the concept defined. We illustrate this idea for *Bosentan* (Figure 6), which is part of a generic medical nomenclature of superordinate and subordinate medical substances, and is decomposed in a partitive chemical formula (Clozel et al., 1994). Those generic and partitive relations are traditionally dealt with by terminology. In our corpus, *Bosentan* is displayed not so much generically or partitively, but instrumentally, in a causal therapeutical chain of medical treatment. Multidimensional definitions ideally take into account the multilayered character of (medical) terms.

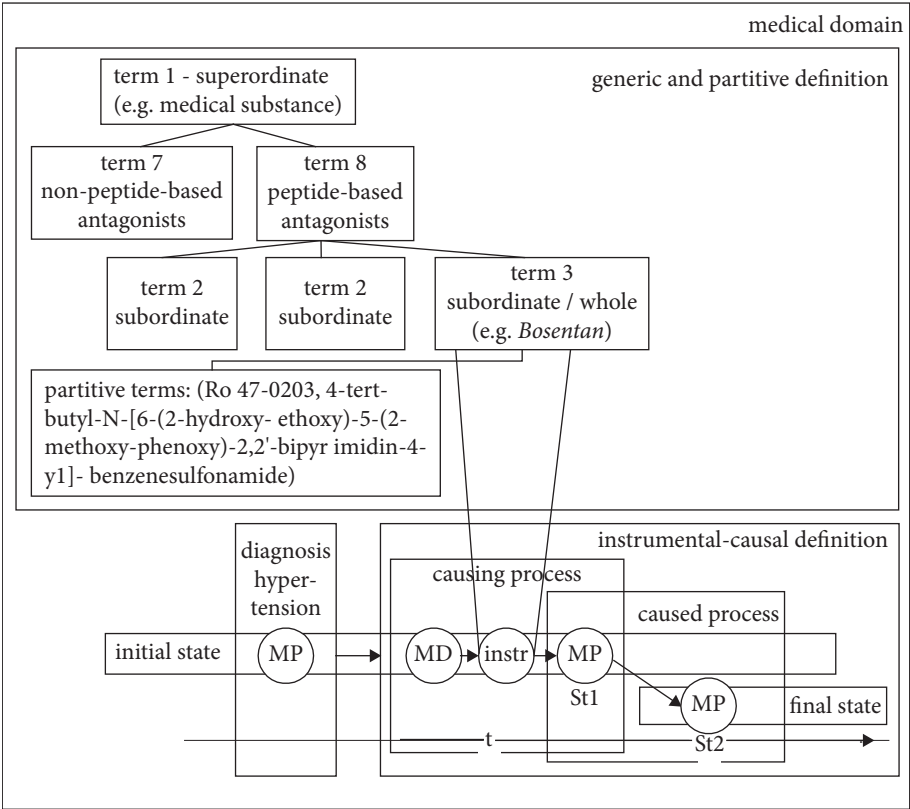


Figure 6. Multidimensional definition of *Bosentan*: generic, partitive and instrumental definition

References

- Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.
- Barlow, M., S. Kemmer (eds.). (2000). *Usage-Based Models of Language*. Stanford: CSLI.
- Busch-Lauer, I. (2003). Perspective in medical correspondence. [English and German letters-to-the-editor]. In: T. Ensink and Ch. Sauer (eds.), *Framing and Perspectivising in Discourse*, 191–214, Amsterdam / Philadelphia: Benjamins.
- Clozel, M. et al. (1994). *Pharmacological characterization of Bosentan, a new potent orally active nonpeptide endothelin receptor antagonist*. *Pharmacology*, vol. 270, Issue 1, 228–235.
- Croft, W. and D. A. Cruse. (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Dirven, R. (2005). Major strands in Cognitive Linguistics. In: F.J. Ruiz de Mendoza Ibáñez and M.S. Peña Cervel, *Cognitive Linguistics and Interdisciplinary Interaction* (17–68), Berlin / New York: Mouton de Gruyter.
- Fillmore, Ch. et al. (2001). *Building a Large Lexical Databank Which Provides Deep Semantics*. Unpublished paper EUROLAN 2001 Summer Institute.
- Fillmore, Ch. et al. (2004). *FrameNet and Representing the Link between Semantic and Syntactic Relations*. Institute of Computer Linguistics 2004.
- Gerzymisch-Arbogast, H. (1996). *Termini inm Kontext. Verfahren zur Erschließung und Übersetzung der textspezifischen Bedeutung von fachlichen Ausdrücken*. Tübingen: Narr.
- Gries, S. (2006). Introduction. In: S. Gries and A. Stefanowitsch (eds.), *Corpora in Cognitive Linguistics. Corpus-Based Approaches to Syntax and Lexis* (1–18), Berlin/New York: Mouton de Gruyter.
- Grondelaers, S. and D. Geeraerts. (1998). Vagueness as a euphemistic strategy. In: Athanasiadou, A. and E. Tabakowska (eds.), *Speaking of Emotions. Conceptualisation and Expression* (357–374), Berlin / New York: Mouton de Gruyter.
- ISO. 2000. ISO 704. *Terminology work – Principles and methods. Travail terminologique – Principes et méthodes*. Geneva: ISO.
- Rogers, M. A. (2005). Lexicology and the study of terminology. In: D.A. Cruse, F. Hundsnurscher et al. (eds.), *Lexikologie. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen* (1847–1854), Berlin / New York: Walter de Gruyter.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar. Theoretical Foundations*. Stanford: Stanford University Press.
- Langacker, R. W. (1991). *Foundations of Cognitive Grammar. Volume 2: Descriptive applications*. Stanford: Stanford University Press.
- Picht, H. (1997). Zur Theorie des Gegenstandes und des Begriffs in der Terminologielehre. *Terminology Science and Research* 8–1/2 (159–177).
- Pustejovsky, J. (1988). The geometry of events. In C. Tenny (Ed.), *Studies in Generative Approaches to Aspect*, Number 24 in *Lexicon Working Papers*. Cambridge, Massachusetts: Center for Cognitive Science.
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, Mass.: MIT Press.
- Rothkegel, A. (2005). Knowledge and text types. In: H. Van Dam, J. Engberg and H. Gerzymisch-Arbogast (eds.), *Knowledge Systems and Translation* (84–102), Amsterdam/Philadelphia: Mouton de Gruyter.
- Sambre, P. (2005). *Conceptualisation et émergence de la définition en langue naturelle. Une étude de cas sur Internet en néerlandais et en français*. Leuven: KU Leuven (Ph.D.).

- Tabakowka, E. (2002). What can the Polish instrumental be instrumental in? In Lewandowska-Tomaszczyk, B. & K. Turewicz (eds.). *Cognitive Linguistics Today*. Bern: Peter Lang (375–396).
- Talmy, L. (2000). *Towards a Cognitive Semantics. Vol. I: Concept structuring Systems*. Cambridge (Mass.): MIT Press.
- Ungerer, F. and H.-J. Schmid. (2006). *An Introduction to Cognitive Linguistics*. London: Pearson.
- Temmerman, R. (2000). *Towards New Ways of Terminology Description. The Sociocognitive Approach*. Amsterdam/Philadelphia: Benjamins.
- Tognini-Bonelli, E. (2002). Functionally complete units of meaning across English and Italian. Toward a corpus-driven approach. In: B. Altenberg & S. Granger (eds.), *Lexis in Contrast: Corpus-Based Approaches* (73–96), Amsterdam/Philadelphia: Benjamins.
- Vendler, Z. (1967). Verbs and times. In Z. Vendler (Ed.), *Linguistics in Philosophy*. Ithaca, New York: Cornell University Press.
- Wermuth, C. (2005). *Een framegebaseerde benadering van classificatierubrieken: cardiovasculaire rubrieken als case study*. Amsterdam: VU Amsterdam (Ph.D.).

Biographical notes

Franco Bertaccini teaches courses on Terminology and Specialised Languages at the School for Interpreters and Translators of the University of Bologna at Forlì, Italy, where he is also the director of the Terminology Research Laboratory. At present, his main research interest is terminology/terminography within Computer-Aided Translation. (e-mail: bertaccini@sslmit.unibo.it).

Cédric Bousquet is a clinical practitioner at the Department of Public Health of the teaching hospital of Saint-Etienne, France.

He graduated in Pharmacy in 2001 and got his Ph.D. in Medical Informatics in 2004. From May 2005 to October 2007, Cédric Bousquet was a clinical fellow at the Medical Informatics department of the European Georges Pompidou Hospital in Paris. Since 2001, he has been working half-time in the pharmacovigilance regional centre of the European Georges Pompidou Hospital and has acted as an expert to the French Drug Safety agency; and half-time at INSERM in research positions. He is now co-ordinating a project on adverse drug reaction terminologies and signal generation in pharmacovigilance which is funded by the French Ministry of Research. (e-mail: cedric.bousquet@spim.jussieu.fr)

Peter De Baer graduated as an industrial engineer in electronics at the *Karel de Grote-Hogeschool* Antwerpen in 1994. In 1995, he earned the additional degree of system analyst in a course directed by Prof. C. De Backer. He is currently employed as researcher at the Department of Applied Linguistics of Erasmus University College Brussels. His main research interests are software development, ontology modelling, computer-assisted translation and knowledge management. (e-mail: peter.de.baer@ehb.be)

Sara Castagnoli has a research contract at the University of Bologna at Forlì, Italy, where she has taught courses on Terminology and LSP and where she has been working mainly on corpora for translation and corpus-based terminology/terminography. She has completed her PhD in Linguistics at the University of Pisa, Italy, focusing on learner translation corpora. (e-mail: scastagnoli@sslmit.unibo.it).

Claudia Dobrina, Ph.D. in English, has been working as a terminologist at Terminologicentrum TNC (The Swedish Centre for Terminology) since 1992. She has been and is currently involved in various terminology and special language

projects. She is especially interested in theoretical and methodological aspects of terminology work and has published a number of papers on the subject. She has a wide experience of terminology standardization and was for many years Secretary of the subcommittee 1 of ISO/TC 37 *Terminology and other language and content resources*. (e-mail: Claudia.dobrina@tnc.se)

Danielle Dubroca Galin, Àngela Flores García, Valérie Collin Meunier and Marc Delborge belong to a group of lecturers, most of them holding a Ph.D. They are interested in questions about translation, particularly the international trade of local products in a linguistic perspective, within the framework of their research project, subsidized by the Region of Castilla y Leon, entitled '*Traducción y marketing: Exportar los productos y servicios de nuestras tierras*' (Nº US 06/04). The group consists of four Spaniards (Danielle Dubroca Galin, Àngela Flores García, Valérie Collin Meunier, Marie-Noëlle Sánchez García, from the University of Salamanca), two Frenchmen (Jean-Marie Flores, University of Pau -UPPA- and Christian Vicente García, University of Haute Alsace UHA) and one Belgian (Marc Delborge, Lessius University College/KU Leuven), all driven by a European spirit, and authors of several publications. They work in Spanish, French and occasionally Portuguese (e-mails: danielle@usal.es, marc.delborge@lessius.eu)

Márta Fischer is a Ph.D. student in Translation Studies at the Eötvös Loránd University in Budapest. Her research is focused on EU language policy, translation and the implications of EU translation work for terminology. She holds degrees in Economics, German specialized translation and European Studies. She is currently teaching at the University of Pécs, at the Faculty of Economics in the Department of Specialized Translation and Terminology. (e-mail: fischerma@tkk.pte.hu)

Jody Foo is a Ph.D. student at the Department of Information and Computer Science, Linköping University. He holds an MSc in Cognitive Science.

His current field of research is Computer Aided Terminology Work, which includes human-computer interaction issues in terminology work, terminology extraction and automatic term spotting. Foo is also CTO for Fodina Language Technology, a company specializing in quality assurance of documentation, which includes terminology extraction, and various applications of terminology, such as standardization and machine translation. (e-mail: jodfo@ida.liu.se)

Ágota Fóris (born: 1970, Hungary), habil. (2006), Ph.D. (2002), linguist. Her field of research includes LSP lexicography, researching technical and scientific vocabulary, and terminology. She is the editor and author of the Hungarian-Italian and Italian-Hungarian Technical and Scientific Dictionary, and the author of the book *Hat terminológia lecke* [Six Lectures on Terminology], and many other publications. She is an elected member of the Dictionary Work-committee of the

Hungarian Academy of Sciences, an elected member of the board of the Hungarian Association of Applied Linguists and Language Teachers, and a founding member of the MaTT (Council of Hungarian Terminology).

She works at the University of West Hungary (Szombathely, Hungary) as Associate Professor and Head of the Terminology Innovation Centre. (e-mail: aforis@t-online.hu)

Mercedes García de Quesada lectures in Terminology and Interpreting at the University of Granada, Spain. Her main research interests are Frame Semantics, Terminography and Quality Studies in Interpreting. Her work has been published in several international peer-reviewed journals specialized in Translation, Terminology and Linguistics, such as *Perspectives*, *Terminology*, *META*, *Babel* and *Journal of Pragmatics*. (e-mail: mgarciaq@gmail.com)

Joaquín García Palacios is Doctor of Hispanic Philology and Lecturer in the Department of Translation and Interpreting at the University of Salamanca. He has taken part in various research projects on Spanish Language, Lexicology, Lexicography and LSP Terminology. He has published papers and monographs on those topics. He is currently leading a number of research projects on terminological neology in Spanish. (e-mail: gpalacios@usal.es)

Sancho Geentjens is presently working as a usability designer at Human Interface Group in Mechelen (Belgium). He left the *Centrum voor Vaktaal en Communicatie* of Erasmus University College Brussels in 2006, where he did a research project compiling an ontology-based specialized translation dictionary in the automotive domain. The terminology extracted from texts in five languages (English, French, German, Dutch and Italian) was structured by means of a conceptual/ontological model. This model was to help solve various problems caused by intra en interlinguistic variation and by differences in automobile makes.

Koen Kerremans is a Ph.D. researcher at the *Centrum voor Vaktaal en Communicatie*, a research centre at Erasmus University College Brussels (Department of Applied Linguistics). His main research interests range from ontologies and knowledge management to translation, special languages and terminology. His work has appeared in journals such as *Terminology* and *Linguistica Antverpiensia*. (e-mail: koen.kerremans@ehb.be)

Vassilis Korkas

Vassilis Korkas is a Senior Tutor in Translation and the programme director for the MSc in Specialist Translation and Translation Technology at the University of Surrey. Among various subjects, he teaches Internet Resources for Translation and Terminology, CAT Tools and Localization. He has a very strong research

interest in terminology and translation technology pedagogy and has spoken at various conferences on the subject. (e-mail: V.Korkas@surrey.ac.uk)

Monica Massari graduated in 2005 from the School for Interpreters and Translators of the University of Bologna at Forlì, Italy, with a dissertation on LSP translation. She currently works as a free-lance translator and terminologist in several technical and scientific fields. (e-mail: monica.massari@tiscalinet.it).

Magnus Merkel is Associate Professor in Computational Linguistics at the Department of Information and Computer Science, Linköping University. He holds an MA in Language Education, a Licentiate Degree in Computer Science and a Ph.D. in Computational Linguistics. Research interests are terminology extraction, terminology validation, translation tools and quality assurance of terminology. Merkel is also CEO of Fodina Language Technology, a company specializing in quality assurance of documentation, which includes terminology extraction, and various applications of terminology, such as standardization and machine translation. (e-mail: magne@ida.liu.se)

Judith Muráth Associate Professor, Dr.phil. habil.

University of Pécs, Faculty of Business and Economics, Institute of Applied Studies in Business and Economics, Department of Specialized Translation. Head of Department, Leader of Programme Specialized Translation and Interpreting, Leader of Terminology Documentation Centre. Research Subjects: LSP, Specialized Lexicography, Translation, Terminology and Translation. Member of TermNet (Member of Board), of MATT (Council on Hungarian Terminology), of Linguistic Association of the Hungarian Academy of Sciences. (e-mail: murath@ktk.pte.hu)

Henrik Nilsson, born in 1970, works as a terminologist at the Swedish Centre for Terminology (*Terminologisentrum* TNC) since 1997. As such he has worked with Eurodicautom and he has administered and participated in various terminology-related MLIS projects (Nordterm-Net and WebIT/EFCOT). In this context, he presented papers at the previous EAFT and TKE conferences and summits. He participated in one of the working groups during the development of the IATE database of the European Commission, software he later evaluated within the national project preparing a national terminological infrastructure in Sweden (named TISS). Within TNC, he takes part in various terminology projects and directs joint group for terminology of the planning and building sector. He is also responsible for marketing activities and is the web editor of TNC's website. He is co-responsible for the contents of the national *termbank*, *Rikstermbanken*, currently under construction. He is also one of the TNC staff teaching terminology to various groups (undergraduates, employees at various private companies and public authorities) and he often presents papers at national and international

conferences. He has been on the EAFT board for six years and has visited a large number of European terminology institutions. He holds a degree in communication science, English and French and is also a trained teacher of English. (e-mail: hni@tnc.se)

Arianne Reimerink lectures in Translation at the University of Granada, Spain. She works as a researcher in the field of Terminology and Translation. Her main research interests are Terminography, Translation Studies, Academic Writing and Computer Assisted Translation Applications. Her work has been published in several international peer-reviewed journals specialized in Translation, Terminology and Linguistics, such as the *International Journal of Lexicography* and *Terminology*. (e-mail: arianne@ugr.es)

Margaret Rogers is Professor of Translation and Terminology Studies and Director of the Centre for Translation Studies at the University of Surrey, UK. During the 1990s, she was involved in a number of EU-supported projects developing software tools for terminology and translation applications. Her most recent publications are in terminological text analysis and information structure in specialist translation. Co-editor of two monograph series on translation, she is also the founder of the Terminology Network at the Institute for Translation and Interpreting and a founder member of the Association for Terminology and Lexicography. (e-mail: M.Rogers@surrey.ac.uk)

Paul Sambre (1969) studied Romance Philology, Semiotics and Pragmatics in Leuven, Bologna and Lyon. He holds a Ph.D. in Cognitive Linguistics from the KU Leuven on definitions in natural language. His research spans French and Italian. His current research interests concern cognitive discourse models, blending theory and multimodality, the relation between cognitive linguistics and French post-structuralism and phenomenology, as well as concept modelling of associative relations like instrumentality and futurity. He teaches discourse studies, Italian linguistics and marketing communication at Lessius University College Antwerp/KU Leuven. (e-mail: paul.sambre@lessius.eu)

Lara Sanz Vicente is a Ph.D. researcher in translation studies at the University of Salamanca, Spain. She has taken two degrees, one in Translation and Interpreting and another in Geography, and is currently working on her Ph.D. thesis on remote sensing terminology. She teaches Terminology in translation studies and her research areas include LSP, neology, technical/scientific translation, and corpus-based terminography. (e-mail: larasanz@usal.es)

Frieda Steurs is a full professor in terminology, technical translation and language technology and the head of the research group 'Language and Computing' at Lessius University College, Antwerp/KU Leuven, Belgium.

She teaches Terminology and Documentation, Media and Translation and Technical and Scientific translation. Her research includes terminology management, language technology and standardization. This has led to several projects with industrial partners and government organizations. She is the founder and former president of NL-TERM, the Dutch terminology association for both the Netherlands and Belgium. She chairs the ISO TC/37 standardization committee for Belgium and the Netherlands and is president of Coterm, the government body on terminology for the Netherlands and Belgium.

She has been the head of the Department of Applied Language Studies at Lessius University College since October 1997. She also teaches as a guest professor at: The University of Angers (UCO), France
The Terminology Summer School (Vienna, Koln)
(e-mail: frieda.steurs@lessius.eu)

Rita Temmerman teaches translation studies and terminology theory at Erasmus University College Brussels, where she co-ordinates the *Centrum voor Vaktaal en Communicatie*, a research centre in applied sociocognitive terminology. Her wider research interests include translation, terminology, knowledge management, multilingualism and cross-cultural communication. She has published a monograph with John Benjamins and multiple articles in collective volumes and journals such as *Terminology*. (e-mail: rita.temmerman@ehb.be)

Marcel Thelen was the interim chairman, and now vice-chairman of the Dutch-Flemish association for Dutch terminology (NL-TERM).

He works at the Department of Translation and Interpreting of the Maastricht School of International Communication of Zuyd University in Maastricht, the Netherlands, where he has been a full-time lecturer in translation and terminology for the past 25 years. Since 2005 he has been Head of the Department. He has published frequently on translation, terminology, lexicology, semantics and quality management. As the technical chairman of the Dutch standards organization (NEN) he contributed to the new European standard on translation services, EN 15038. He is involved in contacts with the industry in his capacity of director of the in-house simulated translation bureau of his Department. He is researcher for the academic chair for International Business and Communication. He is co-organizer of the 5-yearly International Maastricht-Łódź Duo Colloquia on 'Translation and Meaning', and editor-in-chief of the corresponding book series. (e-mail: m.m.g.j.thelen@hszuyd.nl)

Alice Toma

Doctorate, Ph.D.: philology (Bucharest, 2005), linguistics (Geneva); Postgraduate Certificate, PGCE: Contemporary Romanian Language (Letters, Bucharest, 1999); Theoretical and Applied French Linguistics (Bucharest, Languages and Foreign Literatures, 2000), Linguistics (Geneva, 2003); Master's Degree: philology (Bucharest, 1998), mathematics (Bucharest, 2000); Publications: Segmented Constructions (in French) (2008, 170 p), Linguistics and Mathematics (in Romanian) (2006, 2008, 520 p), Interdisciplinary Scientific Vocabulary (in Romanian) (ed. A. Bidu-Vranceanu, 2001, p.141–238), Common Vocabulary, Specialized Vocabulary (in Romanian) (ed. A. Bidu-Vranceanu, 2000, p. 85–118); more than 35 articles.

Maria-Cornelia Wermuth, Ph.D. in Applied Linguistics, has been working as an assistant professor at the department of Applied Linguistics at Lessius University College Antwerp/KU Leuven since 1994. She is teaching German grammar in the Bachelor programme and Medical Translation in the Master in Translation. She is involved in terminological and LSP research as well as in (applied) cognitive linguistics research and frame semantics. She is especially interested in methodological aspects of specialized translation training and in the development of medical ontologies and frame-based models of medical language. She has published a number of papers on the various subjects. She also is experienced in medical terminological standardization. (e-mail: cornelia.wermuth@lessius.eu)

Maria Zimina works as a consultant in document management for Orange Business Services, France Telecom (Paris).

In 1998, she obtained a Master of Letters in Linguistics, Translation Studies and Foreign Language Teaching from Lomonossov Moscow State University. Maria Zimina obtained her Ph.D. in Language Studies and Linguistics from Paris 3 – Sorbonne Nouvelle University in 2004. For the past few years, Maria Zimina has worked as a PostDoc researcher within several institutions in France, such as IRDES (Institut de Recherche et Documentation en Economie de la Santé), INaLCO (*Institut National des Langues et Civilisations Orientales de Paris*) and *Paris Nord* – Paris 13 University. (e-mail: zimina@msh-paris.fr)

Author index

A

Ahrenberg, L. 167, 170, 179
 Aijmer, K. 179
 Albert, S. 42, 45, 48
 Albertz, J. 197, 208
 Alcalá, A.R. 197, 208
 Altenberg, B. 179, 254
 Anderman, G. 135, 230
 Andrassy, G. 50, 57
 Arnaud, A. 160
 Arntz, R. 27, 32, 124, 135
 Athanasiadou, A. 253
 Atkins, S. 97, 99, 118, 119
 Auger, P. 124, 135
 Aussenac-Gilles, N. 187, 188, 192

B

Bach, E. 119
 Bachimont, B. 110, 120
 Baisier, L. 181, 192
 Bajza, J. 37, 45
 Baker, P. 239, 253
 Baneyx, A. 214, 229, 230
 Barlow, M. 239, 253
 Barsalou, L. 101, 118
 Berners-Lee, T. 192
 Berry, L. 230
 Bertaccini, F. 4, 11, 16, 19, 255
 Bertels, A. 185, 186, 192
 Biber, D. 203, 208
 Blampain, D. 192
 Blommaert, J. 120
 Bombi, R. 19
 Borgulya, I. 42, 45
 Bösel, M. 50, 51, 57
 Bourigault, D. 120, 145, 222, 229, 230
 Bousquet, C. 7, 213, 230, 255
 Bowker, L. 185, 186, 192, 202, 208
 Bradean-Ebinger, N. 58
 Branchadell, A. 21, 32
 Brivio, P.A. 197, 208
 Bucher, A.-L. 64, 77
 Budin, G. 62, 77, 127, 128, 135

Bulcaen, C. 120
 Burr, I. 33
 Busch-Lauer, I. 239, 253

C

Cabré, M.T. 14, 19, 42, 45, 89, 94, 98, 100, 118, 151, 160, 185–187, 192, 202, 204, 208
 Castro, R. 119
 Charlet, J. 213, 214, 229, 230
 Chiao, Y.-Ch. 217, 230
 Chikán, A. 39, 45
 Chuvieco, E. 197, 208
 Cifoletti, G. 19
 Clozel, M. 252, 253
 Cobarrubias, J. 32
 Condamines, A. 192
 Cook, M.P. 97, 105, 110, 119
 Corazzari, O. 203, 208
 Corino, E. 45
 Croft, W. 234, 239, 253
 Cruse, D.A. 234, 239, 253
 Cuartero, J. 121

D

Daconta, M.C. 183, 192
 Daille, B. 186, 192, 217, 230
 Dancette, J. 6, 138–141, 145
 Davidson, L. 192
 Davies, J. 146, 193
 De Baer, P. 7, 145, 181, 255
 De Bessé, B. 202, 209
 Déjean, H. 230
 Desmet, I. 185, 192
 De Vicente, A. 145
 Diez-Orzas, P. 146
 Dirven, R. 234, 253
 Dolbey, A. 100, 119
 Dollerup, C. 136
 Domènech, M. 118
 Dragaschnig, E.H. 51, 59
 Draskau, J. 124, 135
 Dróth, J. 42, 45, 50, 58, 59
 Dubroca Galin, D. 6, 149, 158, 161, 256

Dubuc, R. 94, 185, 192
 Dunne, K.J. 179
 Dymetman, M. 179
 Dyvik, H. 177, 179

E

Ellsworth, M. 119, 121
 Emsel, M. 121
 Engberg, J. 253
 Ensink, T. 253
 Erdős, J. 50, 58
 Evans, N. 184, 192

F

Faber, P. 97, 100–102, 104, 105, 107, 113, 119
 Fábíán, P. 37, 45
 Fabre, C. 222, 229
 Fata, J. 54, 55, 57, 58
 Fekete-Silye, M. 57–59
 Felber, H. 29, 119, 134–136
 Feliu, J. 118
 Fensel, D. 146, 193
 Fernandez, M. 137, 145
 Fillmore, C.J. 5, 97–99, 102, 106, 119, 120, 237, 248, 253
 Fischer, M. 4, 21, 23, 32, 50, 58, 256
 Fishman, J.A. 32
 Fleury, S. 230
 Fluck, H.-R. 52, 58
 Fodor, J.A. 102, 119
 Foo, J. 6, 163, 174, 175, 179, 256
 Fóris, Á. 4, 35, 38, 41–43, 45, 51, 58, 256
 Foster, G. 179
 Fracchiolla, B. 230
 Freixa, J. 186, 192
 Fuentes, M.T. 196, 209
 Fuertes Olivera, P. 120
 Fung, P. 217, 230
 Fusco, F. 19

G

Gaivenis, K. 42, 45

- Galinski, C. 62, 77
 Gallardo San Salvador, N. 119
 Gallyas, Cs. 50, 58
 Gangemi, A. 119
 García, J. 97, 102–104, 110, 119, 120, 195, 196, 209, 256, 257
 García de Quesada, M. 97, 102–104, 119, 120, 257
 Gárdus, J. 48, 49, 58
 Garrett, M.F. 119
 Gaudin, F. 185, 192
 Gaussier, É. 230
 Geeraerts, D. 247, 253
 Gerzymisch-Arbogast, H. 52, 53, 58, 239, 253
 Giraud, G. 152, 160
 Gómez-Pérez, A. 137, 145
 González del Rey, I. 161
 Granger, S. 254
 Gries, S. 253
 Grignetti, A. 197, 209
 Grondelaers, S. 247, 253
 Gross, G. 56, 161
 Guerrero, G. 209
- H
 Habert, B. 192
 Hakel, M.D. 120
 Halpern, D.F. 120
 Halskov, J. 113, 120
 Harms, R. 119
 Haugen, E. 13, 19, 32
 Hawkins, S. 185, 186, 192
 Heller, B. 137, 145
 Heltai, P. 32, 42, 45, 50, 53, 58
 Hendler, J. 192
 Herre, H. 137, 145
 Hessky, R. 54, 58
 Hoffmann, L. 52, 58
 House, J. 48, 67, 68, 87, 97, 120, 165, 260
 Hubainé Oláh, Á. 50, 58
 Hundsnurscher, F. 253
- I
 Innocente, I. 19
 Isabelle, P. 166, 179
 Ivanova, J. 218, 220, 223, 230
- J
 Jacquemin, C. 120, 145, 192
 James, G. 230
 Järvinen, T. 168, 179
- Jaulent, M.-C. 214, 229, 230
 Jiménez Hurtado, C. 100, 119
 Johnson, C.R. 97–99, 119–121
 Juristo, N. 145
 Jutrac, J.-M. 179
- K
 Kaalep, H.-J. 179
 Kalliokuusi, V. 64, 77
 Károly, K. 45, 50, 58
 Kavanagh, J. 192
 Kemmer, S. 239, 253
 Kerremans, K. 7, 137, 138, 142, 145, 146, 181, 187, 188, 192, 257
 Kiss, J. 46
 Kittay, E. 119
 Klár, J. 38, 46, 51, 52, 58
 Klaudy, K. 25, 32, 38, 45, 46, 48, 58
 Kloss, H. 23, 32
 Knops, U. 145
 Koit, M. 179
 Koskinen, K. 25, 32
 Kovács, I.J. 50, 53, 58, 59
 Kovalovszky, M. 37, 38, 46, 51, 52, 58
 Krommer-Benz, M. 135
 Kuncova, A. 230
 Kurtán, Zs. 51, 58
- L
 Lamalle, C. 224, 227, 230
 Lande, B. 230
 Lang, F. 121, 136, 254
 Langacker, R.W. 235–237, 239, 240, 243–245, 247, 248, 250, 253
 Lanstyák, I. 41, 46
 Lassila, O. 192
 Lauriston, A. 186, 192
 Lebart, L. 220, 224, 225, 230
 Lehrer, A. 119
 León, P. 119, 208
 León Araúz, P. 119
 Lerat, P. 161, 185, 192
 Lewandowska-Tomaszczyk, B. 254
 L'Homme, M.-C. 110, 120
 Lindig, G. 197, 209
 Loddegaard, A. 136
 Loening, K.L. 209
 Lombard, R. 164, 179
 Lovell, M.W. 21, 32
- M
 Mackintosh, K. 202, 209
 Macklovitch, E. 179
 Magris, M. 19
 Maingueneau, D. 161
 Maisondieu, A. 230
 Malaisé, V. 110, 120
 Maniez, F. 161
 Marello, C. 45
 Márquez, C. 101, 119
 Márquez Linares, C. 119
 Martínez, W. 226, 227, 230
 Mayer, F. 14, 19, 56, 97, 105, 110, 120
 Mayer, R.E. 14, 19, 56, 97, 105, 110, 120
 McEnery, A. 218, 230
 McKeown, K. 217, 230
 Merkel, M. 6, 163, 167, 174, 175, 179, 258
 Meyer, I. 98, 110, 120, 137, 145, 187, 192, 202, 209
 Mille, F. 230
 Mitkov, R. 98, 120
 Montero-Martínez, S. 102–104, 119, 120
 Morel, J. 118
 Moreno, A. 187, 193
 Morin, E. 217, 230
 Muchnik, J. 151, 160, 161
 Muischnek, K. 179
 Müller, F. 33
 Muráth, J. 5, 27, 29, 32, 47–54, 57–59, 258
 Musacchio, M.T. 19
- N
 Nilsson, H. 5, 61, 77, 258
 Nivre, J. 179
 Nolet, D. 94
 Nuk, I. 218, 220, 223, 230
- O
 Obrst, L.J. 192
 Oláh-Hubai, Á. 59
 Old, J.L. 3, 6, 39, 55, 56, 64, 76, 85, 86, 89, 123, 152, 157, 159, 177, 179
 Onesti, C. 45
 Orioles, V. 19
 Östman, J.O. 120
 Ozdowska, S. 222, 230

- P
- Pacifici, S. 89, 94
- Pais, D. 37, 46
- Parkes, C.H. 119
- Paul, S. 196, 197, 199, 209, 233, 259
- Pavel, S. 94, 202, 209
- Pearson, J. 202, 204, 209, 254
- Peeters, W. 146
- Peña Cervel, M.S. 253
- Péntek, J. 41, 42, 46
- Pérez, C. 102, 120, 137, 145, 187, 193, 202, 209
- Pérez Hernández, C. 102, 120
- Perrault, F. 179
- Pescheux, M. 161
- Petruck, M.R.L. 97, 99, 119, 120, 121
- Pettersson, R. 72, 77
- Petterstedt, M. 167, 179
- Picchi, E. 203, 208
- Picht, H. 124, 135, 235, 239, 253
- Piehl, A. 25, 32
- Pisanelli, D.M. 119
- Pogány, I. 51, 59
- Poirot-Zimina, M. 230
- Pozzer, L.L. 105, 110, 120
- Prandi, M. 19
- Prieto, J.A. 104, 119, 120
- Prieto Velasco, J.A. 104, 119, 120
- Priss, U. 177, 179
- Pustejovsky, J. 246, 253
- Pusztai, F. 38, 46, 51, 54, 59
- Pusztai, I. 38, 46, 51, 54, 59
- Pym, A. 25, 32
- R
- Rabchevsky, G.A. 196, 209
- Rádai-Kovács, É. 50, 53, 59
- Raed, M.A. 197, 204, 209
- Rector, A.L. 137, 145
- Rédey, K. 51, 57
- Rega, L. 19
- Reimerink, A. 4, 5, 97, 119, 259
- Réthoré, C. 6, 138–140, 145
- Robinson, R. 129, 135
- Rodríguez, C. 99, 110, 113, 118, 120, 121
- Rodríguez Penagos, C. 99, 110, 113, 120, 121
- Rogers, M. 6, 123, 130, 135, 186, 193, 235, 253, 259
- Rothkegel, A. 233, 253
- Roth, W.M. 105, 110, 120
- Royauté, J. 192
- Rundell, M. 118
- Ruppenhofer, J. 98, 102, 121
- S
- Sadat, F. 217, 230
- Sager, J.C. 21, 26, 31, 33, 124, 136, 187, 193
- Salem, A. 230
- Sambre, P. 8, 233, 235, 239, 253, 259
- Sandrini, P. 27, 28, 33
- Sanz, M.L. 7, 195–197, 203, 209, 259
- Sato, H. 118
- Sauer, Ch. 253
- Scarpa, F. 19
- Schapira, Ch. 161
- Scheffczyk, J. 119
- Schmid, H.-J. 234, 254
- Schmidt, T. 100, 121
- Schmitt, P.A. 27, 33, 134, 136
- Schulze-Kremer, S. 137, 145, 146
- Seibel, C. 102, 121
- Senso, J. 119
- Serra, E. 121
- Short, N.M. 29, 36, 42, 86, 89, 91, 129, 152, 197, 198, 203, 208, 210
- Sintuzzi, S. 19
- Sipos, G. 58
- Skuce, D. 192
- Slodzian, M. 230
- Smith, K.T. 192
- Snell-Hornby, M. 32, 33
- Sobrero, A.A. 13, 19
- Somssich, R. 30, 33
- Sonneveld, H.B. 209
- Steve, G. 119
- Studer, R. 137, 146, 187, 193
- Subirats Rüggeberg, C. 121
- Suonuuti, H. 94
- Sure, Y. 137, 146, 187, 193
- Szabó Mihály, G. 41, 46
- Szépe, Gy. 51, 59
- Szulman, S. 192
- T
- Tabakowska, E. 234, 253
- Talmy, L. 235, 236, 248, 250, 254
- Tapanainen, P. 168, 179
- Temmerman, R. 6, 7, 14, 19, 100, 121, 137, 139, 145, 146, 181, 185, 187, 192, 193, 239, 254, 260
- Tenny, C. 253
- Thoiron, P. 192
- Togni, S. 19
- Tognini-Bonelli, E. 254
- Tolnai, V. 37, 46
- Trosborg, A. 193
- Tummers, J. 145, 192
- Turcsányi, G. 58
- Turewicz, K. 254
- U
- Ungerer, F. 234, 254
- V
- Valeontis, K. 62, 77
- Van Campenhoudt, M. 192
- Van Dam, H. 253
- Van Hamelen, F. 146, 193
- Várnai, J. Sz. 40, 46, 50, 54, 56, 59
- Vega, M. 101, 119
- Vega Expósito, M. 119
- Vendler, Z. 247, 254
- Véronis, J. 217, 231
- Vihonen, I. 25, 32
- Vossen, P. 146
- W
- Wagner, E. 25, 29, 33
- Walker, E.C.T. 119
- Wermuth, C. 8, 233, 239, 254, 261
- Wersig, G. 136
- Wotjak, G. 121
- Wright, S-E. 62, 77
- Wüster, E. 51, 52, 59, 124, 127, 129, 132, 136, 235
- X
- Xiao, R.Z. 218, 230
- Z
- Zajankauskas, S. 43, 46
- Zampolli, A. 119
- Zani, G. 197, 208
- Zimina, M. 7, 213, 224, 226, 229–231, 261
- Zserdin, M. 51, 59
- Zweigenbaum, P. 110, 120, 217, 230

Subject index

A

abstracts 233, 237, 238, 250
 acquis communautaire 24
 acronyms 15–17, 200, 201
 action chains 236
 adverse drug reactions 213, 215, 217, 221, 223, 226
 Agent 101, 103, 188, 189, 236, 241, 246
 align-filter approach 167
 alignment 7, 163–175, 178, 179, 213, 219, 222, 223, 231
 Aristotelian model of genus-differentiae 133
 Armenian 36
 associative relations 141, 187, 233–235, 251, 252, 259
 authoring 6, 163, 164, 166
 automatic term recognition 70, 192
 automotive domain 137, 138, 142, 145, 257

B

backgrounding 235
 Berkeley FrameNet project 5, 97
 bilingual terminological database 196, 202, 208
 BioFrameNet 100, 119
 borrowing 12–14, 19, 195, 201
 Bulgarian 36

C

calquing 120, 201
 Case Grammar 99, 248
 Catalan 198
 categorization framework 138, 142–145, 183, 187–189, 191
 causal chain 235–237
 causality 233, 234
 Chinese 36
 classical terminology theory 52
 Coastal Engineering Event (CEE) 101

Coastal Engineering 5, 6, 97, 99, 100, 101, 107, 109, 119
 cognitive frame 142
 cognitive frameworks 144
 cognitive framing 138
 Cognitive Grammar 8, 233, 235, 236, 251, 253
 Cognitive Linguistics (CL) 234
 cognitively underpinned translation dictionary 139
 collocations 53, 106, 186, 202, 206
 Comité canado-québécois de terminologie de la télé-détection 197
 Commission ministérielle de terminologie de la télé-détection aérospatiale (COMITAS) 197
 communication between EU institutions and European citizens 22
 communication levels 22
 communication within EU institutions 22
 communicative approaches 201, 203
 communicative settings 12, 100, 204
 communicative setting 203
 Communicative Theory of Terminology 100, 204, 206
 comparable corpora 195, 196, 203, 213, 214, 217, 224, 229, 230
 competency management 181, 182
 compound words 199
 concept catalogues 70
 concepts 3, 12, 28–30, 32, 36, 39, 40, 42, 43, 46, 53, 63, 65, 67, 69, 70, 72–74, 84, 86, 100, 101, 113, 119, 129, 132–134, 138, 140, 144, 165, 177, 183, 184, 195, 196, 200, 202, 204, 234, 251
 conceptual domain 202

conceptual EU systems 40
 conceptual frames 137
 conceptually framed terminological dictionary 140
 conceptual scheme 235–237
 conceptual structure 98, 102, 113, 117, 204, 205, 207
 conceptual system of socialist economy 39
 conceptual systems of economics and business and finance 39
 concordances 205, 220
 context-dependent definitions 190, 191
 context 3, 21, 23, 25–27, 30, 31, 50, 52–54, 56, 57, 68, 88, 91, 97, 98, 109, 113, 114, 116–118, 120, 123–125, 127, 128, 130, 132, 135, 138–140, 142, 149–151, 153, 159, 173, 181–183, 190, 191, 196–198, 201–204, 209, 217, 223–227, 246, 251, 258
 contextual autonomy 47
 contextual examples 123, 126, 130–134
 controlled language checkers 164
 controlled vocabularies 70
 corpus-based lexicography 202
 corpus-based terminography 202, 259
 corpus-based terminology 7, 98, 255
 corpus linguistics 179, 209, 213, 218, 231, 233, 237–239
 corpus planning 21, 23, 24, 30–32
 correspondence analysis (CA) 220
 Croatian 36
 cross-cultural translation 139, 140
 cross-linguistic comparison 195, 205

cross-linguistic equivalence 11
cultural transfer 25

D

Danish 85, 233
definition 30, 42, 45, 52, 62, 63,
68, 81, 84, 89–91, 97–99, 102,
104–109, 113, 114, 129, 132–135,
138, 140, 142, 157, 177, 188–190,
202, 206, 207, 233, 251, 252
definitions 6, 8, 43, 68–71,
84–86, 89, 90, 94, 97, 99, 102,
104, 106, 114, 119, 120, 123, 126,
129–134, 166, 176, 190, 191, 233,
251, 252, 259
definitions of terms 43
descriptive-communicative
terminography 202
dictionaries 6, 37, 38, 40–43, 45,
47–50, 53–55, 58, 70, 87, 119,
129, 137, 142, 144, 156, 170, 171,
174, 196, 204, 205, 224
digital terrestrial television 4,
11, 12, 14, 15
direct loans 201
discursive autonomy 139
domain loss 21, 63
dominant language 198
Dutch 1, 5, 6, 137, 142, 144, 182,
188, 189, 257, 260

E

EAFT Brussels Declaration – for
international cooperation on
terminology
economic organization 39
economy 35, 37, 39, 41, 47, 51, 216
e-government 65
e-HRM 181, 182, 191
elaborated instrumentality 236
electronic Human Resource
Management (e-HRM) 181,
182
English 3, 4, 6, 7, 11–13, 15–17,
19, 25, 28, 32, 36, 40, 42, 43,
50, 58, 59, 63, 69, 84–86, 88,
90–92, 95, 99–101, 107, 120,
137–139, 141, 142, 144, 164, 165,
171, 172, 175, 178, 179, 182, 188,
193, 195–198, 200–206, 209,
213–219, 222–224, 227–230,
249, 253–255, 257, 259
English as lingua franca 19, 195
equivalent terms 196, 202, 209

EU conceptual system 4, 27,
28, 30
EU language 21, 23, 256
EU terminology 21, 24, 28,
40, 55
events 25, 42, 69, 85, 107, 219,
234, 236, 239, 253
Event 101–103, 113–115, 117, 118,
235, 236, 243, 247
everyday Hungarian 41
extensional definitions 133
extract-align approach 166, 167
extraction of definition-like
sentences 99

F

false friends 201
filtering 126, 163, 167, 174
Finnish 27, 36, 64, 85
force dynamics 235
foregrounding 235
form and content 123, 125, 126
four different understandings of
'theory' 128
frame 97–100, 102, 105–110, 112,
113, 117–121, 142, 160, 218, 220,
237, 257, 261
frame-by-frame analysis of the
lexicon 106
Frame Element (FE) 97
frame-like structure 97
FrameNet-like descriptions 100
FrameNet 5, 97–100, 102, 106,
109, 118–121, 253
Frame Semantics (FS) 99
French 4, 6–8, 11, 12, 14, 15, 17,
25, 27, 28, 32, 36, 37, 59, 85, 91,
137–142, 144, 149, 150, 152–159,
182, 196–198, 201, 209, 213–215,
217–219, 222–224, 227–230,
233, 237, 238, 240–248, 250,
251, 255–257, 259, 261
frequency lists 205, 206
full text alignment 167
functional synonymy 16

G

General Theory for Terminol-
ogy (GTT) 51, 124

German 5, 6, 8, 25, 27, 28, 32, 36,
37, 47, 50–57, 59, 85, 100, 101,
107, 134, 137, 142, 144, 193, 196,
201, 209, 215, 233, 237, 238,
240–246, 248, 250, 251, 253,
256, 257, 261
German FrameNet 100
glossaries 17, 43, 63, 65, 70, 74,
197, 204, 205
Greek 36, 128, 199
Gypsy (Lovari and Beas) 36

H

half translations 201
horizontal terminology infra-
structure 62
Hungarian 4, 5, 27, 35–59,
256–258
Hungarian National Corpus 40
Hungarian Standards Insti-
tute 48
Hungarian terminology 4, 5,
35–40, 42–46, 49, 257, 258
HUTERM 49

I

IATE 30, 31, 61, 62, 71, 76, 77, 258
IATE Best Practice 30
IATE terminology database 30
i-Model 233–235
inconsistent terminology 164
information modelling 65,
82, 86
information models 70
instrumentality 8, 233–252, 259
instrumental prototype 247
instrumental role 233, 234, 236,
241, 243
intensional definitions 133
inter-conceptual term-trans-
fer 21, 26, 28, 32
interco-operative terminology
networks 74
International Classification of
Diseases (ICD) 216
international trade 6, 149, 256
intra-conceptual term-trans-
fer 21, 26, 29, 32
intracultural translation 25
isomorphisms 137
Italian 4, 6, 11–17, 19, 35–37, 85,
137, 142, 144, 196, 197, 209, 215,
254, 256, 257, 259
itchy fingers syndrome 125

- i-Term 233, 252
 Itools suite 168–170, 178
 IWT-TETRA funding programme 182
- J**
 Japanese 36, 100, 215
 JOGSCOT 66
 joint terminology groups 66, 74
- K**
 knowledge management 137, 146, 192, 193, 233, 255, 257, 260
 knowledge patterns (KPs) 113
 knowledge poor contexts (KPCs) 98
 knowledge rich contexts (KRCs) 98
 Korean 36
 KPC 98
 KRCs 98, 110, 113, 115, 118
- L**
 language equivalents 197, 201, 205
 Language for Special Purposes (LSP) 42, 130
 language planning 23, 31, 32, 35, 37, 40, 41, 46, 82, 87
 language policy 21, 22, 31, 256
 Latin 36, 37, 197, 198, 200
 legal translations 25
 lexical borrowings 11
 Lexical Semantics 99, 102, 118
 Lexical Syntax 104
 lexical units (LUs) 105
 lexicons 40, 70, 120, 169, 177, 224, 227
 LGP 42
 Lithuanian 43, 45, 68, 76
 loan translations 201
 loan-word 17, 195
 loan-words 195–198, 200, 201
 LSP 35, 37, 38, 40–42, 46–50, 52, 53, 70, 130, 198, 202, 233, 239, 255–259, 261
- M**
 management of translations 22
 MarcoCosta (MC) 100
 marketing communication 149, 259
 medical ontology 233
- Medical Subject Headings (MeSH) 216
 medical texts 186, 213, 214, 220, 229, 250
 medicine 8, 85, 92, 103, 137, 213, 216, 233
 meronymy 106, 235
 MHP platform 13
 minority languages 23, 36, 63, 69, 70
 MLIS-projects 65
 Multilingual Categorization Framework Editor (MCFE) 183, 188
 multilingual primary term formation 21
 multilingual terminology 5, 26, 145, 163, 182, 214, 215
 multilingualism 3, 22, 260
 multimodal corpus 97
- N**
 name lists 70
 National Board of Health and Welfare 61, 74, 75
 national centre for terminology 64, 74, 81, 83
 national term bank 5, 61–63, 65–69, 71, 76, 77
 natural language processing 179, 213, 214, 229
 near-synonymy 142
 neologism 56, 195, 201
 neologist movement 37, 38, 46
 nomenclatures 70
 non-isomorphism 138, 141, 144
 Nordic master programme in terminology 76
 Nordterm-Net 65, 71, 77, 258
- O**
 official languages 4, 21, 23, 24, 26–28, 31, 32
 one-to-one correspondence 47, 52
 O*NET 188
 onomasiological approach 27, 29
 onomasiology 129
 ontologically underpinned translators dictionary 138
 ontologies 70, 82, 119, 120, 145, 146, 183, 233, 257, 261
- ontology-based specialized translation dictionary 137, 257
 ontology engineering 7, 137, 187
 OOV project 144
- P**
 parallel corpora 25, 167, 170, 179, 213, 217, 230
 pathological synonymy 16
 Patient/Result 102
 permeability of the Italian language to English words 12
 PERTOMed project 7, 213, 217, 218, 223, 229, 230
 pharmacovigilance 7, 213–215, 217, 229–231, 255
 phraseological dimension 202, 205
 physiological synonymy 16
 PoCeHRMOM project 181–184, 188
 Pointer project 62, 65, 66
 Polish 36, 234, 254
 polysemes 123, 131–133
 Portuguese 156, 158, 196, 197, 209, 215, 256
 practice and theory 47, 49
 prepositional instrumentality 242, 251
 primary term formation 21, 26
 problem-solving 123, 125, 128, 130, 131
 procedural languages 24, 25, 28
 Process 12, 13, 24–29, 31, 43, 50, 51, 53, 71, 89, 101, 110, 114, 119, 126, 129, 131, 132, 150, 153, 163–169, 172, 174, 176, 178, 184, 198, 199, 201, 204, 206, 207, 216, 223, 226, 234, 236, 243–245, 247, 248
 promotional terminology 6, 159, 160
- Q**
 Quality Assurance (QA) 7, 75, 163, 165, 178, 256, 258
 quasi-synonymy 138, 141, 144
 query processing procedures 81, 82, 86, 87
 query 5, 81–89, 93–95, 107, 183
- R**
 real texts 201–203, 208

- remote sensing 7, 195–204, 207–210, 259
- response 67, 81–83, 88–95, 127, 215
- retailing domain 138, 142
- revision 26, 43, 51, 71, 163, 167, 175
- Rikstermbanken 5, 61, 62, 66–71, 74, 76, 77, 258
- role of translators 25, 32
- Romanian 36, 41, 42, 154, 261
- Russian 7, 36, 85, 213, 214, 216–224, 228–230
- Ruthenian 36
- S
- SALSA project 100
- schematic instrumentality 236
- secondary term formation 21, 26, 29
- SELPER 197, 204, 209, 210
- semantic mirroring strategy 177
- semantic network 107, 138
- Semantic Web 7, 146, 181, 182, 183, 184, 191, 192, 193, 251
- semasiological approach 98
- Serbian 36
- SFN 17, 18, 100, 107
- shared knowledge scheme 137
- Slovakian 36, 41
- Slovenian 36
- small languages 3, 35
- social pension 47, 55, 56
- social science 47, 57
- society 1–4, 35, 39, 45, 53, 54, 62–64, 68, 69, 72, 76, 81, 93, 119, 151, 159, 196
- Sociocognitive Theory of Terminology 100
- socio-terminological approach 11, 12
- Spanish 5–7, 36, 97, 99–101, 107, 118, 120, 121, 149, 150, 152, 153–155, 157–159, 161, 195–201, 203–206, 208, 215, 256, 257
- Spanish Association of Remote Sensing (AET) 204
- Spanish FrameNet 5, 97, 99, 100, 121
- specialized dictionaries 40, 47–50
- specialized language 11, 12, 14, 100, 118, 185
- Specialized Lexicography 48, 100, 120, 123, 258
- specialized translator 48, 127, 128
- spell and grammar checkers 164
- standardization 6, 11, 14, 16, 19, 41, 66, 73, 82, 163, 166, 183, 195–198, 201, 203, 208, 235, 256, 258, 260, 261
- standardized term bank 163–167, 175, 178
- standardized terms 163
- standards 39–41, 48, 64, 75, 86, 87, 179, 215, 229, 260
- state of necessity 12, 13
- status planning 21, 23, 24
- subject field 29, 30, 81, 85, 87–90, 92, 95, 124, 206, 207
- Swedish 5, 27, 36, 61–73, 76, 77, 81, 83–88, 90–92, 94, 95, 171, 172, 175, 178, 255, 258
- Swedish Centre for Terminology 5, 61, 64, 81, 255, 258
- synonymy 4, 11, 12, 14, 16, 106, 129, 131, 138, 141, 142, 144, 177, 186, 193
- syntactic structures 200, 202
- T
- taxonomies 70
- TDCNet 65
- technical Hungarian 41
- term banks 6, 66, 163, 165, 177–179
- term bases 47–50
- term candidates 163, 164, 167, 169, 172, 174–178, 203, 205, 223
- term-concept distinction 129
- term-concept relations 123, 129, 132
- term extraction 6, 7, 163, 174, 213, 222, 223, 229
- term formation 21, 26, 29, 201
- TermIK (Terminológiai Innovációs Központ / Terminology Innovation Centre 44
- terminographic definitions 97, 99, 102, 104, 120
- terminography 98, 100, 102, 105, 120, 123, 145, 185, 187, 192, 202, 203, 208, 255, 257, 259
- terminological awareness 62–65, 76, 77, 81, 83, 94
- terminological consequence analysis 75
- terminological dictionaries 137, 144
- terminological harmonization 40
- terminological information on demand 81
- terminological phraseme (TP) 103
- terminological query service 5, 81, 82, 89, 93, 94
- terminological resources 7, 65–67, 86–89, 100, 186, 187, 202, 213–216, 229
- terminological standardization 14, 261
- terminological transfer 153
- terminological variation 7, 14, 181–183, 185, 186, 187, 191
- terminologically reliable resource 81
- Terminologicentrum 61, 64, 81, 255, 258
- terminology 1–9, 11, 13, 14, 16, 19, 21, 24–28, 30–33, 35, 36–53, 55, 57–59, 61–69, 71–79, 81–83, 85–87, 94, 97–100, 113, 118–121, 123–133, 135–138, 142–146, 149–151, 153–155, 158–160, 163–166, 179, 181–187, 191–193, 195–198, 200–204, 206, 208, 209, 211, 213–215, 217, 223, 224, 226, 227, 229, 230, 233–235, 239, 251–260
- terminology co-ordination 61, 66, 72–75, 77
- terminology co-ordination programme 61, 66, 72–75, 77
- terminology co-ordinator 72–74
- terminology extraction 163, 179, 213, 217, 256, 258
- terminology infrastructure 3, 5, 61–63, 65, 66, 67, 72, 76
- terminology management for translation purposes 123
- terminology management system (TMS) 125
- terminology portal 5, 61, 66, 67, 69
- terminology projects 65, 74, 82, 118, 258
- Terminology Studies 123, 124, 126, 135, 259

-
- terminology theory (and practice) 123
 - terminology training 6, 61, 66, 67, 75
 - terminology work 4, 5, 19, 24, 27, 31, 32, 40, 42, 47–51, 53, 57, 61, 64–68, 72, 74–77, 94, 130, 253, 256
 - Termontography 6, 7, 137, 142, 145, 146, 181, 183, 187, 188, 192
 - Termontography Workbench (TW) 142
 - terms in context 197, 198, 209
 - Term to Text 48
 - text comprehension 7, 195, 196
 - text level and system level 47, 53
 - textometric browsing 224, 229
 - text production 7, 195, 196
 - text selection criteria 203
 - thesauri 70
 - TISS programme (terminology infrastructure for Sweden) 65
 - TNC 5, 61, 64–72, 74–77, 81–83, 85–89, 92–95, 255, 256, 258, 259
 - traditional translation dictionary 139
 - translation 1–6, 16, 21, 22, 24–33, 38, 40–42, 45–51, 53–55, 57, 58, 63, 66, 71, 74, 85, 120, 123–125, 127–131, 135–142, 144, 145, 149–151, 157, 163–168, 172, 173, 175, 178, 179, 186, 189, 191, 193, 195–198, 201, 202, 216–218, 224, 226, 230, 231, 253, 255–261
 - translation dictionaries 6, 137, 142
 - translation in terminology 26, 31
 - translation memories 163, 164, 166, 167
 - translation of specialized texts 47
 - translation of terminology between and within conceptual systems 25
 - translation-oriented terminologist 127, 128
 - translation-oriented terminology work 5, 47, 49–51, 57
 - translation policy 3, 21, 22, 24, 31
 - translation process 31, 50, 51, 53
 - Translation Studies 42, 45, 123, 256, 259–261
 - translator training 22
 - trans-national inter-lingual communication 27
- U
- unambiguousness 47, 52
 - units of language 205
 - units of meaning 205, 254
 - units of understanding 187
 - univocity principle 14
 - usage-based linguistic typology of instrumentality 233
- V
- variation 4, 7, 11, 12, 14, 31, 138, 181–187, 189, 191, 192, 197, 200, 201, 249, 250, 257
 - visual contexts 97
 - vocabulary of the Social Sciences 53
- W
- word alignment 163, 164, 166, 167, 168, 171, 172, 173, 178, 179
 - WordSmith Tools 205
 - working languages 23
 - World-Health Organization-Adverse Reaction Terminology (WHO-ART) 215, 216, 227

In the series *Terminology and Lexicography Research and Practice* the following titles have been published thus far or are scheduled for publication:

- 13 **THELEN, Marcel and Frieda STEURS (eds.):** Terminology in Everyday Life. 2010. vi, 271 pp.
- 12 **NIELSEN, Sandro and Sven TARP (eds.):** Lexicography in the 21st Century. In honour of Henning Bergenholtz. 2009. xi, 341 pp.
- 11 **FUERTES-OLIVERA, Pedro A. and Ascensión ARRIBAS-BAÑO:** Pedagogical Specialised Lexicography. The representation of meaning in English and Spanish business dictionaries. 2008. ix, 165 pp.
- 10 **GOTTLIEB, Henrik and Jens Erik MOGENSEN (eds.):** Dictionary Visions, Research and Practice. Selected papers from the 12th International Symposium on Lexicography, Copenhagen 2004. 2007. xii, 321 pp.
- 9 **YONG, Heming and Jing PENG:** Bilingual Lexicography from a Communicative Perspective. 2007. x, 229 pp.
- 8 **ANTIA, Bassey (ed.):** Indeterminacy in Terminology and LSP. Studies in honour of Heribert Picht. 2007. xxii, 236 pp.
- 7 **GÖRLACH, Manfred:** English Words Abroad. 2003. xii, 189 pp.
- 6 **STERKENBURG, Piet van (ed.):** A Practical Guide to Lexicography. 2003. xii, 460 pp.
- 5 **KAGEURA, Kyo:** The Dynamics of Terminology. A descriptive theory of term formation and terminological growth. 2002. viii, 322 pp.
- 4 **SAGER, Juan C.:** Essays on Definition. With an introduction by Alain Rey. 2000. viii, 257 pp.
- 3 **TEMMERMAN, Rita:** Towards New Ways of Terminology Description. The sociocognitive approach. 2000. xvi, 258 pp.
- 2 **ANTIA, Bassey:** Terminology and Language Planning. An alternative framework of practice and discourse. 2000. xxiv, 265 pp.
- 1 **CABRÉ CASTELLVÍ, M. Teresa:** Terminology. Theory, methods and applications. Edited by Juan C. Sager. Translated by Janet Ann DeCesaris. 1999. xii, 248 pp.

