Journal of the
International Institute for Terminology Research
– IITF –

TERMINOLOGY SCIENCE & RESEARCH

Vol. 16 (2005)

Publisher:
Address:
Redaktion: International Institute for Terminology Research (IITF)
Secretary General: Christer Laurén

CONTENTS

GENERAL FOREWORD

Volume 16 of the Journal Terminology Science and Research contains papers submitted to the editorial board by speakers at the TAMA conference, organized by TermNet in Cologne, December 2004. The title of the conference was Terminology in Advanced Management Applications. 7th International Conference on Multilingual Knowledge and Technology Transfer. The focal points of the conference and the preceding workshop were terminology for electronic applications as well as future-directed aspects of mobility: scalability, methodology, and global design.

The Board of the International Institute for Terminology Research has decided to publish this and future volumes of the Journal of Terminology Science and Research electronically via IITF's portal. It is also planned that with intervals of a couple of years, selected articles will be published in book form. The+ portal with the guidelines for authors etc. can be found under the websites of the University of Vaasa, see http://lipas.uwasa.fi/hut/svenska/iitf/. The possibility to use e-publishing will mean that improved word-processing and layout-facilities will be available.

As from the volume, we are two editors who shall be working together to compile and prepare the journal: Nina Pilke (University of Vaasa) and Birthe Toft (formerly Bertha Toft, University of Southern Denmark). Please submit articles to the editorial board via one of our e-mail addresses (see below).

Vasa and Kolding, November 2005

Nina Pilke
nina.pilke@uwasa.fi

Birthe Toft
toft@sitkom.sdu.dk

Silvia Cerrella Bauer
Head of Translation Services
SIS Swiss Financial Services Group AG
Zurich, Switzerland
silvia.cerrella-bauer@sisclear.com

TERMINOLOGY AS AN ASSET FOR KNOWLEDGE SHARING AND TRANSFER IN A LEARNING ORGANISATION - A CASE STUDY

Abstract

This paper focuses on the approach to terminology practice adopted by the in-house language service department of a medium-sized company, its aim being to highlight that terminology represents a key factor for enhancing knowledge sharing and transfer in a learning organisation and thus can improve the quality of its multilingual deliverables.

The case study hereby presented equally puts emphasis on the commitment to build a terminology database that suits not only the language service provider's needs but those of other users within the organisation as well, endeavouring to optimise a concrete internal process: multilingual document production. The approach for terminology practice adopted by the translator-terminologists is targeted at developing a corporate language based on the consensus of all internal departments involved in the above-mentioned process as one of the key measures for consistent internal (and external) communication. In this "cooperative" scenario, the translator-terminologists are called to play a pro-active role for achieving effective corporate knowledge utilisation and leveraging at the organisation level. Thereby, the use of appropriate methods and modern tools also becomes fundamental.

1 INTRODUCTION

Our case study refers to the terminology practice of an in-house language services department working exclusively for a Zurich-based holding company, the SIS Group, and its subsidiaries, which provide clearing, settlement as well as securities safekeeping and management services for Swiss and foreign banks. The group is active at international level and releases its business-related documentation in German, English and French. The members of the language services department are multi-task professionals; in other words, they assume the role of "translator-terminologists".

An organisation producing a certain amount of multilingual deliverables (paper or electronic) has to streamline its procedures, which inevitably involves a certain degree of automation as far as the production and maintenance (update etc.) of multilingual documentation is concerned.

Taking into consideration the ever increasing complexity of the company's business jargon, the translator-terminologists team acknowledged the need for creating and maintaining its own language resources that would serve internal and external information and communication purposes and would be geared to improving corporate knowledge sharing and transfer, first within the team, later across the whole organisation.

As a consequence, the team of translator-terminologists introduced in 2000 computer-assisted translation (CAT) technology consisting specifically of a translation memory (TM) tool and a terminology management system (TMS). As of then, the team started creating a multilingual terminology database (TDB) from scratch.

2 TERMINOLOGY & CORPORATE KNOWLEDGE SHARING AND TRANSFER

2.1 General contribution of terminology to corporate knowledge management

In the context of this paper, "terminology" embraces the terms - and an extensive set of data complementing them - used within the company and pertaining to its various business fields.

According to Wieden & Weiss (2004), terminology plays a pivotal role in the internal knowledge development cycle, especially in terms of conceptualisation. In fact, terminology acts as an instrument that renders concepts more tangible, allowing to distinguish conceptual properties due to a concepts transformation into language. As for knowledge management, terminology paves the way for easy access to conceptualised expert knowledge, for the standardisation of technical terms and hence for the crossing of linguistic and cultural barriers.

In the present-day information and knowledge society, a concerted approach is vital when it comes to handling acquired and shared knowledge respectively, as the way knowledge is managed and particularly the means and pace of its transmission have undergone considerable changes.

Moreover,

- the degree of specialisation is on the increase in all fields of knowledge
- a high rate of innovation has driven the need for ever more differentiated specialist vocabulary (i.e. terminologies)
- dialogue is becoming increasingly difficult, not only between experts and laymen but also among experts of one and the same discipline
- technical language presupposes precision and unambiguousness
- vocabulary errors can lead to distortions in communication and thus engender competitive disadvantages and extra costs!

Since corporate knowledge is transported via language in written and oral form, i.e. the company's technical language, a multilingual harmonisation of the technical documentation published in the various languages must be granted in globally active companies. First and foremost, systematic, conceptual and coordinated terminology work makes for unequivocal communication.

The collection, management and continual updating of the technical terms that constitute the basis of terminology practice promote the communication between the parties involved in knowledge generation and sharing, and ensure the consistency of digital products (content1). Therefore, it is fair to say that terminology practice is part and parcel of a company's knowledge management.

2.2 Specific contribution of translator-terminologists to sharing and transferring corporate "explicit" knowledge

The translation-targeted terminology practice as pursued by translator-terminologists originates a vast repository of conceptualised corporate knowledge and results from a thorough selection and classification of information contained in texts to be translated or in other corporate publications. In the course of this procedure, service, process and product-specific knowledge is gathered and processed in three languages and can be made available by means of a simple query.

As systematic collectors of multilingual, contextualised data, the translator-terminologists assume the role of knowledge owners.

They thereby deal with so-called "corporate explicit knowledge" - in other words, knowledge relevant for the organisation, which has already been documented in different forms and can therefore be adequately transferred with electronic tools.

In terms of knowledge ownership, the focus is not on the information contained in the selected terminological data, since this information existed already in other internal or external sources. The accent is more on the way the data was retrieved directly and/or re-elaborated according to specific principles and finally put together to build a "corporate knowledge repository". It is by making it available to all members of staff that the translator-terminologists contribute to corporate knowledge sharing and transfer throughout the organisation.

As highlighted in **figure 1**, knowledge sharing and transfer represents one of several interrelated activities presented in Probst's model for knowledge management (Probst *et al.*, 2000).

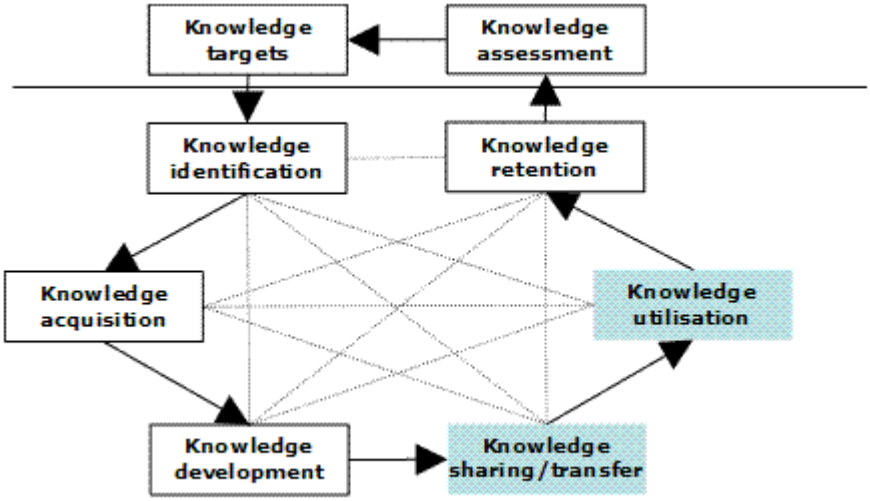


Fig. 1 - Knowledge management processes according to G. Probst et al., 2000

Since the creation and sharing/transferring of knowledge would be purposeless if the latter was not effectively used, it is of fundamental importance to make sure that the intended recipients/users of the corporate knowledge repository make systematic use of it. Knowledge utilisation, another element in the chain of knowledge management processes, also highlighted in figure 1, relates to the de-facto application of the accumulated corporate knowledge for task processing/completion and/or business processes.

It may be argued that ensuring the regular use of corporate explicit knowledge at organisation level surpasses the duties of translator-terminologists, especially in this case, where they assume the role of internal service providers but are not decision-makers as far as business processes are concerned.

Though, thorough terminology practice linked to a pro-active and cooperative approach of the translator-terminologists in their interaction with other specialists within the organisation can largely help establish best practices as far as corporate knowledge leveraging is concerned. This matter is discussed in detail in the following sections.

Considering the fact that all knowledge management processes are closely interrelated and that a change to any of them can have an effect on the others, translator-terminologists can be regarded as knowledge managers or, in other words - and as pinpointed by Peter Drucker -, as individuals "making knowledge productive" (Frappaolo 2002:24) for their specific contribution to knowledge sharing and transfer.

3 TERMINOLOGY PRACTICE IN A LEARNING ORGANISATION: A PHASED APPROACH

As previously stated, the translator-terminologists from the very outset considered terminology practice as an ongoing task that had to be systematically carried out within their "learning" team.

The targets were clear: the continuous enlargement of the multilingual TDB leading to terminology consistency and therefore higher translation quality. At a broader level, this new ongoing assignment in the translator-terminologists' task portfolio was geared to promote knowledge sharing and transfer not only at a team level, but organisation-wide as well.

Furthermore, the translator-terminologists ambitioned to apply this phased approach to their interaction with the business specialists from the various internal departments, especially those involved in the production of multilingual documentation. The translator-terminologists nurtured the vision of promoting a cooperative and interactive working environment, so that they considered themselves as part of a "learning organisation" in which all members would view its success in the future as being based on continuous learning and adaptive behaviour2.

For this reason, the translator-terminologists wished to promote awareness of the added value that terminology represented for the whole organisation as a means of achieving more efficiency in multilingual document production with the aim of increasing the multilingual deliverables' quality and, consequently, productivity.

As a matter of fact, once the multilingual TDB is made available to all members of staff, its regular use by the latter will also contribute to establishing a consensus-based corporate language. Subsequently, the organisation's external image will be strengthened as well.

For achieving it, the translator-terminologists adopted a phased approach. Clear targets were defined, specific measures taken and resources allocated for the various phases, the last of which - still in progress - is aimed at maximising knowledge utilisation within the framework of specific tasks completion.

Figure 2 illustrates the given timeframe for completion of the various phases, which are explained in detail in the following sections.

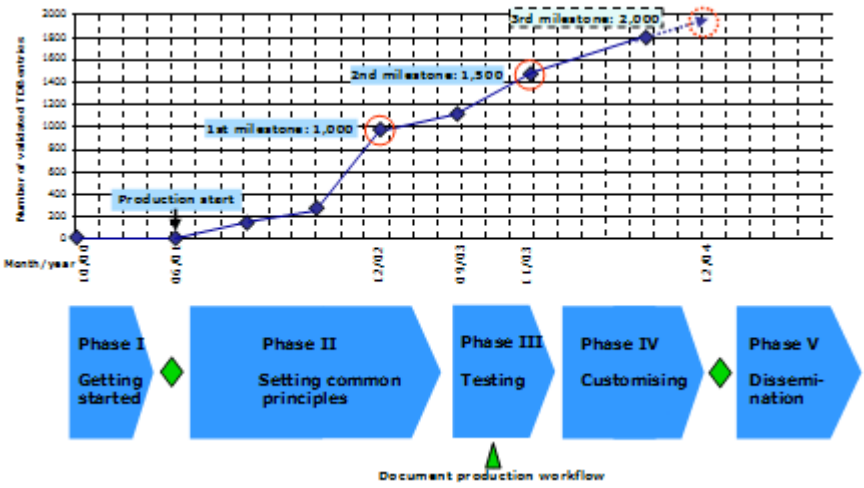


Fig. 2 - Phases in terminology practice: from beginning till present

The actual timeframe for each of the described phases is not of much relevance in the present case, since the translator-terminologists carry out terminology-related tasks on a best-effort basis; in other words, when they have the time and no pressure to deliver translations.

3.1 Phase I: Getting started before terminology capture

Before actually starting capturing terminology, the translator-terminologists defined terminology targets for knowledge sharing and transfer purposes at team level.

In this first phase, the main target was to define the most suitable TDB structure and to ensure that the first entries covered the most relevant terms of the organisation's business areas, especially the official denominations of products and/or services - though without specifying a threshold number of entries to be reached. Ensuring this was not difficult, since the translator-terminologists entered in the TDB mainly the terms contained in the texts to be translated and their equivalents in the other languages.

Furthermore, the translator-terminologists carried out a certain number of preliminary tasks for optimum assignment steering, such as:

- Build-up of a document corpus (representative texts, magazines, specialist publications etc.)
- Consultation of specialists within the organisation for defining the subjects to be covered
- Revision of the subjects
- Modelling of the terminology database including the definition of a model for data input
- Processing/revision of an existing glossary of some 550 records
- Import of the glossary into the TDB
- Assignment of responsibilities and organisation of terminology practice

During this phase, lasting a good six months, particular attention was given to setting general guidelines for future terminology practice; in other words, terminology capture, extraction, validation, update, etc. These guidelines for professional terminology management were aimed at guaranteeing or, at least, largely contributing to overall consistency of the terminological data.

3.2 Phase II: Setting common principles for terminology practice

The translator-terminologists gave themselves the following three guiding principles for terminology practice, geared to developing the TDB:

- **Quality comes before quantity, with special emphasis on systematic update, thorough validation and exclusive use of well-established sources of terminological data[3]**

The quality of the terminological data is more important than its quantity, whereby all the subjects represented in the TDB are to be covered by the relevant entries and the concepts and/or terms entered correspond to the currently used specialist jargon.

Furthermore, the so-called "6-eye principle" is applied: Each entry is proofread by two translator-terminologists other than the one who created it. Particular attention is also given to selecting only reliable sources and confirming content veracity by multiple source comparison, where necessary.

- **Terminological data has both descriptive and prescriptive character and is consensus-based**

The TDB is to become the official terminological reference for the organisation as far as corporate language is concerned. In order to establish a given corporate terminology, certain terms are defined as preferred whereas others are considered as synonyms or even disregarded, as a result, for example, of internal use considerations, language spelling conventions, etc.

Last but not least, and since the TDB is meant to be a tailor-made corporate knowledge repository, the translator-terminologists address internal field specialists when it comes to defining certain complex terms.

- **TDB languages are handled on an equal footing**

To be regarded as completed and thus undergo the validation process, the TDB entries must contain all relevant mandatory concept-related and/or term-related data in the three handled languages (German, English and French). As each of the translator-terminologists in the team has sufficient general know-how to enter data in the three languages and is involved in the validation process according to his specific language competencies, only one model for data input has been defined.

For consistency's sake and as an instrument for knowledge sharing and transfer at team level, the translator-terminologists resolved to write what could be called a "white paper" for terminology practice. This regularly updated and/or enlarged compendium of all the knowledge continuously acquired through terminology practice. The handbook is richly illustrated with some 108 extracts of validated TDB entries and 9 complete TDB entries. Through its 125 pages, it comprehensively highlights and defines the following aspects of terminology practice related to the TDB:

- Use and scope of all data categories and data elements
- General procedures for entering terms
- Handling of the most frequent problems of multilingual terminology practice (e.g. overlapping or partial equivalence of term meaning in the different languages; terminology gaps; accepted non-synonymous designations for signifying a term due to usage reasons; mixture of different language registers to signify specific terms etc.)
- Applied methods (e.g. procedure for entry validation, TDB management and planning, including statistical monitoring of terminology production)
- Complementary useful information (e.g. language and bibliographical codes, accepted format of internal and external publication titles and details used as a source in the TDB entries)

This second phase came to an end when the first milestone in terminology practice was reached: The TDB contained its first validated 1,000 multilingual entries.

3.3 Phase III: TDB's user acceptance test

Knowledge has to be incorporated in an action by a human being to be considered as such: It would be pointless to maintain an instrument for knowledge sharing and transfer without the intended recipients using it! By letting key staff members test the TDB, the translator-terminologists also sought to raise awareness of the overriding role a consensus-based corporate knowledge repository plays for improved internal knowledge exchange, thus increasing the quality of multilingual deliverables.

3.3.1 Scope of TDB test

In view of the declared goal of building up a corporate knowledge repository, it was important for the translator-terminologists to know whether they were on the right track as far as the adopted approach to terminology practice from the perspective of other potential TDB users was concerned.

Therefore, as soon as the TDB contained 1,000 validated entries, the time was ripe to assess TDB acceptance as a future corporate knowledge repository by letting key staff members test it.

More precisely, the purpose of the test was to measure:

- the quality of the terminological data in the test users. view' especially to find out whether the displayed data was sufficient or too comprehensive, whether the collected data corresponded in most cases to the specialists' understanding of the relevant terms and, finally, whether the organisation's specialised terminology was actually represented (measured in percentage of successful search results);
- the degree of user-friendliness offered by the current graphical user interface (data display design, need for user training, etc.)
- the performance of the terminology management system (stability and search response time).

Moreover, the test users were requested to comment on any other requirement they wanted to be given consideration to for the future TDB development.

The test ran during a whole month, after which the translator-terminologists gathered and classified the test users' feedback on the above-mentioned criteria. As a matter of fact, their feedback confirmed that, by continuing terminology practice with the current approach, terminology would have undoubtedly contributed to improved knowledge sharing and transfer organisation-wide. The translator-terminologists

committed themselves to considering as much individual input as possible from test users for the read-only TDB version to be implemented on the organisation's Intranet in future.

3.3.2 Multilingual document production as the underlying business process for determining test user selection

The translator-terminologists decided to address especially the staff members involved in the multilingual document production process, since it was clear that the latter would be the prospect users of a corporate knowledge repository such as the TDB for their task completion.

The current workflow for the production of one of the most important multilingual documents released by the organisation, the Guide to Market Information (hereinafter "Guide"), was analysed in order to ascertain potential enhancements, one of which would clearly be regular TDB query by the various parties involved.

In end effect, the latter can be regarded as "knowledge carriers" from different areas since a large percentage of terminological data in the TDB is based on the information contained in the various texts delivered for translation by them.

Figure 3 shows the main sections and the length of the Guide as well as the various internal departments giving input for its production and update.

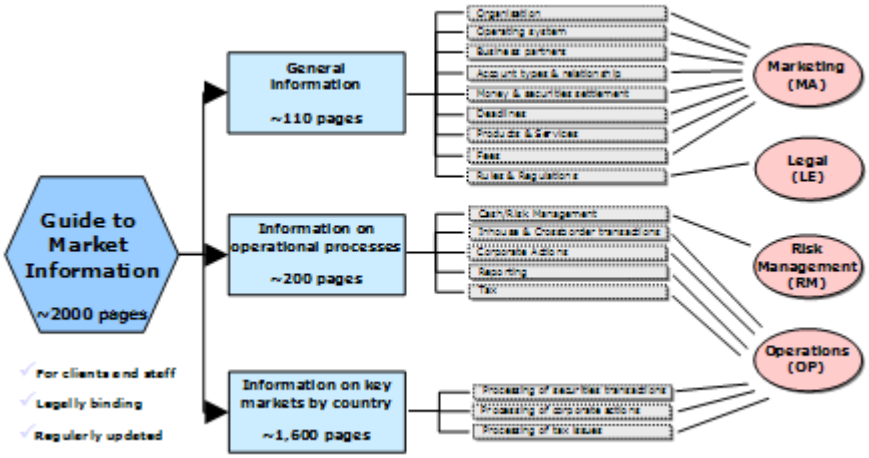


Fig. 3 - Guide to Market Information: volume, sections and internal departments involved in its production

The Guide is published on the organisation's Internet and regularly updated. The intended readers are not only the organisation's clients worldwide but also the employees in charge of processing and providing reliable information to partners and clients outside the organisation. The Guide contains a comprehensive description of the products and services provided by the organisation as well as information about the operational procedures related to the single business processes and to the individual market specifications. A considerable amount of information undergoes regular adjustments or is only valid for a short period of time and therefore needs to be replaced and updated on short notice. The Guide consists of three main sections of a total of over 2,000 printed pages and has legally binding character for both the company and the clients.

Given the variety and significance of its content, input on single sections of the Guide is provided by as many as 11 units within the marketing, legal, risk management and operational issues departments. Key representatives of these departments were thus invited to test the TDB.

3.3.2.1 Current document production workflow

The current workflow for the production of the Guide consists of a processing sequence involving other units as well, as represented in **figure 4**.

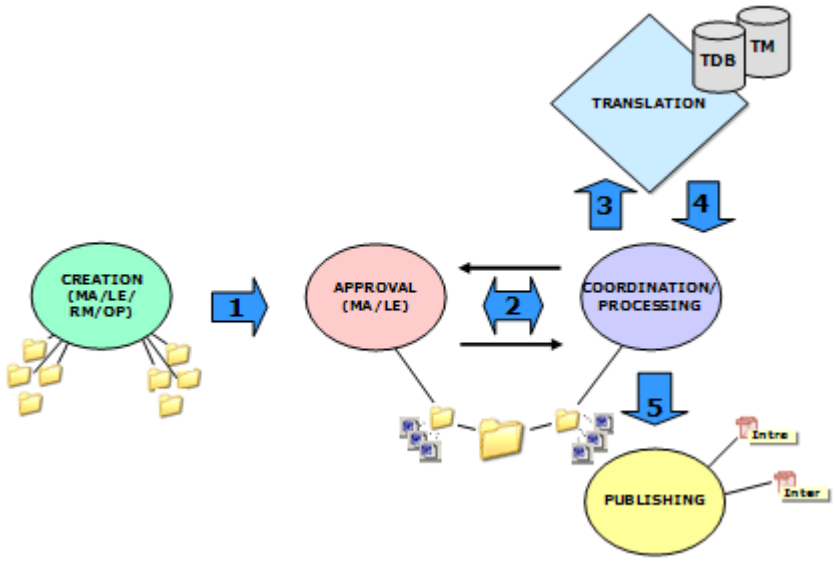


Fig. 4 - Workflow for the production of the Guide to Market Information

Once an author has drafted a text, it is passed on to the instance responsible for checking legal compliance and content truthfulness; the text is corrected where necessary. After this first approval, the process continues with the intervention of the coordination/processing instance that is also in charge of checking language correctness in the text's original version (mostly in German language). If any adjustment is made, the text goes back to the previous instance for renewed approval. Only then, the text is passed on to the translator-terminologists, who deliver the translation by using CAT technology and consulting the TDB for terminology checking. The translated versions of the original text are again submitted to the coordination/processing unit that often passes them on (but not systematically) to the approval unit for checking by specialists who are native speakers of the respective languages. If any adjustment proposal is submitted to the translator-terminologists, the latter approve or reject them and - whenever necessary - immediately proceed to data updating in the translation databases and/or the TDB.

Although the workflow has been designed to ensure proper quality control, the target of making key business information available to staff and clients timely is, unfortunately, not always achieved due essentially to terminology and/or content inconsistencies at source. Moreover, authors lack specific training on technical writing and no controlled language methods are applied. The mentioned shortcomings generate continuous document changes and adjustments in the process and in most cases a renewed approval by the single units becomes inevitable. As a matter of fact, most misunderstandings occur due to the lack of a common consensus-based terminology.

Furthermore, the absence of a common file storage repository results in inefficient document management and makes it difficult for the different units, of which each has its single folder setup, to swiftly get to the relevant document version. Since several units within the different departments are solely responsible for individual sections or sub-chapters of the Guide, the necessary updates are not simultaneously carried out throughout the entire document, which results in content contradictions or outdated information in single document sections and, in many cases, even in invalid - sometimes even inexistent - references to terms and/or titles.

As the last link in this long chain, the translator-terminologists are basically the only instance able to ensure overall consistency: by dint of their TDB.

3.3.2.2 How to optimise the document production workflow

This process analysis clearly highlights that document quality does not exclusively depend on the output from the translator-terminologists.

Appropriate input at source; in other words, no terminology inconsistencies, use of precise wording and broad consensus as to content are elements that accelerate the workflow and help improve the quality of multilingual deliverables.

The test users unanimously voiced their interest in availing of the TDB as a working tool that would ease preparing and checking information to be delivered for the production not only of the Guide but also of any other document. In this manner, current and validated information could be re-used for several purposes.

The test users also requested to be able to give feedback as to the TDB entries (entry of terms not included in the TDB, adjustments of term-related entered data, etc.) to the translator-terminologists in future. This cooperative approach corresponded to the translator-terminologists. vision of setting priorities for fostering corporate language together with the specialists of the relevant fields.

Acknowledging the need to avert duplication, the test users even agreed to charge the translator-terminologists with the review, adjustment and import of relevant data contained in existing glossaries individually build up and stored by the various departments, so that the TDB could become the only terminological reference for all staff members.

These conditions being met, the concept of the learning organisation could be materialised in this process. The growth of the corporate knowledge repository would no longer be the exclusive responsibility of the translator-terminologists but of all parties involved in the dissemination of corporate information; the translator-terminologists provide their expertise, whereas the experts ensure data truthfulness and its effective use. The "lessons learned" could be an asset for better understanding and steering future processes.

It goes without saying that other departments involved in the production of other multilingual documents face similar challenges when it comes to using a consensus-based corporate language to disseminate information outside the organisation.

Making the TDB available to all and implementing a suitable single-source document repository such as a document management system (DMS4) (see **figure 5**) would lead to:

- better content quality already at source
- validated unambiguous and unequivocal language
- consistent and consensus-based terminology constant reuse of approved information

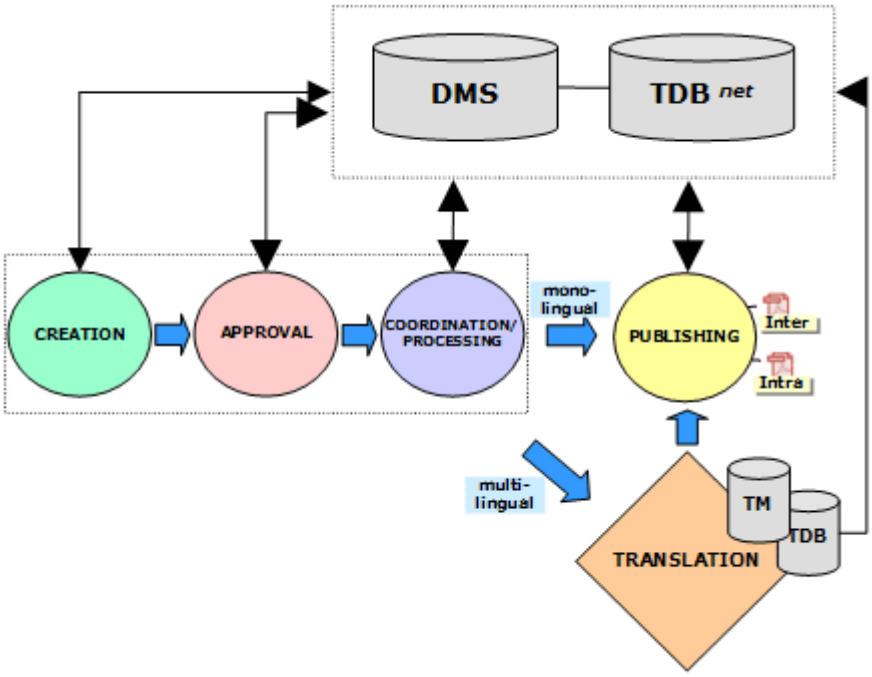


Fig. 5 - Optimised workflow for document production

More efficiency could be attained for both monolingual and multilingual document production workflows: The coordination/processing unit would come into play only for the production of documents in their original version; for the production of the foreign language versions, it would become superfluous since the units responsible for the content creation could get back directly to the translator-terminologists by means of the DMS version control functionality, should any adjustment be necessary. The coordination/processing unit could then devote more of its time to other, more valuable tasks commensurate with its specific language and computer skills.

Cooperation among the different "knowledge carriers" resulting in an optimised resource allocation would obviate the need to change and/or adjust the relevant document versions at different stages and result in much shorter production processes in a short- to medium-term time span. Circulating and leveraging corporate knowledge at organisation level would consequently contribute to higher productivity.

3.4 Phase IV: Customising the TDB's graphical user interface and on-demand bulk terminological data integration

3.4.1 Customising the TDB's graphical user interface

Most of the test users. feedback concerned the TDB.s graphical user interface (GUI). Therefore, customising a truly user-friendly GUI was very important to ensure that future users could feel at ease querying the TDB on the organisation's Intranet and, consequently, use it regularly.

Customising the TDB's GUI to the users' needs meant essentially to provide:

- an easy search function consisting of two fields with dropdown menus for defining the source and target language and a field for term entry. Any other search criteria such as subject etc. to restrict the search were not regarded as necessary; a fuzzy search feature was, on the contrary, considered as a must.
- a "compact" terminological data display in order to avoid unnecessary scrolling, whenever possible: For instance, by displaying only relevant data elements without displaying the corresponding data categories, whenever the latter were self-explanatory.

- two different data display layouts: A default one appearing immediately after any successful search and displaying only the basic entry data such as concept-related data categories with the relevant data elements, all designations in the three languages and the concept definition in the source language chosen by the user; an optional one displaying all entry data. The latter would appear only in case of the user wanting to extend the search.
- an automatic e-mail generation feature to send a request to the translator-terminologists for requiring/suggesting adjustments to terminological data on a specific entry.

For the translator-terminologists, another customising goal was to ensure no user training for TDB querying would be necessary. Apart from a simplified search function as described above, the TDB's read-only version had to comply with the organisation's corporate identity guidelines and design. Furthermore, some navigation links to useful TDB-related information were added to its Intranet page, such as:

- the number of available terminology entries (this figure has to be updated as frequently as possible);
- a disclaimer;
- all navigation links to the pages containing useful information for the readers such as: A description of the scope, content and languages of the terminological data collection (including applied internal guidelines for each of the handled languages etc.), a description of the entry structure and entry fields, a list of the available subjects, the TDB's facts and figures (development, statistics, etc.) and, last but not least, a detailed description of the search feature (data displays, fuzzy search, etc.), definitions of terminology, terminology practice and the relationship between terminology and knowledge management.

Furthermore, a link to the translator-terminologists mailbox is provided for any terminology-related request.

3.4.2 On-demand terminological data integration

Following their cooperative approach and their commitment of avoiding the proliferation of disparate glossaries on the organisation's Intranet, the translator-terminologists immediately included the integration of data from existing glossaries in the TDB among their priorities for terminology production and started reviewing and adjusting it.

In this process, they found out that approximately 60% of all data had already been entered in the TDB by them. This confirmed the adequacy of the choices they had made when it came to defining the terms to be entered first, before effectively starting with terminology capture. As for the remaining eligible data, the translator-terminologists committed themselves to fully import it in the TDB as soon as they could before making the TDB available on the Intranet.

3.5 Phase IV: Organisation-wide dissemination of the terminology collection

3.5.1 Timing for the dissemination

This challenging phase started in January 2005. At that moment, the translator-terminologists. third milestone in terms of terminology capture, 2,000 validated entries, was reached. Although the TDB is only in its fledgling stages and, in quantitative terms, does not stand comparison with other similar products managed by large organisations, the number of entries that it contains - though modest - is a sound basis to make it an efficient search and working tool for all staff members.

The TDB is likely to prove useful since terminological data closely related to the organisation's business activities was given preferential treatment for entry and is in line with linguistic practices of internal and external specialists. Two other major reasons underpin the need to release the collected data and to put the TDB at a larger user group's disposal:

- Timeliness of the terminological data
- Validity of spelling rules, especially in the German language

In an ever changing business environment such as the financial one, terms and/or data related to them can become outdated after a relatively short time (e.g. a company name changes as a result of a merger or a reorganisation; or a product name is replaced by another one since a company no longer supports the former version and/or applications; e.g. a company is no longer active in a certain business field or a programme or application replaces functionalities naming; or SWIFT compliant messages change designation, etc.).

In addition to that, the introduction of the German language reform represents another stumbling block for translator-terminologists when it comes to defining adequate spelling in order to ensure term consistency; this holds true especially for the Swiss German business environment where the tendency to reuse English terms, sometimes hyphenated, sometimes uppercased (whereby spelled exactly as in the original language), prevails. Since authors do not have common spelling guidelines, translator-terminologists have to cope with repeated inconsistencies and assist authors with a sensible and consistent spelling alternative that they then reflect in the corresponding TDB entries.

The choice of an appropriate communication strategy to entice potential users to acknowledge the existence of the new tool is also crucial: The translator-terminologists have already written the relevant announcement - a concise message in the form of a pop-up-window - to appear on the Intranet homepage for a period of about two weeks. By telling the users that their involvement in the process of defining corporate language is welcome and valued, they expect to appeal to the staff members' curiosity and have them click on the corresponding link to the TDB and subsequently become active users of this corporate knowledge repository. The chain reaction will hopefully result in a productivity increase over time, leaving the staff members with time for more value-added tasks.

3.5.2 Next steps

It goes without saying that the translator-terminologists. phased approach to terminology practice does not end with this phase. The next step they commit themselves to make is to assess to which extent the product "TDB" will effectively contribute to knowledge sharing and transfer for knowledge seekers that wish to extract knowledge from this corporate knowledge repository for personal purposes (internalisation) or for disseminating it in the form of deliverables to a larger recipients group (externalisation).

The first aspect will definitely be more difficult to assess since it has to do with the understanding and acknowledgement of the organisation's language in the minds of all staff members - including those with whom the translator-terminologists do not maintain any (regular) working relationship. The second aspect could be appraised by analysing how much more efficient the processes for which the terminology collection represents a supporting tool became within a given timeframe.

In order to monitor the number and kind of queries of the TDB, the translator-terminologists intend first of all to implement a statistic feature in the TDB Intranet site. In addition to that, they will gather and classify all users' requests and keep the latter informed about their status (also via the TDB site) as well as prepare a user survey on the TDB.

The timing for initiating this assessment phase will depend on the reaction from users. In all likelihood, however, a whole year will elapse from the point of terminological data dissemination until assessment becomes viable.

4 HOW TRANSLATOR-TERMINOLOGISTS CAN CONTRIBUTE TO CORPORATE KNOWLEDGE UTILISATION IN A LEARNING ORGANISATION

By using translation-related and terminological knowledge (on the human side) and the appropriate infrastructure (on the technical side), it becomes possible to share a sort of ready-to-use knowledge with the company staff. This fact allows the continuous repurposing of certain or all contents for purposes they were not originally intended for.

In the presented corporate setting, terminology becomes an asset for knowledge sharing and transfer. As a matter of fact, the described multilingual document production process, as well as other ongoing processes such as the localisation of a proprietary software, effectively demand the use of a corporate knowledge repository such as the TDB. Particularly in this respect, the translator-terminologists seek to become real knowledge facilitators by making specific knowledge available on demand.

As far as human intervention is concerned, the translator-terminologists envision contributing to set the grounds for a more cooperative environment, similarly to a learning organisation in which practices, culture and systems promote the continuous sharing of experiences and lessons learned (Frappaolo 2002).

According to Garvin (1993) a learning organisation has the skills to create, acquire and transfer knowledge and to change its behaviour based on the new knowledge and insights gained. By continuing to adopt a pro-active approach in their relationships with other agents across the organisation, the translator-terminologists will seek due recognition as central language regulators for communication processes and, more specifically, as qualified partners of the business specialists for language and terminology matters. The ultimate goal is to contribute - even if minimally - to the cultivation of an open attitude towards changes and new approaches in the minds of the other agents in the organisation as well.

In order to give input on how existing or future gaps could be filled and on how the use of controlled languages or machine translation tools for future complex processes demanding a higher degree of automation could be integrated, the translator-terminologists will need to keep up-to-date with the latest trends in language processing solutions and techniques. They will thus promote active TDB utilisation.

5 ACKNOWLEDGMENTS

I would like to thank my team colleague, Hans Martin Jörimann, certified conference interpreter, for his valuable support in proofreading and lay-outing this paper.

[1]"Content" as a technical term refers to material (data) saved in one or several media to support knowledge acquisition and transfer (e.g. content on Websites or other digital media as manuals, reports, internal and external corporate publications, etc).

[2]http://www.nelh.nhs.uk/knowledge_management/glossary/glossary.asp#I

[3]According to C. Frappaolo (2002:10-11), the main challenge when facing explicit knowledge is "to manage its volume, ensure its relevance and quality, and make it easily accessible - in a phrase, handling infoglut".

[4] A DMS is a software system based on an underlying database, in which unstructured objects (i.e documents) are indexed and tracked. A DMS monitors security, logs access to files and maintains a history of file content (Frappaolo 2002).

6 REFERENCES

CERRELLA BAUER, S. (2004): Die Übersetzer-Terminologen als Wissensmanager. In: Hieronymus 3/2004, term bulletin of ASTTI (Association suisse des traducteurs, terminologues et interprètes), Bern, 3-16.

CERRELLA BAUER, S. (2003): Handbuch für die Terminologiearbeit mit SISTerm. Internal document, SIS Swiss Financial Services Group AG, Zurich, 125 p.

CERRELLA BAUER, S. & E. YUSTE RODRIGO (2004): Circumstances, challenges and consequences of a quality-geared and technology-aided process of translating: a case study. In: Hieronymus 2/2004, term bulletin of ASTTI (Association suisse des traducteurs, terminologues et interprètes), Bern, 42-54.

CHAN FATT CHOW, D. (2002): Is Your Company a Learning Organisation? School of Information & Communication Technology (ICT), http://www.ict.np.edu.sg/icthome/ictweb-2002/publications/Articles/LearningOrganisation.htm

FRAPPAOLO, C. (2002): Knowledge Management, Capstone Publishing, Oxford.

GARVIN, D. A. (1993): Building a Learning Organization, Harvard Business Review, July-August, 76-91.

ILCHMANN, C. (2003): Investigation of Technologies in Knowledge Management Context. School of Computing . Dublin Institute of Technology, http://www.comp.dit.ie/rfitzpatrick/MSc_Publications/2003_Carsten_Ilchmann.pdf

Ruth Feil & Lotte Weilgaard Christensen
Department of Business Communication and Information Science, Kolding
University of Southern Denmark
Denmark

THE ThT SYSTEM - A MULTILINGUAL NORDIC SEARCH INTERFACE

Abstract

The aim of this paper is to present the results of a project the purpose of which is to develop and test a prototype of a multilingual Nordic search interface which may, in combination with existing search engines, give monolingual access to information in several Nordic languages at the same time. The project is carried out by a Nordic research network called NorNa (Nordic Navigator). The languages involved are Norwegian, Finnish, Swedish, Danish, and English, which allows subsequent addition of non-Nordic languages. The ThT system (th standing for thesaurus and t for term), which has been developed specifically for this project, is based on a term bank application provided with additional modules such as a module containing search concepts. The data on which the project is based has been obtained via terminological analysis of a number of parallel corpora, one of them in the domain of the Arctic environment. The overall aim of the project is making the results obtained via research directly useful to trade and industry as well as public institutions in the form of a language technology tool which may contribute to optimum utilisation of Nordic knowledge resources. The project has received financial support from the Nordic language technology research programme of NorFA, the Nordic Academy for Advanced Study of the Nordic Council of Ministers.

1 INTRODUCTION

The aim of this article is to present the approach of a Nordic research network to multilingual information retrieval in the Nordic languages. The purpose of our network project is to develop and test a prototype of a multilingual Nordic search interface which may, in combination with existing search engines, give monolingual access to information in several Nordic languages at the same time.

The background for the NorNa project is that we in the Nordic countries – or at least in the Scandinavian countries – can understand each other even though we speak different languages. To most of us it is not unusual that non-Nordic people react with surprise when overhearing Scandinavians in a lively discussion in which all participants speak their respective mother tongues. The starting point for our network has thus been that a Dane may ask a question in Danish and get the answer in Norwegian. A Nordic search engine should be able to do the same.

The network has received financial support from the Nordic language technology research programme of NorFA, the Nordic Academy for Advanced Study of the Nordic Council of Ministers.

At our first network meeting in September 2002, we decided that the name of our network should be NorNa which stands for Nordic Navigator. Later we discovered that Norna also is the name of a rare orchid growing in the northern parts of Sweden. We have therefore made this orchid our logo.

2 APPROACH TO MULTILINGUAL INFORMATION RETRIEVAL

The participants in the network all represent institutions focusing on specialist communication, i.e. the Norwegian School of Economics and Business Administration in Bergen and the University of Bergen (Norway), the University of Vaasa (Finland), the University of Southern Denmark, Kolding and the software developer TERMplus ApS in Copenhagen (Denmark), including observers from Stockholm (Sweden) and Ventspils (Latvia). The participants all adhere to Wüsterian terminological principles in one form or another.

This means that we have a very strong focus on LSP texts and that our multilingual approach is embedded in an onomasiological perspective. The fact that we use a concept-based method is an obvious consequence of our need to find texts which describe the same concept in different languages – rather than the different senses of a given lemma.

Our overall approach has been very pragmatic. We have chosen to define a somewhat restricted task for our project work, as reflected in the fact that our documentary basis is quite limited and at the same time clearly defined. On the one hand we want to make sure that the texts we analyse are exemplary, and on the other hand we do not want to end up with a too comprehensive material given the time and the resources available for the project.

The title of this paper begins with the phrase 'ThT system' where 'th' stands for thesaurus whereas 't' stands for term. We have chosen this name to indicate that together with terminology, I&D (Information and Documentation) plays a crucial rôle in our project. Thus the letters ThT express the basic idea behind our approach: not only is the termbase we are constructing based on documents; it is also intended to function as a multilingual retrieval system in which the terms serve as search terms.

When it comes to software the aim has been to develop an exemplary prototype with a view to future product maturing. For this purpose we have decided not to develop new software from scratch and instead chosen to:

- either start from existing software and develop it further
- or exploit facilities of existing software to meet our needs

The central piece of software is a further development of TERMplus, a terminology management system originally named DANTERM for Windows. The use of existing software in our project includes in particular modules in MSWORD and MSACCESS.

The Nordic languages that have been analysed are Swedish, Norwegian Bokmål (one of the two official Norwegian languages), Danish, i.e. the Scandinavian languages, as well as Finnish, since the Finnish network group in Vaasa has native speaker competence in both Swedish and Finnish; English is also included, initially thought of as 'pivot language'. However, in our onomasiological perspective it is more correct to designate the concept as the pivot, which implies that all languages are equal. English has rather come to be viewed as a "bridge language" in the project, since English is used in order to facilitate communication with people outside the Nordic region.

3 WORKING PROCEDURE

The overall working procedure for developing a fully implementable search interface consists of the following stages:

- the establishment of parallel specialist corpora / full-text databases
- the establishment of a classification
- the establishment of an index with search concepts in one language followed by the inclusion of equivalents in other languages
- the development of a search interface for the selected domain

3.1 Establishing corpora

Three corpora have been included in our project. In the initial project phase the network group worked with a small test corpus consisting of a TV manual, whereas our two main corpora are:

- a document on the Arctic environment in the Nordic countries. The text was selected 1) because the documentation was available in all project languages AND 2) at the same time the Arctic region is of central concern and a very topical area in the Nordic countries. The Danish network participants are "primary providers", i.e. have taken on primary responsibility for this subcorpus.

- Nordea's Year-end report. This domain has been selected since several of the project participants have specialised in economic-administrative terminology. The Norwegian School of Economics and Business Administration has for instance initiated the establishment of a knowledgebase within these domains. Here the Norwegian network participants are primary providers.

These two subcorpora have been provided with header information and basic XML mark-up.

The size of the two main subcorpora can be tabulated as follows:

The Arctic regions

Danish	65,000 tokens
English	72,000 tokens
Swedish	60,000 tokens
Norwegian	63,000 tokens
Finnish	45,000 tokens

Nordea's year-end report

approx. 9,000 – 10,000 tokens in each language

In the text on the Arctic regions, the word count is roughly equal for Danish, Norwegian and Swedish, while the number of tokens in Finnish is markedly lower, with a difference of 20,000 tokens between the Danish and the Finnish version.

Nordea's Year-end report consists of texts of approximately 10,000 tokens. This is not a very large number of tokens. Still, it is near the magic figure of 10,000 described e.g. by Khurshid Ahmad and Margaret Rogers (1994) at the University of Surrey as being sufficient to conduct some forms of LSP corpus research, such as term extraction.

3.2 Establishing classifications

When the subcorpora have been established, it is necessary to analyse and determine the content structure in the texts to develop a classification. Only then can the actual work to establish an index of search concepts for the various subcorpora begin.

The NorNa classification is based on the specific content structure of the texts. This content structure is revealed by the table of contents or, in the case of the Arctic subcorpus, by the preface, which proved to be a valuable source of inspiration for working out a classification. The latter is a strong indication of the document-based approach we have applied.

The classification of the text on the Arctic regions has resulted in the structure shown in fig. 1.

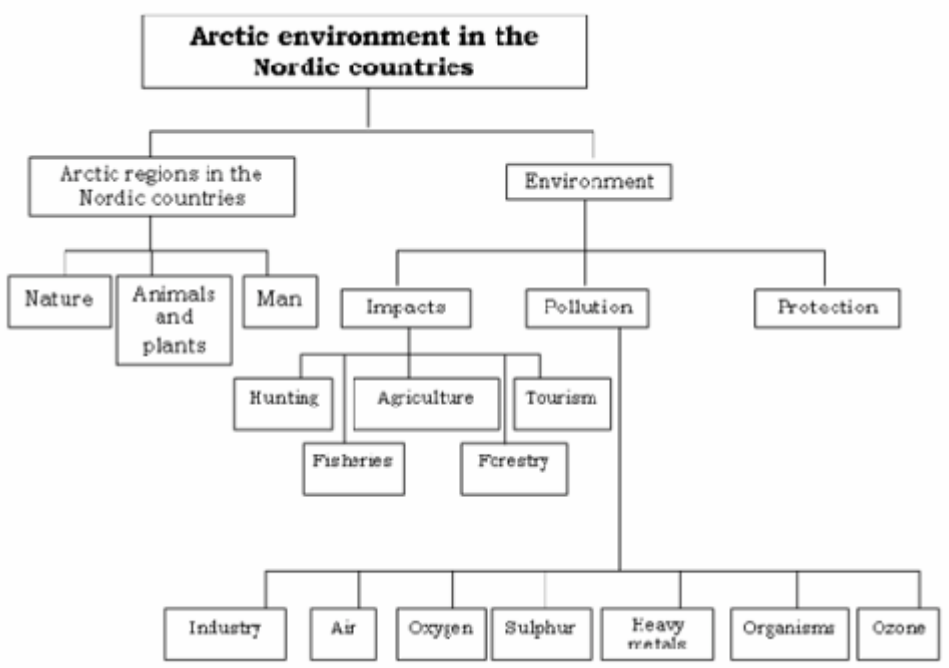


Fig. 1

3.3 Establishing indexes with search concepts

The indexes consist of search concepts represented by the terms in the five languages. The process of establishing indexes includes the following stages:

- An interactive extraction of possible search concepts (using WORD index as a tool)
- An interactive systematisation of these concepts (using a module called TREE VIEW as a tool)
- The search for equivalents (using five-language alignment as a tool)

Our extraction of search concepts is based on an interactive extraction of potential search concepts, for which we have used WORD's index facility. Possible search concepts are marked manually, and then WORD automatically marks all matches of the given string of words in the respective corpus text. The search concepts will then be grouped automatically in an alphabetical list at the end of the process. The extraction stage thus results in an alphabetical index.

The task of the primary providers of a subcorpus has been to extract search concepts in their own language(s) and distribute the list of these search concepts to the other groups, i.e. the above mentioned alphabetical index. Thus one language will be used as the point of departure when the index is established.

When an alphabetical index has been prepared for one of the languages, the extraction of equivalents in the other languages may begin. This is done by aligning the five languages. To aid the extraction of equivalents, an alignment of the texts from which the search concepts have been selected is used, that is, either the Arctic texts or the Nordea texts in all five languages.

In our terminology a search concept is represented by a term, which means that when search concepts are extracted from texts, this selection will be in the form of terms.

The five-language alignment is illustrated in fig. 2.

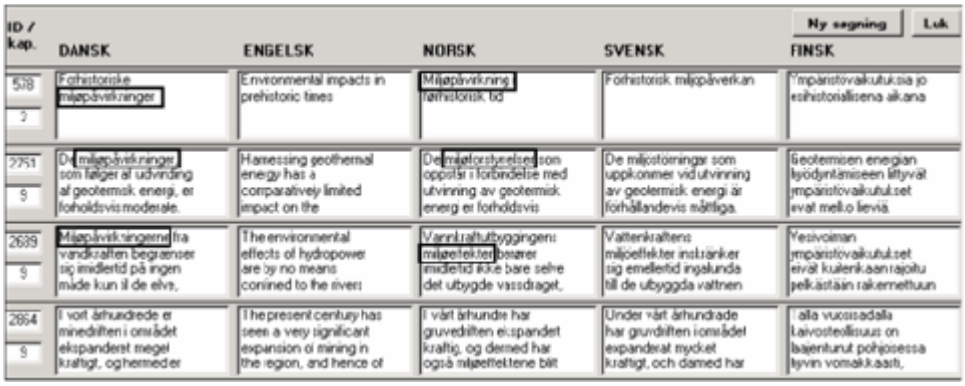


Fig. 2
The figure shows a search into the five-language alignment, which has been prepared in an ACCESS database. The search is made for the search concept environmental impact in Danish. All sentences which include this search concept are displayed together with aligned sentences in the other languages. In this way it is possible to find the equivalents in the other languages. Moreover, the method in question has proved extremely useful when it comes to recording synonyms. The example illustrates how synonyms may be recovered by means of the aligned texts. As shown in the aligned Norwegian sentences, there is a terminological inconsistency in Norwegian – in this example alone the concept environmental impact is translated in three different ways in Norwegian. It should be mentioned, though, that the translation miljøeffekt is strongly influenced by English and is considered untypical of Norwegian.

The list with the terms in all five languages is then imported into the terminology management system TERMplus, which has been enhanced with a module called TREE VIEW. In TREE VIEW it is possible to 'drag and drop' the search concepts manually to establish a hierarchical structure.

In fig. 3 an index of the conceptual structure of the text on Arctic regions is shown. This structure corresponds to what was shown in fig. 1.

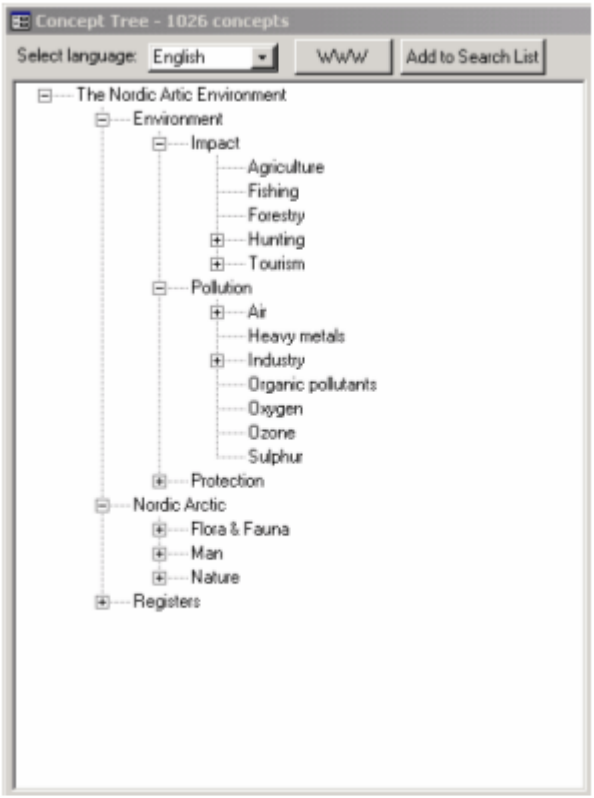


Fig. 3

3.4 Development of search interfaces for the selected domains

Based on the alphabetical index a simple index structure is built in the system, more specifically in the module named TREE VIEW in which the conceptual relations will simply appear as hierarchical, i.e. without any specification of the nature (generic or partitive) of the relation, which means that neither causal nor any other type of relations will be revealed in the structure. The relation type will, however, be specified in the termbase, to the extent that definitions or other content-based information on the search concepts have been recorded in the termbase.

In TREE VIEW it is possible to select the other four languages as well so that the above index may be shown in for instance Finnish.

4 SEARCH INTERFACE

In our opinion, it is an absolute requirement for the search interface to allow the user to identify the search concepts by widening or narrowing searches horizontally and/or vertically.

The **horizontal** dimension involves a single concept, to which one or more terms may be associated in each language, in addition to orthographic variants, i.e. a synset. The example given in fig. 4 shows the search concept waste management in English together with its equivalents in the Nordic languages.

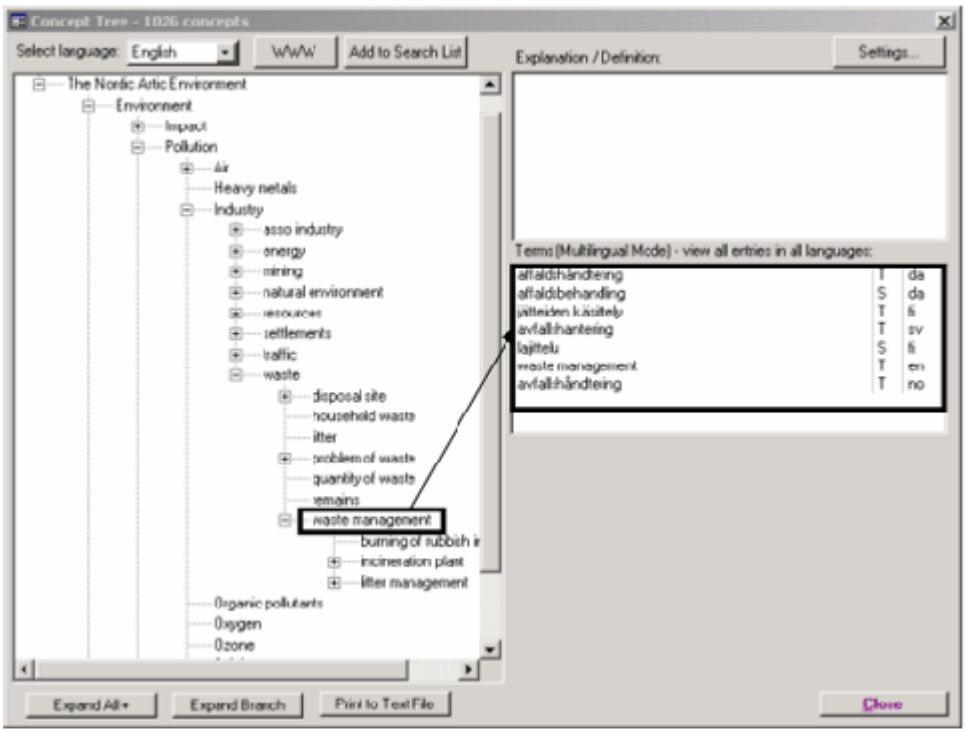


Fig. 4

Vertical identification implies that the searches may be enlarged by including superordinate concepts, coordinate concepts and subordinate concepts as shown in fig. 5.

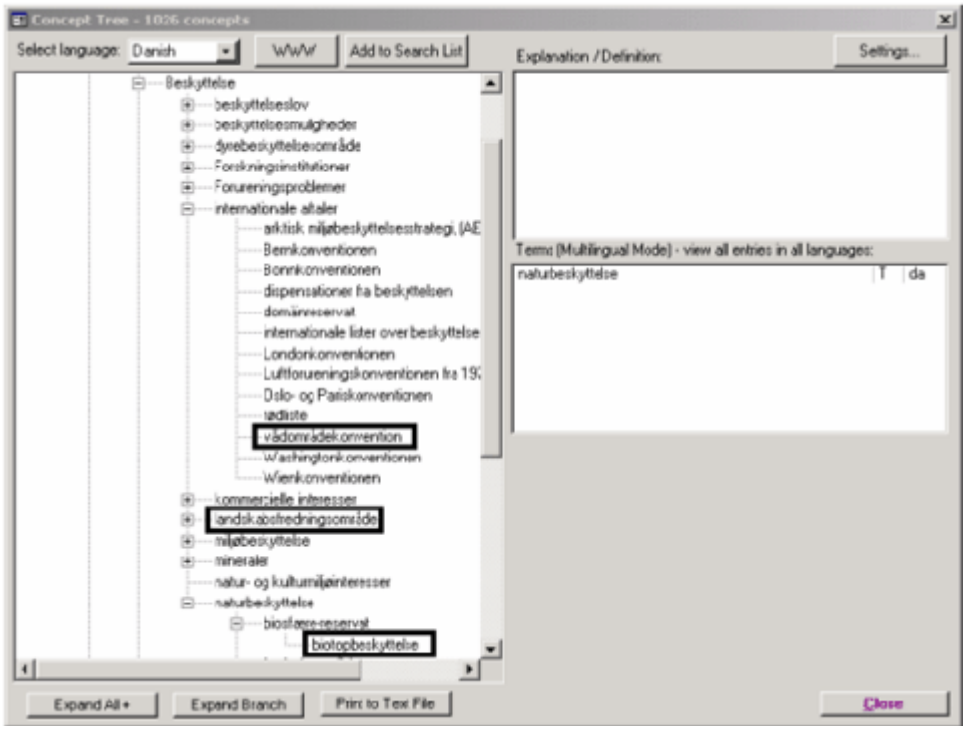


Fig. 5

Fig. 6 illustrates the search interface that contains the search concepts in several languages. First the search concepts to be searched for are selected, and then they are added to a search list, which is submitted to a search engine e.g., Google.

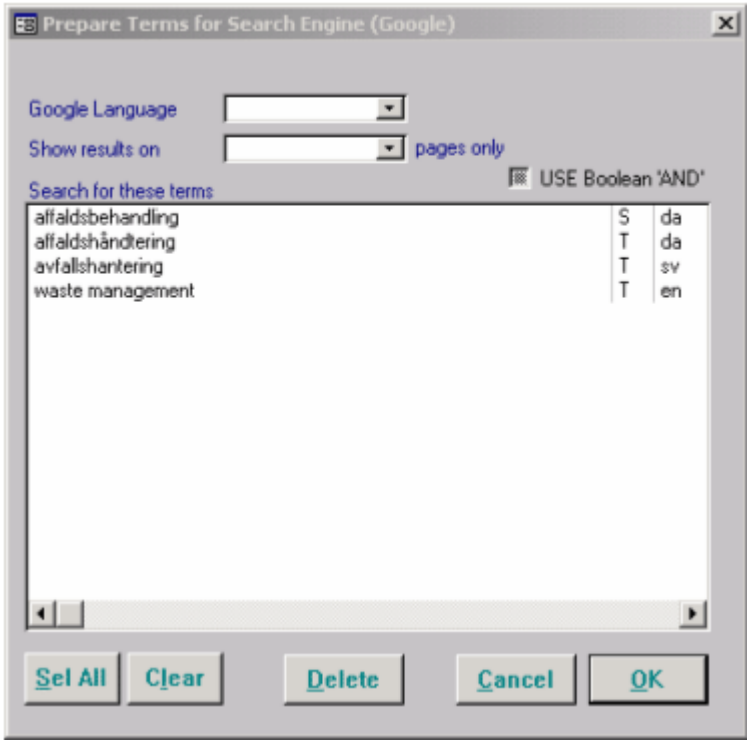


Fig. 6

At the same time it is also important to be able to validate the search concepts. This is done by consulting content related information in the terminology management system, primarily definitions and explanations, to ensure that the selected search concept holds the sense one is interested in and is searching for. Recorded definitions and explanations of the search concept are shown in the right part of

the window, as illustrated in fig. 7. After validation search concepts or selected parts are sent to a search engine.

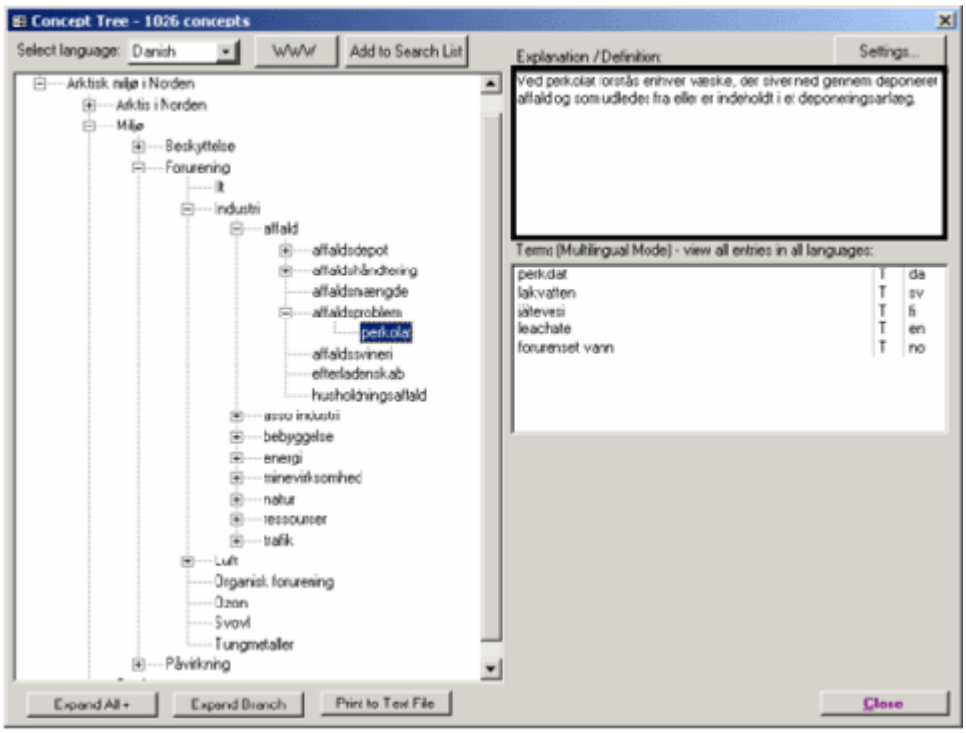


Fig. 7

5 THE THT SYSTEM IN RELATION TO I&D AND TERMINOLOGY

To conclude we should like to compare documentation thesauri, which comprise I&D search terms, with termbases. The comparison takes as its point of departure an article by Øivin Andersen (2005), our colleague at the University of Bergen, Norway. We have supplemented his comparison with the ThT system. In this article we have selected only a few of the aspects presented by Øivin Andersen.

Fig. 8 shows the similarities and differences among documentation thesauri, termbases, and the ThT system.

Documentation Thesauri	Termbases	The ThT System
Information	Communication	Primarily like I&D, but also T
Finite and relevant documents	LSP corpora	Like I&D
Incomplete hierarchies	Ideally complete hierarchies	Like I&D
No distinction as to type of relation	Distinction as to type of relation	Like I&D

Fig. 8

1. According to Øivin Andersen's article, in I&D information is regarded as the crucial aspect whereas terminology focuses on communication. In the ThT system, both are crucial since we use a term management system as a platform for cross language information retrieval.

2. One of the main differences between the two systems is that documentation thesauri are based on a finite and relevant set of documents, whereas LSP corpora for terminological purposes will comprise several central texts of a domain.

In the NorNa network we have so far only been able to work with a single document in each of our corpora, which means that for the time being, our approach is more similar to that of I&D.

3. One of the main differences is that the hierarchies of documentation thesauri are incomplete compared to those of terminological conceptual systems. In the ThT system, the hierarchies will also be incomplete, mainly as a result of our document based approach, which implies that our work is based on a single document not representative of the entire subject field.

4. Unlike terminological conceptual systems, documentation thesauri usually provide no distinction between generic and partitive hierarchies, nor between any other types of relations. As mentioned above, the user of the ThT system may also verify the exact type of relation by consulting the termbase.

What we have just said may suggest that the I&D aspect is predominant in our approach, but it is important to keep in mind that though differences exist among hierarchies, both fields nevertheless use hierarchies as their point of departure. Besides, our approach is more similar to that of the terminologist than that of the documentalist.

In our project I&D and terminology supplement each other. The point of departure of the ThT system was information retrieval, but with a terminological approach.

6 FINAL REMARKS

The project network has used the term management system TERMplus as a platform and developed it further into a cross language information retrieval (CLIR) system for the Nordic languages.

In the spring of 2005, we plan an evaluation phase in which the coverage of the ThT prototype will be tested via a Google search. In addition, we shall increase the amount of data in the termbase. On the one hand, we will first of all add more definitions and explanations in order to be able to validate the search

concepts, thus strengthening the terminological side of our approach. On the other hand we will add more descriptors, i.e. add key words the way it is done in I&D to the terms we have recorded in our document based terminology approach, i.e. the terms functioning as search concepts in the prototype. This means that in this respect, too, we plan to further develop the system both on the I&D and the terminology side.

6 REFERENCES

AHMAD, K. & M. ROGERS (1994). The analysis of text corpora for the creation of advanced terminology databases, Kolding.

ANDERSEN, ØIVIN (2005). NorNa-tool som et hybridsystem. In: Holmboe, Henrik (ed.): Nordisk Sprogteknologi, Årbog for Nordisk Sprogteknologisk Forskningsprogram 2002-2004, Årbog 2004, Museum Tusculanums Forlag, Copenhagen, 81 - 95.

BERNES, CLAES (1996). Arktisk miljö i Norden: orörd, exploaterad, förorenad, Nordiska ministerrådet, Köpenhamn, Naturvårdsverket, Stockholm, ISBN: 92-9120-897-3 (Nordiska ministerrådet), Nord 1996:21, ISBN: 91-620-1168-5, Monitor 15

WEILGAARD CHRISTENSEN, LOTTE & GERT ENGEL (2004). Datafangst ved hjælp af en tværsproglig, nordisk søgemaskine - NorNa. In: Holmboe, Henrik (ed.): Nordisk Sprogteknologi, Årbog for Nordisk Sprogteknologisk Forskningsprogram 2002-2004, Årbog 2003, Museum Tusculanums Forlag, Copenhagen, 243 – 247

Karina Martínez Ferber
euroscript Deutschland GmbH
Berlin
Karina.Martinez@euroscript.de

DECONTEXTUALIZATION IN A MULTILINGUAL ENVIRONMENT

Abstract

Content reusability and thus decontextualisation is a main requirement within CMS scenarios. Source content creation processes have to change as terminological and stylistic consistency become ever more critical. Optimization of translation processes is often seen rather as a "natural" and automatic side effect than as a pre-requisite for the effectiveness of a multilingual content system. This lecture presents the adaptation of translation processes to the requirements of the CMS as a challenge for the (in-house or external) translation service provider, which can nevertheless only be mastered by close cooperation between the translation unit and technical writers or content authors. What are the "typical" scenarios which the translation service provider encounters within a CM workflow, what aspects of translation services are affected, and how does the service provider interact with content authors, translators and proof-readers in order to ensure the success of the multilingual content management system? Against the background of terminology management as a central issue, typical initial situations, questions and procedures will be dealt with, along with the "classic" pitfalls and fallacies.

INTRODUCTION

Against the background of modularized documentation and content management, translators and translation service providers are faced with changing requirements.

Is this statement really true?

To a certain extent. The fact is that many language service providers are only gradually beginning to see the fundamental changes that are affecting their work. And these changes are even less obvious to companies buying technical translations, that is, to industrial enterprises.

Basically, we will be talking about quality here - but about a *different quality of quality*.

The goal of this presentation is to outline the new quality requirements that result from the modular work process and the typical scenarios that they entail for customers as well as for providers of translation services. At the same time, we will take a closer look at current situations at the beginning of the process and pinpoint conflicting aspects and "gaps" between the quality requirement and the realities of production.

In the concluding considerations, the focus will be on professional terminology work as the linchpin to ensure this new type of quality.

A "DIFFERENT QUALITY OF QUALITY"

An example from the automotive industry: the fact that the parts used to build your automobile are standardized and can be used in several different models and model variants does not in itself increase the quality of your vehicle. Perhaps you would even prefer a prototype specifically manufactured for you that better meets your requirements.

However, your overall advantage (so to speak, the "total benefits of ownership") is undoubtedly greater in the case of a standardized product.

The benefits include lower manufacturing costs and, thus, lower prices; improved availability, better repairability (almost) regardless of the workshop you take your car to, the availability of replacement parts and accessories, etc.

If we look at technical documentation again, we are talking about a kind of text quality that meets the basic functional standard: we are assuming that the contents and the choice of words fulfil the subject requirements and – figuratively speaking – that our "car" has the properties in demand by the market.

However, these texts also share some properties with the standardized components of an automobile. These properties arise when we deconstruct complete documents into functional units in the context of modularized documentation.

Just as in the automotive industry, the aim of the business is not to produce single, customized products, but rather to achieve high-quality mass production.

And, of course, to reduce costs.

When a particular vehicle model is phased out, some of its components can continue to be "re-used" in other models without being changed. The same applies to text components and their translation: they can potentially "survive" their original context of creation (if there ever was one) for an infinite amount of time.

We could be translating for eternity!

TYPICAL SCENARIOS AND CLIENT EXPECTATIONS

There are Content Management projects in which the company takes the aspects of terminology and translation into account well before the project starts.

Unfortunately, however, such projects are the exception.

In the majority of the cases, decisions about which system and process to use in the CM environment are made largely independent of any requirements relating to language structure and processes associated to multilingual document management.

This cost-intensive investment is – of course – made with the expectation of a high degree of process automation, resulting in a quick return on investment.

Even if linguistic aspects are considered, requirements are generally only defined for the language of authorship, that is, the source language. Sometimes style guides and glossaries are created or the technical writers are trained in the creation of structured information.

The properties of the information modules which have been created in this way are then expected to be passed on "automatically" to the translations of these texts.

Whatever is reusable in the source language should be reusable in the target language as well. However, the principles of reusability are often not clearly defined or known with regard to source modules. In these cases, individual decisions of the technical writers are the rule. Their decision whether to rewrite or to reuse is often based on the effort necessary for a retrieval compared to that necessary for a new creation. Unsuitable structuring principles, such as information modules, which are too small, lead to time-consuming searches and to redundant text production.

As the new system and workflow are expected to become operational as quickly as possible, existing documentation material is often transferred without any re-editing. Frequently, the traditional method of document creation continues in parallel with single source publishing for quite some time. So to speak, the engine is replaced "at full speed".

Consequently, extensive post-editing is often necessary in the automated process. This procedure may be acceptable for a limited period of time in view of the pressure of circumstances, but it is really a paradox situation: a complex CMS is basically used to continue production in the old way.

TYPICAL SITUATIONS FOR THE SERVICE PROVIDER

For translation service providers, various forms of data transfer are possible for the CMS output.

Sometimes, the published source document (in the target DTP format) is handed in for translation, occasionally complemented with some inserted text blocks that already exist in the target languages. The task of the service provider then is to overwrite the texts that have not yet been translated, usually in a DTP format. If a CAT system is to be used in this process, the texts to be translated must first be extracted from the complete document.

A preferable output method is provided through so-called CAT interfaces. They output the text units to be translated as XML or SGML. A preparation of the data as in the first case does not take place.

However, these interfaces usually do not take into consideration the properties of the software used and the process requirements of the translation. For instance, the sequence of the output XML fragments can deviate from that in the finished document, or the output units exceed the maximum file size that can be processed by the CAT tool editors. In such a case, the service provider usually must split larger units into smaller blocks prior to beginning the translation and then recombine them after the translation. These tasks require a high level of specialized knowledge.

Apart from the text components to be translated, translators also need the surrounding text so that they can understand the overall context. Some CM systems already offer the possibility to output the whole document as a PDF file in which the existing translations are, for example, highlighted in colour. If, additionally, the complete document is available in the source language, then translators can (at least) actively search for and research terminology and ensure that the new sections to be translated are consistent with the surrounding ones.

If translation memory systems were used for the respective customer before the CMS was introduced, a memory might be available as an additional resource. However, due to the tagging differences between the previously used DTP program and XML, the segments' value may be reduced. In other words, the same source sentence is no longer considered a "100% match". Therefore, fuzzy matches cannot be accepted without adapting the tags so that the usable value and any associated cost savings are reduced until a sufficient number of new language pairs has been created again.

In some cases, the customer uses the transfer to a CMS as an opportunity for insourcing more and more of the translation management. Customers might, for instance, want to host translation memories themselves so that they can carry out their own analyses. As a consequence, they will most likely carry out automatic pre-translations before passing the remaining segments that have not yet been translated on to the service provider. These remaining text fragments have, thus, been filtered twice at this point: first on a modular level and then again on a sentence level.

From the standpoint of terminology management, the introduction of modular technical documentation unfortunately does not usually bring about any changes.

Unlike translation memories, which are virtually created as a "by-product", the building-up and maintenance of terminology databases require additional manual processing. They are, therefore, often neither requested nor paid for and left up to the translator's discretion.

Consistency on a terminology level is supposed to be achieved by the interaction of translation memory, supplied reference material and the individual terminology maintenance of the translator.

As mentioned before, the service provider is often unprepared when he learns about the modified production environment, and he is not given the chance to play a constructive role in the planning of new processes and specifications at an early point in time.

Frequently the customer does not see the need to inform the service provider about the basics of the new information structure, the processes, the applicable rules and regulations for the source language (if any) or the objective for reusability.

THE ROLE OF TRANSLATION

Now the translators deserve our more detailed attention, since they are expected to ensure both types of quality we have previously discussed.

To summarize it again briefly, they must

- Create accurate information with correct contents (just as the technical writer for the source language)
- Maintain terminological consistency with regard to any possible context while doing so
- Ensure the reusability of the functional components in multiple contexts

But do they have the necessary information and chances to influence the process?

In the traditional setup, a translation service provider works with a number of internal and external (free-lance) translators. Although it is very important to quality-conscious service providers to maintain a regular working relationship with their free-lancers, they cannot avoid using different translators for the projects of a specific customer, due to personnel availability or fixed deadlines.

Moreover, a service provider frequently supplies more than one language combination.

So the number of translators used for a customer over an extended period of time can be quite large.

As a natural consequence, the amount of time required to transfer information, communicate knowledge and perform checks increases.

Let’s have a quick look at the typical participants within a translation workflow:

The *technical writer* is the author of the source texts. Typically, the technical writer is also the person responsible for the terminology to be used in the source language and thus the "guardian" of in-house knowledge.

Another role that is assigned at the customer's end is that of the *"translations" project manager*; in other words, the customer's interface person who sends jobs to the translation service provider(s).

The translation service provider, in turn, generally provides a project manager who receives jobs and data, clarifies specifications and distributes the tasks to the translators and other resources. This task involves the efficient channelling of data and information streams.

Generally, the translator is the one solely responsible for transporting the text contents into the target language. In most cases, they are also the terminologists for their respective target language because they have to define terms they encounter in the translation process for themselves and decide on an suitable translation.

After the translation is completed, the data records are saved in the system and the document is automatically published or formatted manually. In the context of content management systems, this task is more and more often performed by the customer himself or by another one of his service providers. Ultimately, the goal is to automate the publishing process to a large extent.

After publishing the document – sometimes even immediately following the translation – the next step in the workflow may consist in a translation quality check. Frequently, this is where the customer's foreign subsidiaries are involved in order to confirm the correctness of the translation with regard to content or terminology and to make specific modifications. Their special task within modular documentation processes will be defined later on.

Translators who are familiar with CAT systems have the know-how required for editing tagged formats such as HTML or even special formats created through filtering processes (e.g. Frame Maker or Quicksilver to Trados ttx-format).

However, unlike these tagged formats, XML offers much more comprehensive options for structuring and thus also more options for restrictions. For example, while translators can delete or duplicate internal formatting tags created out of DTP formats without any major consequences, such actions may result in a loss of validity when dealing with XML. Complex structures that define, for example, the sequence of specific elements of a section of text can hinder the natural flow of text in a foreign language.

Consider the following example of a XML structure:

> <source>Improper operation can lead </source>
> <consequence> to dangerous situations. </consequence>

Syntactical or simply stylistic rules in certain languages may require an inverse sequence of the logical elements of the phrase. However, simply switching the sequence of the structuring tags was impossible, because the rule "source precedes consequence" had been defined as a criterion for validity.

Therefore, at minimum, translators need to have advanced training and their work should be checked in the beginning. In cases like this, the project manager of the service provider must assume the function of a multiplier.

Besides transferring skills and information to the translator, there must also be a feedback loop to the technical writer. XML structures that should be avoided (as shown in the above example) due to the characteristics of some languages must be identified and solutions must be worked out together. This is the only way to ensure that the system will work for all languages that are concerned.

Apart from the technical restrictions, however, the translator also must observe content-related rules that go far beyond what is usually taught in translation curricula. Merely using CAT systems limits possibilities for translators.

In translation theory, "text" is generally defined as the "original linguistic sign" (Hartmann 1981: 9). This implies that the "correctness" of a translation is measured by the congruence of the content of the overall text with the original document. The number, structure and content of individual elements of the text, like sentences, can differ from those in the original document without adversely affecting the overall content.

As a matter of fact, my personal experience at university was that translations were considered particularly good whenever their structure deviated from the original so far that the characteristics of the target language were optimally represented while the meaning remained unchanged. And depending on the target language, these characteristics can differ considerably from those of the source language.

The presence of different grammatical and stylistic principles in different languages is a fact. But even without the presence of such particularities in the target language, a translator working on individual sentences may not always see the need to translate them one-to-one to correspond with the source language sentences.

Let's take a look at the following example:
Source text:

> The lid is held by four screws. Remove the screws beginning with screw no. 1.

A correct translation of this short text into German could read as follows:

> „Der Deckel ist angeschraubt. Lösen Sie die vier Schrauben und beginnen Sie dabei mit Schraube Nr. 1."

If you look at the individual sentences by themselves, it becomes clear that, indeed, contents can be moved between sentences without changing the overall meaning of the text. Regardless of whether such a transfer of content is necessary or purely a matter of taste – from the translator's point of view, there's basically nothing wrong with taking this liberty.

However, as the work in a CAT system is segment-based, translators are getting more and more used to decontextualized translating. In the "traditional" CAT process, however, translators generally continue to have the whole text in front of them.

They are used to modifying full or fuzzy matches and re-reading the text completely at the end so that they can compensate for possible deficiencies in consistency and other weaknesses.

Working with text modules takes away these possibilities. In the idealized case of a fully automated process, the translator will – so to speak - never see the final product.

It is amazing how the qualification of the translator for these tasks is taken for granted, while technical writers in many cases are provided, at minimum, with training and with a learning phase.

THE ROLE OF REVISERS

It is not only common to overlook the translators, but also those persons who might be doing proof-reading, e.g. in foreign subsidiaries.

And here is yet another danger in the system that we cannot afford to ignore:

- It takes a considerable amount of effort and time to build a functioning chain from technical writing to translation to publishing.
- It takes system support through terminology databases.
- It takes training and qualification of all people involved.
- It takes a constant, well-organized flow of information.

Texts that have been created within such a strictly governed system conform to many more rules and regulations than defined by the natural languages.

The reviser constitutes just an additional element of this system.

It is obvious that these participants must follow the same rules and specifications the translator used before. They must be provided with the same information, training measures and systems.

Experience shows, however, that in most cases the proof-reading of translations is only a marginal task of these individuals. Most of them do neither have an appropriate basic qualification nor do they have the necessary technical aids and time at their disposal.

Therefore, the most important question is: what goal should be accomplished by integrating a proof-reading step into the process?

Just imagine the potential costs of changing a single technical term in hundreds of Translation Memory segments and a large number of modules within a CMS. Should a decision to change a specific term – entailing substantial costs – be left to the reviser or the translator alone?

A clear decision structure is an absolute necessity in this respect, and the responsibility cannot be passed to the foreign subsidiary or to the service provider alone.

A modular system won‘t work without every participant observing a quite sophisticated set of rules. One of the crucial questions is therefore: Will every participant have the necessary information, time and technical support to deal with these requirements? Whatever the answer and the corresponding workflow decision may be, the point is to ask these questions well before implementing the system.

CONSISTENCY AT THE SOURCE AND THROUGHOUT THE PROCESS

The changed requirements regarding the quality of source and target texts and the new processes will be reflected in an expanded and newly defined range of services offered by translation companies. Few

client companies have in-house staff who are able to provide the necessary expertise and take care of all the required coordination tasks.

Translation service providers must meet the challenge of an expanded concept of quality and convey it to their in-house and free-lance employees. At the same time, the service provider has to define processes that allow him to ensure quality in rendering the new services.

Some of the crucial questions are:

- How can we ensure consistent, equivalent terminology work for a large number of languages at the same time?
- How can we ensure the maintenance and updating of translation memories and who should be responsible for that task in the process?
- What can we do to restrain the translator from "unpermitted creativity"? How do we communicate to translators and revisers the possible quality problems and costs that a simple terminology change entails?
- How should feedback loops to the customer's technical writers be set up?
- How can feedback be ensured and organized for a large number of language directions at all?
- How do we choose and train our translators?
- How do we train the employees who interface between customers, translators and engineers?

The changed requirements should be reflected in an expanded and newly defined range of services offered by translation companies. Consequently, "Service level" or framework agreements could become more important in the cooperation with the customer.

In these agreements, the processes and services should be described in detail. Among other things, such an agreement should include:

- Who delivers what at which stage in the process and how?
- What services are included in the "translation"?
    - Terminology work (to what degree and in which form)?
    - Terminology management (maintenance, provision...)?
    - Context check when files have been pretranslated, or no checking of pretranslated segments whatsoever?
    - Feedback for system optimization?
    - Observance of complex editorial rules; possibly (automated) conformity checks and validations?
    - Checking documents in the layout after an automatic layout creation?
    - Checking, evaluating and incorporating corrections from foreign subsidiaries?
    - Special technical tasks (e.g. programming interfaces, performing country-specific adaptations in the XML structure, etc.)?
    - Supervision and control of all steps and information by project management?
    - Consultative services?
- Which properties of the translation must the service provider ensure?
- Which assurances should be given with regard to the qualification of the employees involved?
- Who is, in fact, liable for a complete document, generated from modules? Is there a warranty that text modules created at different stages of the product life cycle will fit together and make up a usable documentation?
- Decision-making structures: Which employees have the right to make decisions with regard to terminology, style and structure? What do the decision-making processes look like - in other words, what has to happen for a terminology change with potentially far-reaching consequences and costs to be implemented?

CONCLUSION

The meaning of quality in connection with the service "translation" is shifting in the environment of modularized content management: quality characteristics with respect to language content remain an indispensable requirement, but are being expanded considerably by aspects relating to form and form of contents.

As a side-effect, translation quality should become a clearly defined and thus measurable feature. Translators and translation clients all over the world are going to hear this with relief.

Under present conditions, however, knowledge concerning the linguistic part of the system requirements is building up very slowly: system integration and language service are usually completely independent of each other within the overall process. But consistency, which is one of the most important requirements, can only be ensured if, for example, terminology work is performed consistently and professionally as a link in the process chain.

An entirely new spectrum of services is being created around the linguistic conformity of texts which go well beyond the traditional meaning of the term "translation".

Translators and translation service providers should embrace this challenge as a chance to enlarge their ranges of activities and to gain efficiency and performance in their work.

And finally: consulting as well as accompanying the process and providing knowledge transfer between the parties involved in the process are areas of tasks for which the language service provider, the specialist in "building bridges", is predestined.

Face the challenge!

REFERENCES

HARTMANN, P. (1981): The Text as a Linguistic Object. In Stempel, W. D. (ed.): Contributions to the Linguistics of the Text. München.

Bodil Nistrup Madsen
DANTERM & Department of Computational Linguistics
Copenhagen Business School
Copenhagen

IMPLEMENTATION OF RESEARCH RESULTS IN THE DEVELOPMENT OF AN INTERNET TERMINOLOGY AND KNOWLEDGE MANAGEMENT SYSTEM

Abstract

The Danish Centre for Terminology (DANTERM) and the Department of Computational Linguistics at the Copenhagen Business School (CBS) are working closely together in research and development projects.

DANTERM has developed an Internet terminology and knowledge management system, i-Term, which can give users access to company-specific terminology from anywhere in the world. The development of i-Term is based on experience gained from co-operation with Danish companies as well as on the results of the CAOS project (Computer-Aided Ontology Structuring), which is carried out at CBS.

i-Term has many advantages, one of these being the graphical module, i-Model, which allows the user to create a domain specific ontology (concept system), comprising all kinds of relations between concepts, characteristics of concepts and subdivision criteria, as recommended in the ISO standards for terminology work. This module builds on some of the principles that have been developed in the CAOS project.

Other advantages of i-Term are the flexible search facilities as well as the layout facilities which allow users to customize the interface according to their own preferences as to types of information given about a concept. i-Term has XML import/export functions and a look-up facility from Word, PowerPoint and other applications. A demo version of i-Term and i-Model is available at www.i-Term.dk.

In this paper I will focus on the underlying principles of the system, especially those of the graphical module for concept systems.

INTRODUCTION

First I will briefly mention some of the relevant principles used in the research project CAOS - Computer Assisted Ontology Structuring, and give an explanation of central concepts. Then I will briefly introduce the Terminology and Knowledge Management System i-Term, and go into more detail with the concept modelling tool i-Model.

PRINCIPLES USED IN CAOS

The aim of the CAOS project is to develop a computer system designed to enable semi-automatic construction of concept systems, or ontologies. The system is intended to be interactive and presupposes an end-user with a terminological background (terminologist or professional translator), cf. Madsen 2002b. The backbone of the concept modelling in CAOS is constituted by characteristics modelled by formal feature specifications, i.e. attribute-value pairs. The use of feature specifications is subject to a number of principles and constraints. Madsen 2004 describes ten important principles.

The concept modelling tool i-Model does not fulfil all the requirements laid down in CAOS. However, it is compatible with the principles of CAOS. Here I will focus on the principles that are most relevant for i-Model. As an example ontology I present part of an ontology for printers in Figure 1, which will be our starting point for describing the principles of CAOS.

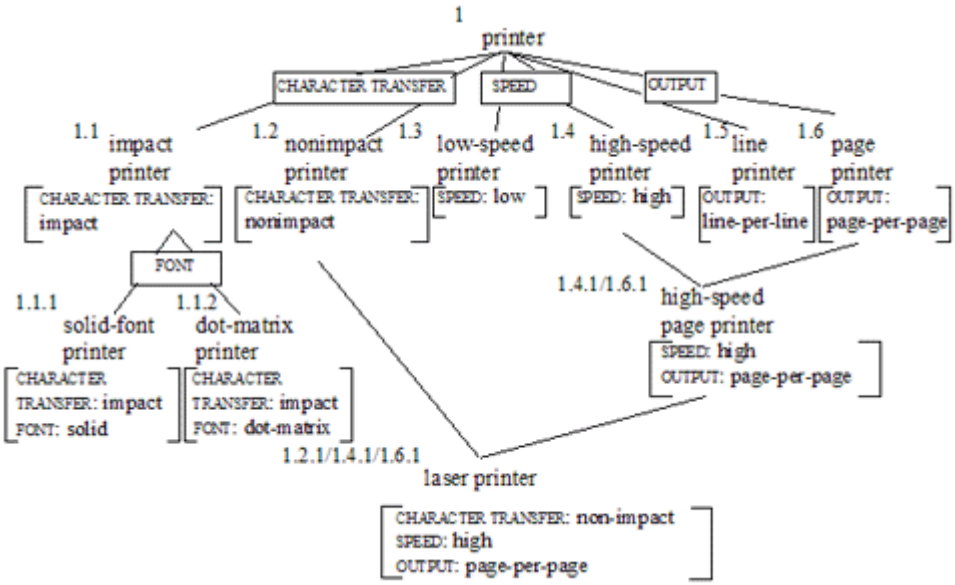


Figure 1: Part of an ontology for printers

Feature specifications

When making an analysis of a concept the terminologist will always look at the nearest superordinate concept and distinguish the concept from this and from any coordinate concepts by at least one delimiting characteristic. Feature specifications are the formal modelling of the terminologist's characteristics. Both 'impact printer' and 'nonimpact printer' are types of 'printer', and they are distinguished by means of the feature specification containing the attribute CHARACTER TRANSFER with the different values: 'impact' and 'nonimpact'. CHARACTER TRANSFER means 'way of transferring characters from printer to paper', and the value IMPACT means 'striking the paper'.

Function from attributes to values

A concept is characterized by means of a set of feature specifications, called a feature structure. A feature structure is formally defined as a (partial) function from the set of attributes to the set of values (cf., for instance, Carpenter 1992). This means that an attribute may only be associated with one value in a feature structure. Thus one concept must not be characterized by two feature specifications having the same attribute, but two different values, for instance: [CHARACTER TRANSFER: impact] and [CHARACTER TRANSFER: nonimpact]. In CAOS procedures to check this have been developed.

Inheritance of feature specifications

The principle of inheritance means that we have to distinguish two kinds of feature specifications: primary and inherited. A primary feature specification is assigned directly to a given concept, whereas an inherited feature specification is inherited from the concept's superordinate concepts. In Figure 1 the feature specification on 'solid-font printer' [FONT: solid] is a primary feature specification, whereas [CHARACTER TRANSFER: impact] is inherited from the concept 'impact printer' and thus it is not a primary feature specification.

Dimensions

A dimension is an attribute whose possible values allow a distinction between some of the subconcepts of the concept in question. In graphic illustrations I represent dimensions by means of boxes covering the relevant branches between a superordinate concept and its subconcepts as in Figure 1 (CHARACTER TRANSFER, SPEED, OUTPUT and FONT). These will often prove helpful to the user because they help significantly to give a clearer overview of concepts within a certain domain.

Subdividing dimensions

As can be seen from **Figure 1**, subdividing dimensions group sister concepts according to the attributes contained in their delimiting feature specifications. Just as feature specifications are the formal modelling of the terminologist's characteristics, subdividing dimensions serve to model the terminologist's subdivision criteria. One or more of the dimensions of a concept must be chosen as subdividing dimensions.

Consider the fragment of the ontology for printers in **Figure 2**, in which three dimensions have been registered for the concept 'printer'. Formally all three dimensions qualify for the role of subdividing dimension, but the terminologist will choose CHARACTER TRANSFER because she will realize that this is the essential one because the characteristics associated with this dimension determine the other characteristics: the characteristics of being noisy or quiet and the capacity for single or multiple copies are consequences of the kind of character transfer employed. The choice of this subdividing dimension entails that impact printer will be associated with a definition like the following: 'a printer that transfers characters from printer to paper by striking the paper'.

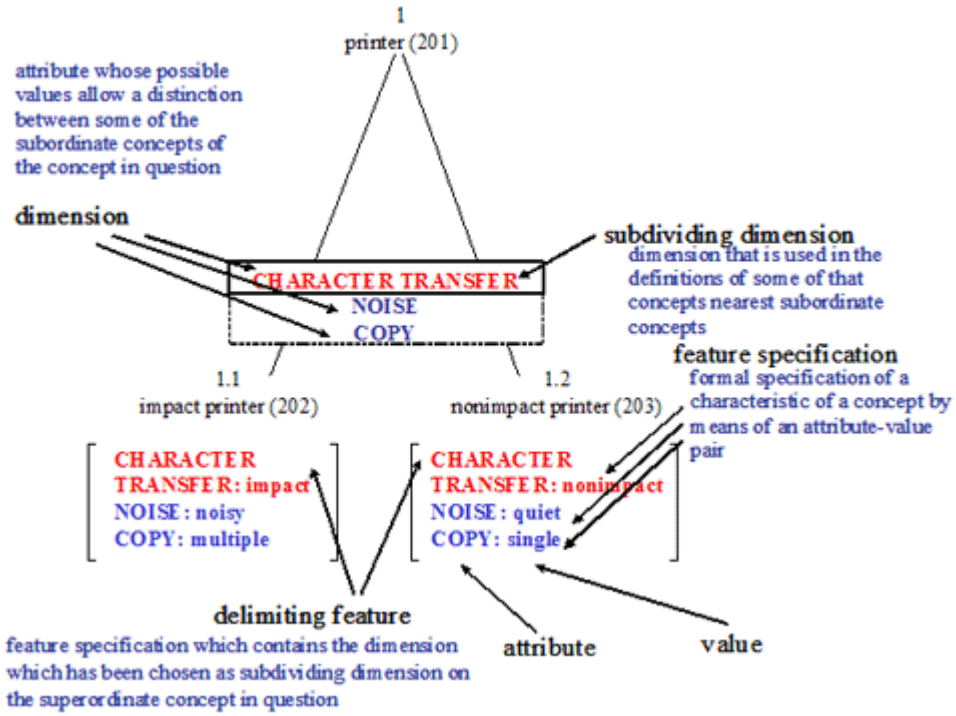


Figure 2: Three dimensions

Polyhierarchical structure

A concept in CAOS may have more than one nearest superordinate concept. This is the case when the concept has feature specifications in common with two superordinate concepts. In Figure 1 the concept 'high-speed page printer' is a subordinate concept to both 'high speed printer' and 'page printer', and thus inherits the feature specifications [SPEED: high] and [OUTPUT: page-per-page] from the two superordinate concepts respectively. In this case the concept 'high-speed page printer' will be associated with a definition, comprising the combination of the two feature specifications. Before going into more detail with i-Model I will very briefly introduce i-Term.

I-TERM – AN INTERNET TERMINOLOGY AND KNOWLEDGE MANAGEMENT SYSTEM

The terminology and knowledge management system i-Term has been developed for registering and maintaining company- and institution-specific terminology. Some of i-Terms main features are:

- Accessibility via the Internet
- Wide selection of search combinations (same qualities as a free text database application)
- Customization of the user interface according to the preferences of the individual user (selection of fields to be displayed in search results and in the data entry interface)
- Different user levels according to access rights
- User defined pick lists for word classes, subjects, status, and languages
- Storage and presentation of multimedia such as image files, video and audio files
- Storage of information on references relevant to the terminology entries
- Hyperlink management
- Import and export using XML
- Special look-up facility: i-Find which allows the user to look up a term in the database from any other application on his or her pc
- Concept modelling tool: i-Model

I-MODEL – THE CONCEPT MODELLING TOOL OF I-TERM

The concept modelling tool i-Model allows the user to interactively produce a graphical representation of a concept system. It is possible to enter all kinds of concept relations, using special symbols for type relations, part-whole relations, temporal relations and other relations, which may be named specifically by the user.

The user may also enter concept positions, feature specifications and subdivision criteria (subdividing dimensions).

In **Figure 3** concepts from the terminology of the typewriter keyboard are presented (in no special order) as a result of a search for all concepts within the subject field Office machines. Terms are grouped by concept to show the connection between synonym terms. Thus the terms 'correcting key' and 'correction key' are synonyms (same background colour).

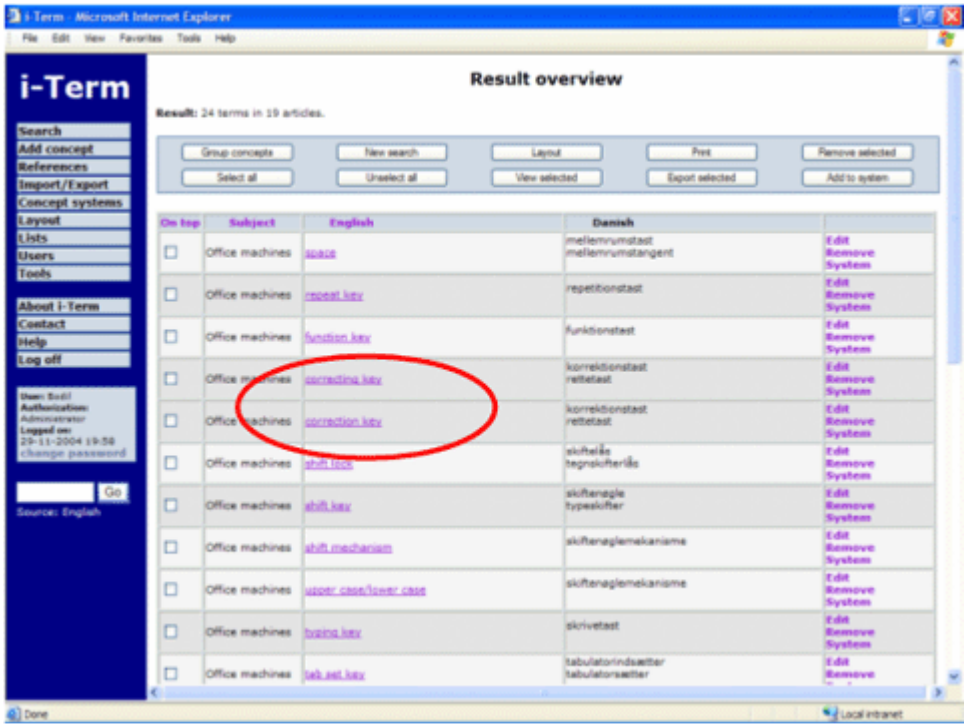


Figure 3: Typewriter keyboard concepts

In **Figure 4** a draft concept system for the concepts from the Result overview of the previous slide is presented. This system only comprises type relations (type of) and part-whole relations (part of). The 'correction key' is a type of 'function key'. The 'shift key' is a part of 'shift mechanism'.

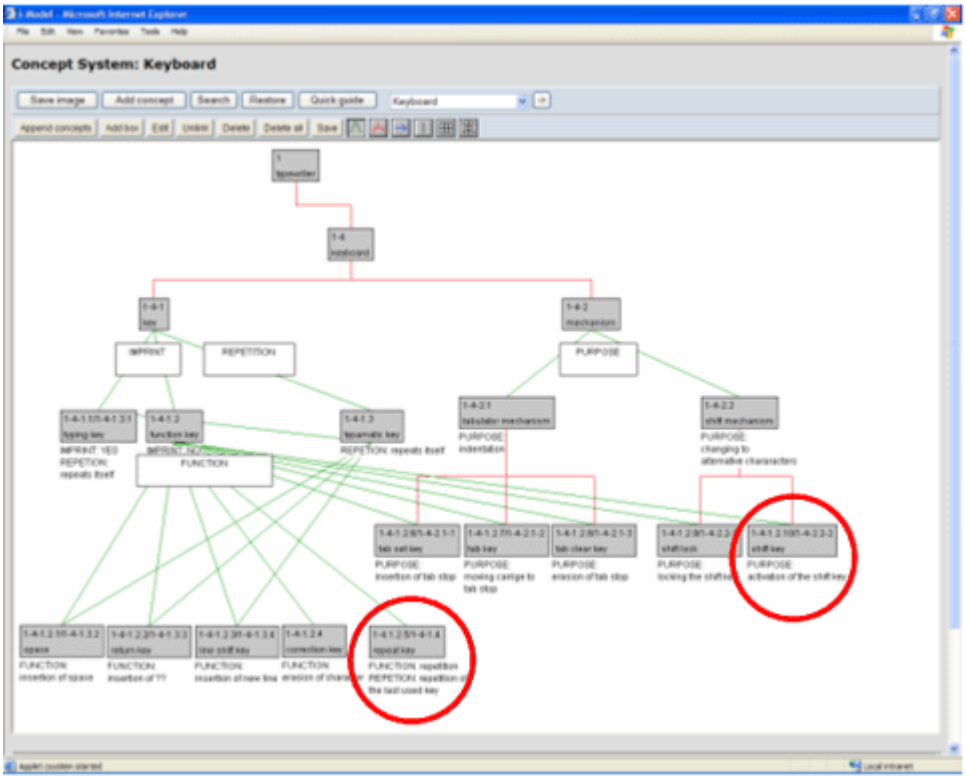
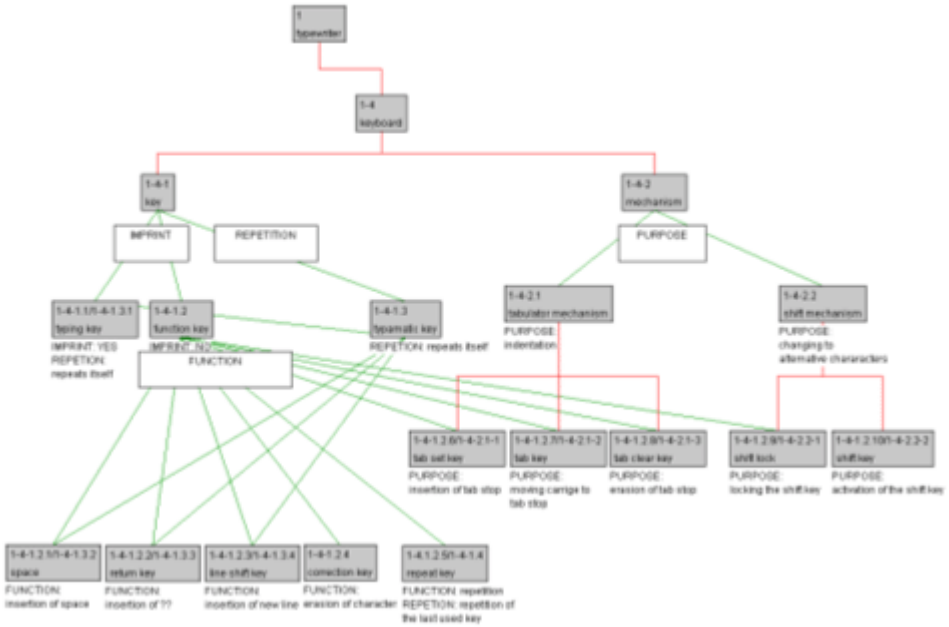


Figure 4: Concept system for typewriter keyboard concepts

In **Appendix A** the same concept system is shown as a GIF file produced in i-Model.

Appendix A



This concept system comprises:

- concept positions (e.g. 1-4-1)
- feature specifications (e.g. [IMPRINT: yes])
- subdivision criteria (e.g. IMPRINT in a box).

All information from the concept system is automatically available in the articles containing information about the concepts, cf. **Figure 5**.

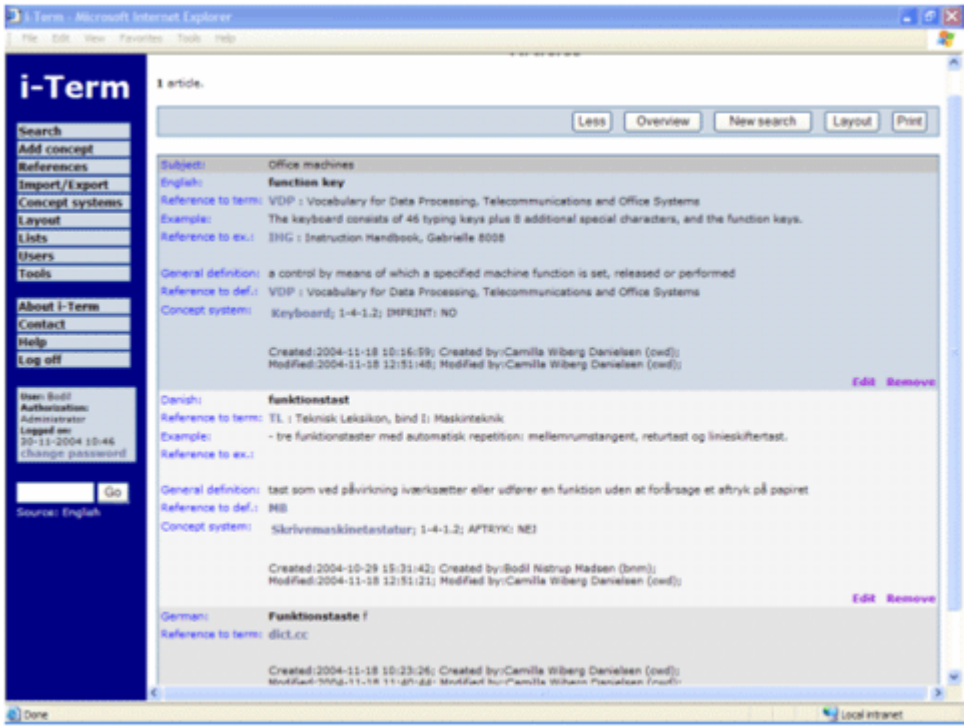


Figure 5: Article in i-Term containing information about 'function key', including information about concept system

It is possible to click on one concept in the concept system to display information about it, cf. **Figure 6**, in which information on 'shift mechanism' is presented.

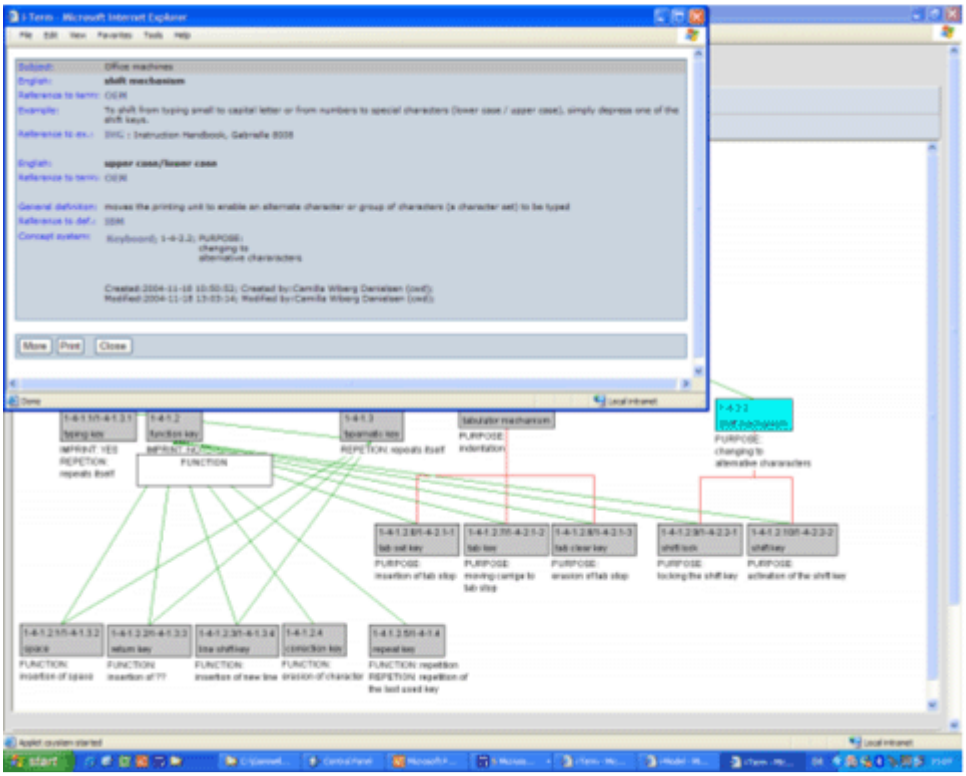


Figure 6: Presenting information about one concept in the concept system

In **Figure 7** a draft concept system for windmills is presented. The 'vertical-axis windmill' and 'horizontal-axis windmill' are types of 'windmill', and 'American multi-blade windmill' is a type of 'horizontal-axis windmill'. The 'rotor' is a part of 'wind mill'.

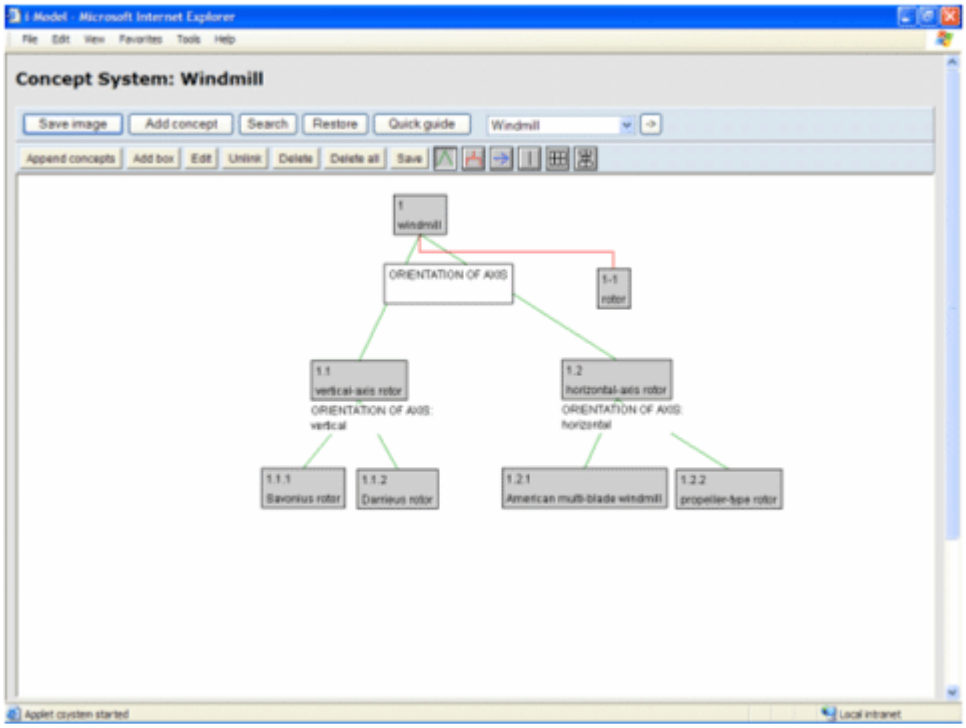


Figure 7: Unfinished draft concept system for windmills

The user may choose to add concepts to an existing or a new concept system from the data editing or the data entry interface. The concepts will appear as 'boxes' in the concept system and the user may reorganize the concepts and introduce concept relations etc., cf. **Figures 8 - 11**.

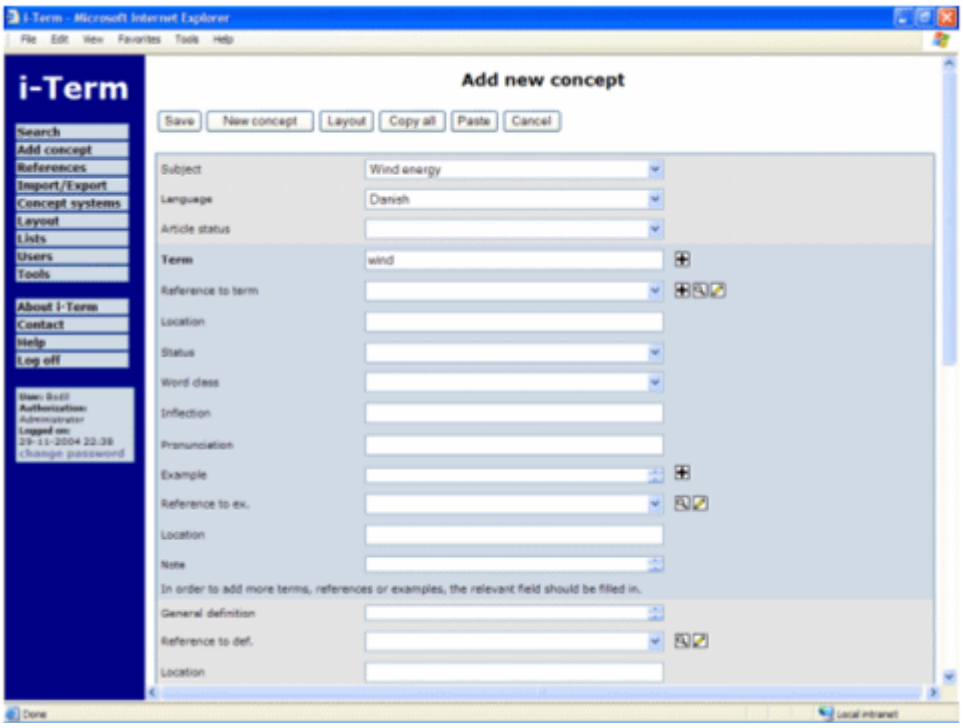


Figure 8: Addition of a new concept 'wind' to i-Term

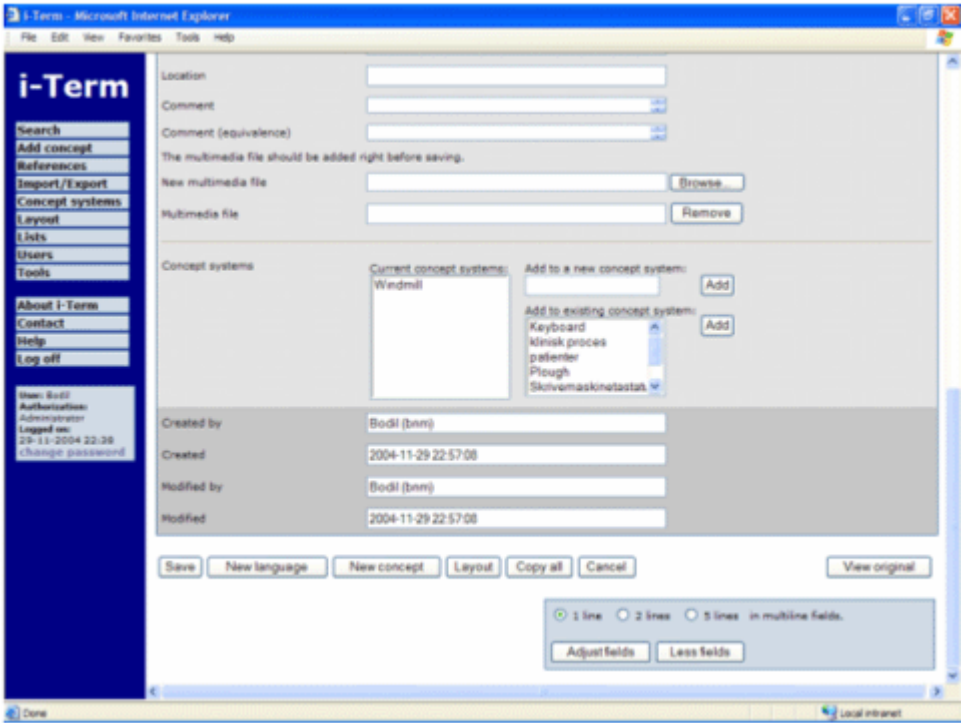


Figure 9: Addition of the concept 'wind' to the concept system Windmill

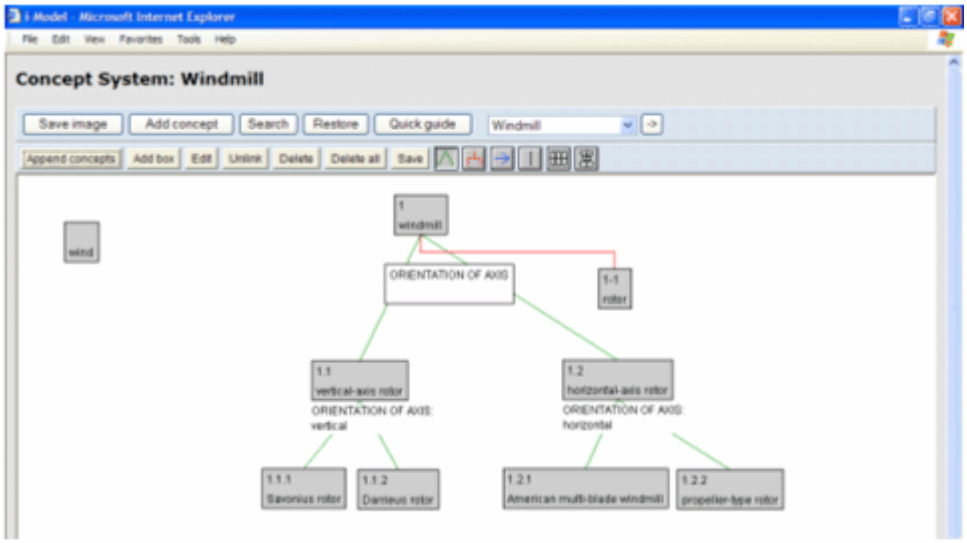


Figure 10: The concept 'wind' appended to the graphical concept system Windmill

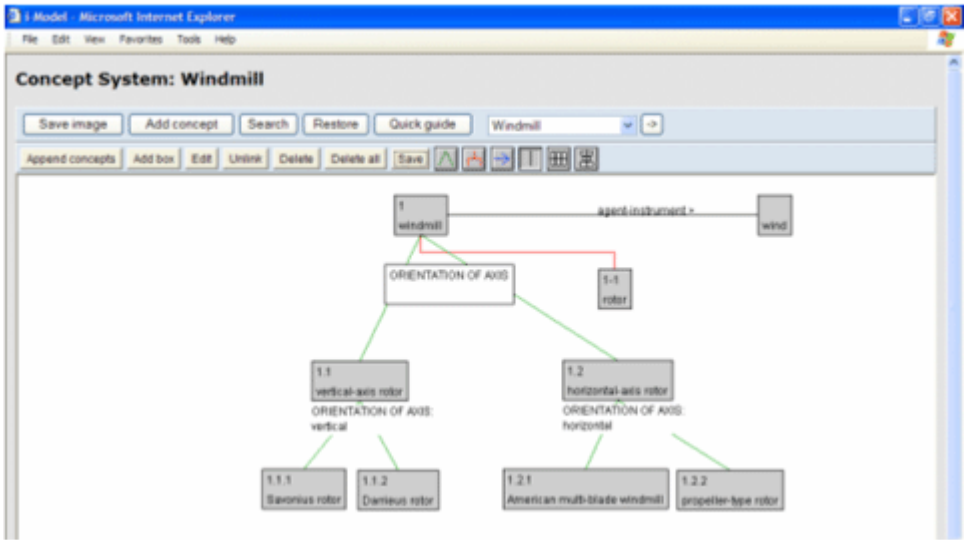


Figure 11: Associative relation (agent-instrument) between 'windmill' and 'wind'

When the terminologist builds the concept system, she may choose between four types of concept relations by pressing the relevant button before drawing the link:

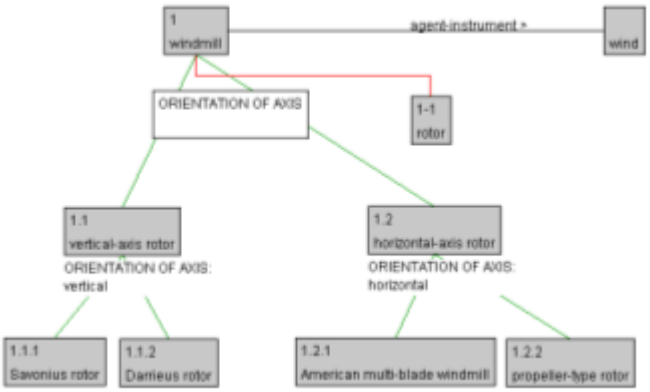
- type relation – symbol: 
- part-whole relation – symbol: 
- temporal relation – symbol: 
- associative relation – symbol: 

In **Figure 11** the relation between 'windmill' and 'wind' is an associative relation (agent-instrument), which is indicated in the concept system by adding the name of the relation.

See more about associative concept relations below.

In **Appendix B** the same concept system is shown as a GIF file produced in i-Model.

Appendix B



ASSOCIATIVE CONCEPT RELATIONS

Figure 12 contains an overview of some associative relations defined in the Danish interdisciplinary research project OntoQuery – Ontology-based Querying, which is an interinstitutional project with several partners, cf. Madsen 2002a and www.ontoquery.dk.

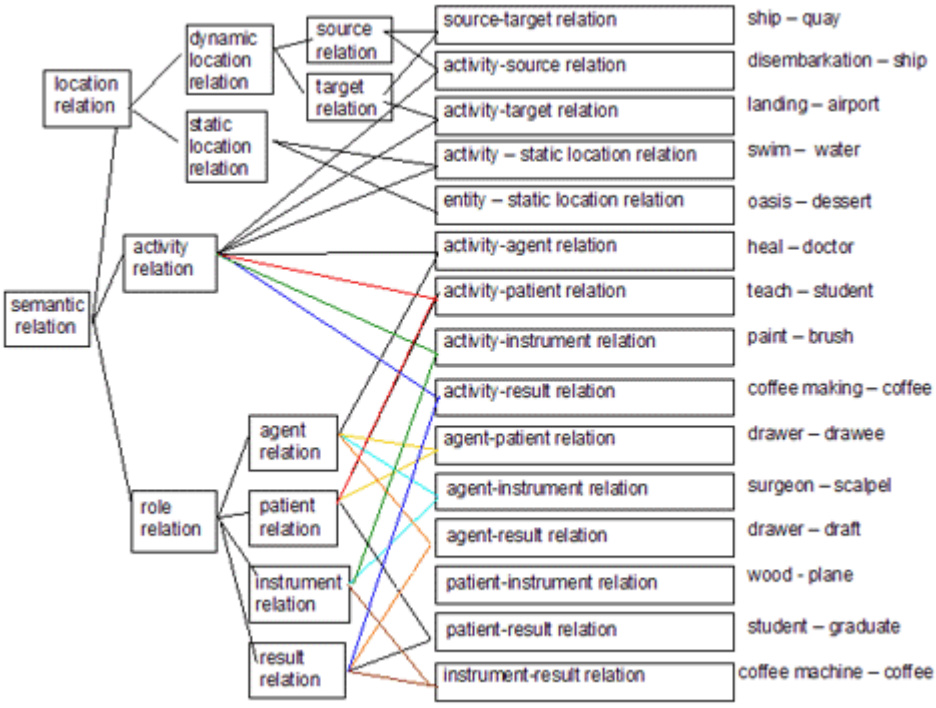


Figure 12: Overview of some associative relations

BUILDING A NEW CONCEPT SYSTEM

As we have already seen, it is possible to add a concept to an existing concept system. It is also possible to select some or all concepts in the result overview, and add them to a (new) concept system. In the

following example all Danish concepts belonging to the subject Soil preparation have been selected and added to a Danish concept system for Ploughs ('Plove'), cf. **Figure 13**. The concepts may now be reorganized and the terminologist may add characteristic features and notations, cf. **Figures 14 - 16**. In Appendix C the English concept system for ploughs in **Figure 17** is shown as a GIF file.

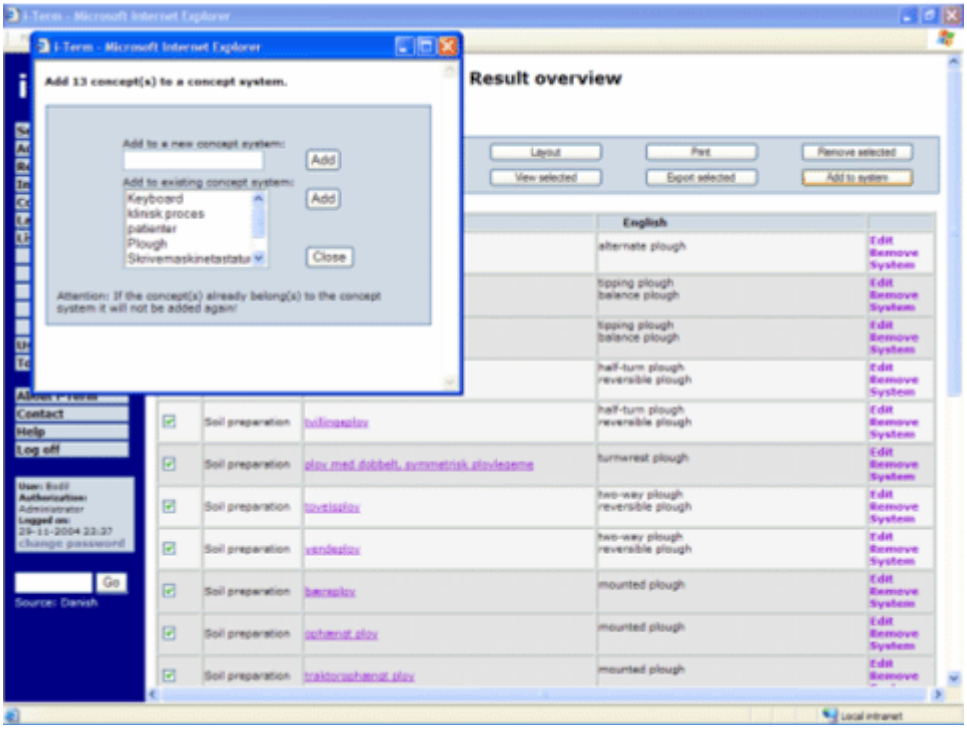


Figure 13: Danish concepts belonging to the subject Soil preparation

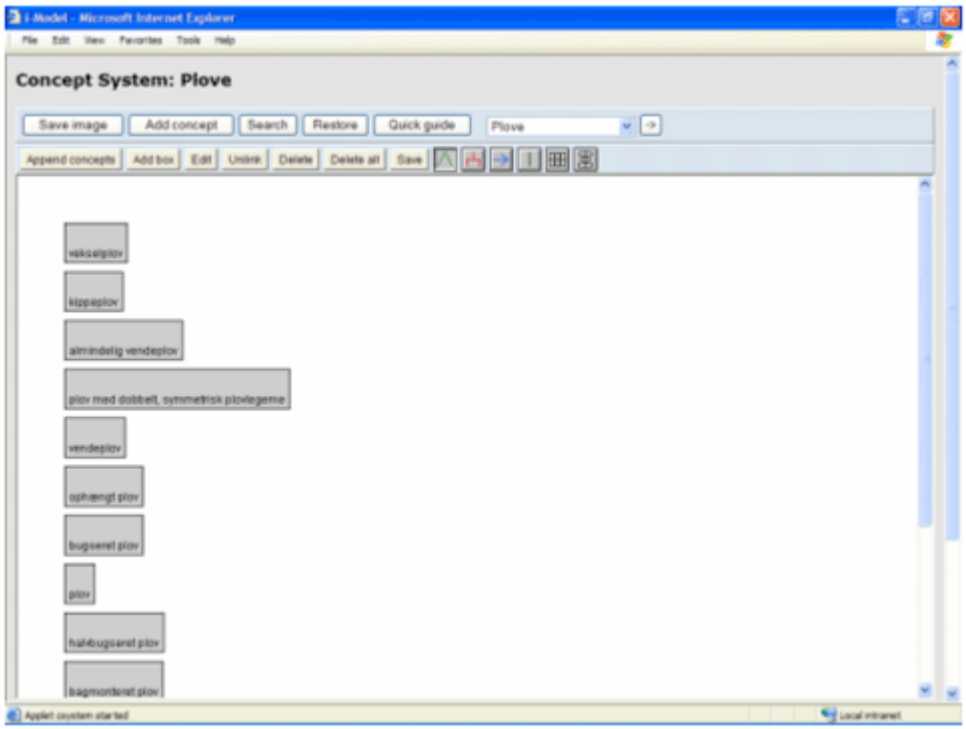


Figure 14: Danish concepts added to the concept system 'Plove' (Ploughs)

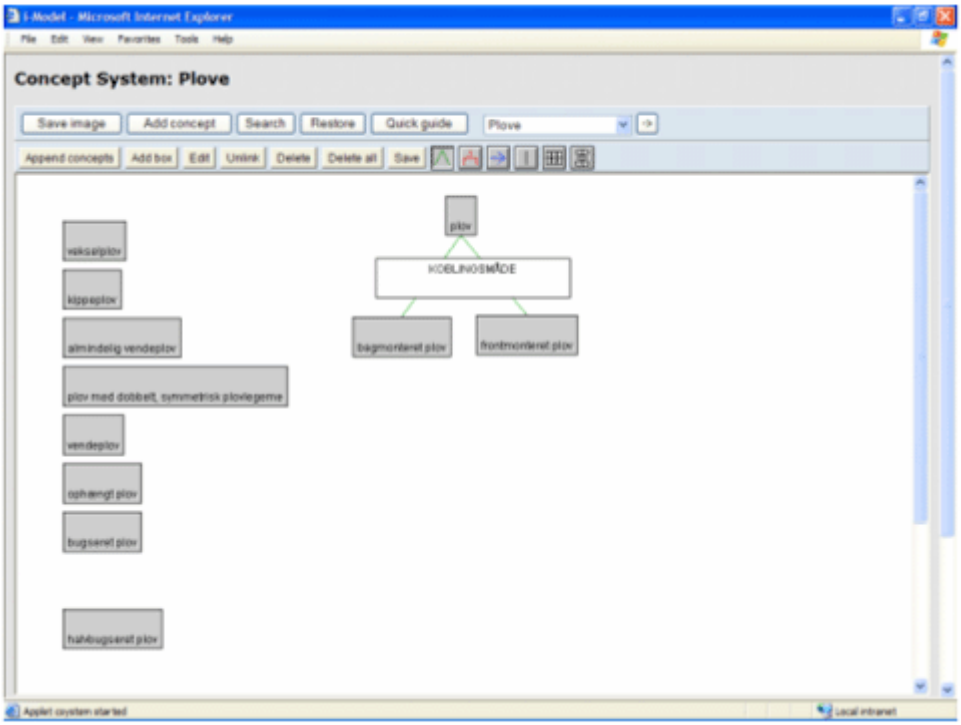


Figure 15: The 'top' of the concept system 'Plove' (Ploughs)

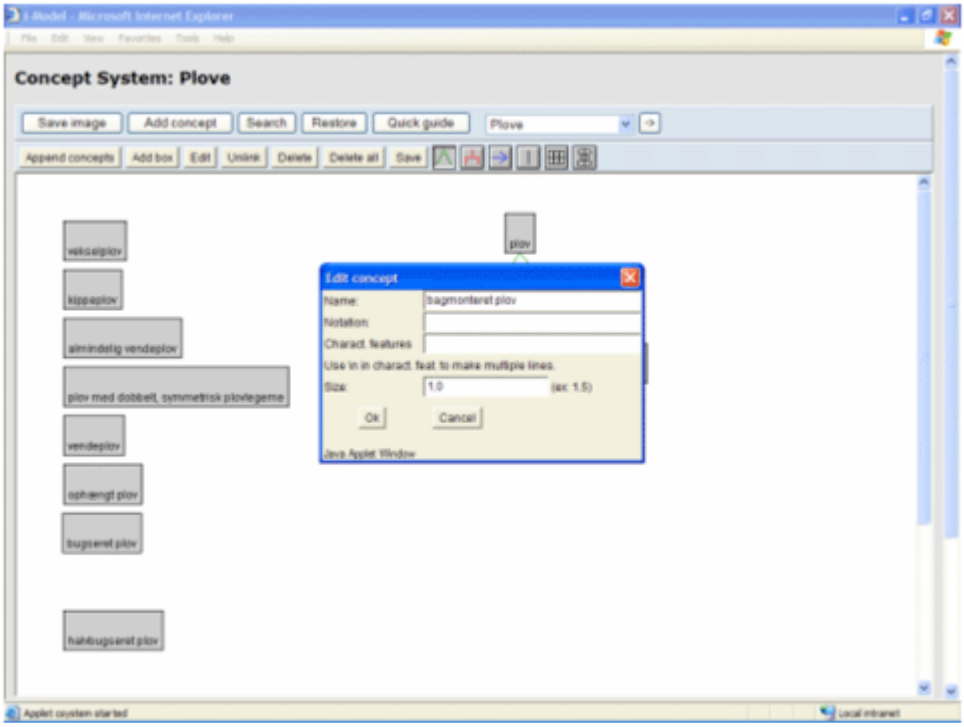


Figure 16: Addition of notation and characteristic features

Appendix C

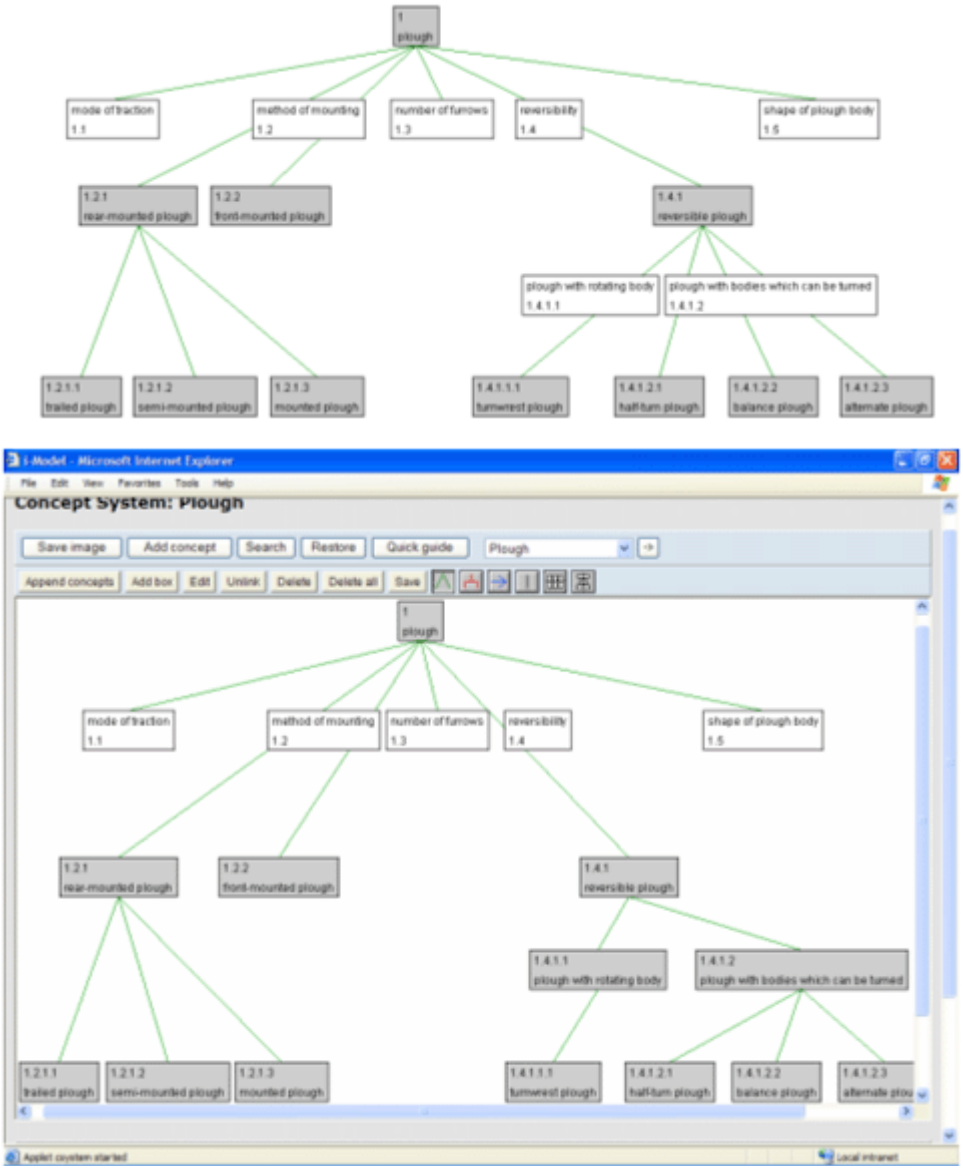


Figure 17: Draft version of the English concept system for ploughs

The concept modelling tool i-Model offers flexible drawing facilities. For example it is possible to select a number of concepts and to move the whole group simultaneously. In **Figure 18** the concept 'horizontal axis rotor' and its subconcepts from Figure 11 have been moved to the left of the concept 'vertical axis rotor' and its associated subconcepts. The systematic notations should be adjusted accordingly.

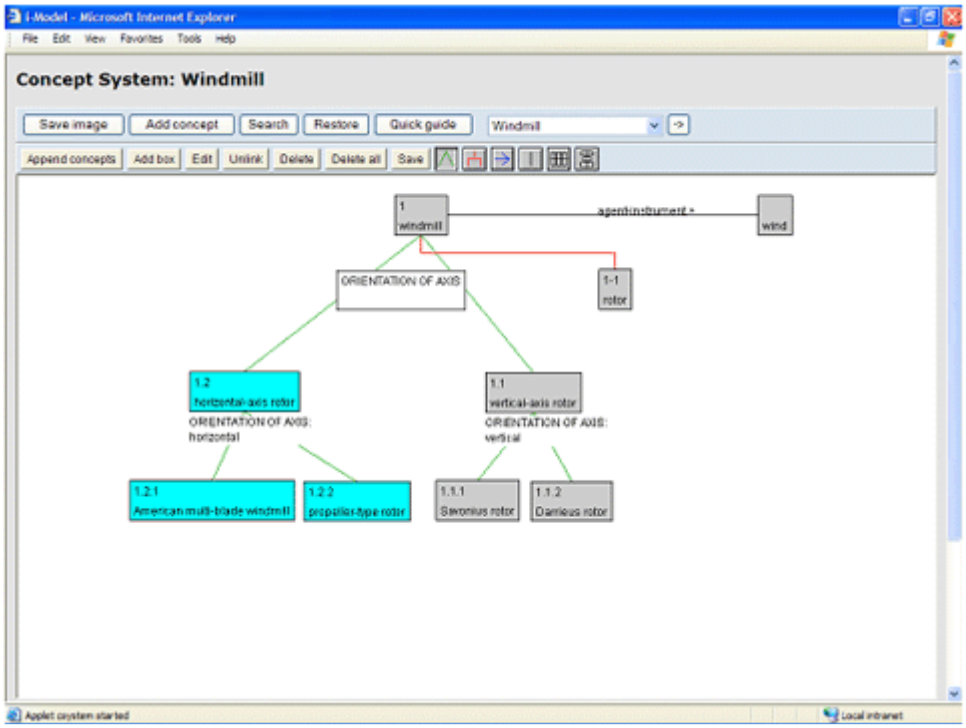


Figure 18: Moving a group of concepts in the system for windmills

PRINCIPLES OF CAOS IMPLEMENTED IN I-MODEL

Compared to the principles of CAOS, i-Model has the following facilities:

It is possible to add feature specifications to concepts in the ontologies in i-Model. There is, however, no checking of the principle laid down in CAOS, that an attribute may only be associated with one value in a feature structure (cf. Function from attributes to values). This principle means that one concept must not be characterized by two feature specifications having the same attribute, but two different values.

It is possible to introduce the principle of inheritance in i-Model, but there is no automatic procedure for doing it.

The graphical representation of dimensions has been implemented in i-Model, and the terminologist may enter two dimensions in one box, and then later decide upon one of them.

The principle of polyhierarchy is implemented in i-Model.

Compared to the present version of the CAOS prototype, i-Model further offers advanced graphical drawing facilities, which enable the user to move concepts or groups of concepts.

CONCLUSION

As a conclusion I will briefly list the features of i-Model. It

- helps the user in setting up complicated concept systems in an easy way
- offers a formalization of traditional terminology principles
- contributes to consistency in the concept system
- contributes to the elaboration of consistent definitions on the basis of feature specifications

REFERENCES

CARPENTER, BOB (1992). The Logic of Typed Feature Structures. Cambridge, Massachusetts: Cambridge, University Press.

MADSEN, BODIL NISTRUP, BOLETTE SANDFORD PEDERSEN & HANNE ERDMAN THOMSEN (2002a). Semantic Relations in Content-based Querying Systems: a Research Presentation from the OntoQuery Project. In: K. Simov and A. Kiryakov (Eds.): Ontologies and Lexical Knowledge Bases. Proceedings of the 1st International Workshop, OntoLex 2000. OntoText Lab., Sofia, 72-82

MADSEN, BODIL NISTRUP, HANNE ERDMAN THOMSEN & CARL VIKNER (2002b). Computer Assisted Ontology Structuring. In: Melby, Alan (ed.): Proceedings of TKE '02 - Terminology and Knowledge Engineering, INRIA, Nancy, 77-82.

MADSEN, BODIL NISTRUP, HANNE ERDMAN THOMSEN & CARL VIKNER (2004): Principles of a system for terminological concept modelling. In: Lino, Maria Teresa; Maria Fransisca Xavier, Fátima Ferreira, Rute Costa, Raquel Silva (eds.): Proceedings of the Fourth International Conference on Language Resources and Evaluation. ELRA, 15-19.

Oliver Streiter
NUK, National University of Kaohsiung
Leonhard Voltmer, Isabella Ties and Verena Lyding
Eurac research, European Academy Bolzano

STRUCTURING TERMINOLOGICAL DATA: THE BISTRO PROPOSAL

Abstract

For more than 15 years the computer is employed in terminology work. In the beginning terminology systems have been simple programs for inserting and modifying data stored in a static format. During the last couple of years the potentialities of computer assistance became obvious. These days systems that offer the possibility to flexibly access, link and display information are much asked for. In this paper we will present the BISTRO approach for a dynamic terminology system. We will point out requirements that current implementations have to tackle, provide a comparison with alternative approaches and discuss the strengths of our implementation.

1 INTRODUCTION

When managing complex terminological data, the limitations of term-banks become apparent soon. The list of requirements is long: Terminological data should be unambiguous, economical, non redundant, modular and formalized. They have to be transparent and reflect relevant differences in meaning or usage. They should serve different users with different information needs, different cultural and linguistic backgrounds. The data should be suited for on-line consultation with different protocols, the creation of CD-ROMs and paper dictionaries and should support the process of document creation, translation and management as good as possible. Term-banks should be embedded into a multi-user environment and support simultaneous consultation and updating. A term-bank failing on one or more of these criteria will sooner or later lead the terminologist into severe difficulties.

While older term-banks clearly cannot cope with these requirements, a closer look behind the scene of modern systems reveals that one of the main hindrances which impede progress in the field has not been addressed. Terminological data are still fragmented in so-called entries and emulate a file card system which dates from the early days of human writing. Although this model has obvious strong points as its modularity and its clear arrangement, further reflection will highlight its inadequacy to fulfill the above mentioned requirements. As shown in Streiter and Voltmer (2003a) and Streiter et al. (to appear), this model is not suited for the storage and insertion of data. It complicates or even prohibits their management and it makes a user-adaptive term presentation or presentation in different media, protocols or devices intolerably cumbersome.

Modern term-banks embrace XML, as XML promised to separate the storage and presentation of data and to provide a solution to the aforementioned requirements. At the same time, however, they stick to tree-like data structures, a design which reflects a fundamental misconception of human cognition and communication. Conceptual structures are circular networks. They are communicated however in tree-like structures, leaving it to the receiver to recompute the reentrances with the help of principles and conventions. This wisdom is embodied in almost all syntactic theories. They assume a non-tree-like underlying (conceptual) structure and map it on a tree-like surface structure (cf. Meaning <=> Text Theory, LFG, Relational Grammar, GB and its successors).

What may seem a theoretical subtleness may become dramatic when complex and valuable data are at stake. Using tree-like structures to store knowledge is like using a natural language expression for the storage and to hand over the task of understanding to the computer. But how can computers correctly recalculate the relations from a complex sentence like: "I gave Mary the flowers which I had bought 3 days before the birthday of her son, so that she could give them to him for me in time."?

Net structures have a promising future. Some the leading projects in the periphery of terminology, let's name WordNet, EuroNet, FrameNet, MindNet and RDF, have abandoned tree-like structures already a decade ago, c.f. Powers (2003). Tree-structured terminology deemed them a dead-end road.

Under these circumstances, EURAC investigated and implemented an alternative data design to overcome the limitations of entry-based terminology. In our BISTRO, the Juridical Terminological Information System Bolzano, project we handle legal terminology for Austria, Germany, Switzerland and Italy, with special focus on the legal and linguistic situation in South Tyrol, an autonomous province with 3 languages and legally binding translation relations between them.

The remainder of this paper is structured as follows. In Section 2 we will present our abstract data model. In Section 3 we will motivate the actual implementation and discuss the respective roles we assigned to the relational database and XML. Section 4 is dedicated to so-called term tools (term extraction, term recognition, text acquisition through agents, text classification, KWIC) and the status they have in our model. Section 5, finally will discuss matters of term presentation and how we think to orient the user in a richer and more informative environment. In the final conclusions we will come back to our initial claim and show that the break-up of an entry-based term model indeed can solve many of the problems current terminological work is struggling with.

2 DATA MODEL FOR TERMINOLOGY WITHOUT ENTRIES

The model is designed to flexibly make statements about entities which are relevant for a specific terminology. Statements are formalized as predicates. Unary predicates identify entities. E.g. <tt>term(213)</tt> states that there is a term called 213. Binary predicates state something about an entity. <tt>denomination(553,'legge provinciale')</tt> states that there is a denomination 553 which is written as 'legge provinciale'. In the same way predicates of higher order define entities and establish relations between different entity types.

Following these principles we can model a small fragment of a terminology by defining three predicates: <tt>grammar(665,'Nf')</tt>, <tt>denomination(553,'legge provinciale',665)</tt>, <tt>term(213,553)</tt>. The binary predicate <tt>grammar(665,'Nf')</tt> identifies a grammatical entity called 665 that describes a female noun. The trinary predicate <tt>denomination(553,'legge provinciale',665)</tt> thus states that the denomination 553, which is spelled as 'legge provinciale', is a noun of the grammatical gender female. In virtue of this <tt>term(213,553)</tt> states thatthe term 213, spelled as 'legge provinciale', is a female noun.

Using upper case characters (e.g. TERM, DENOMINATION, GRAMMAR) as variables for entity names and lower case characters for defining attribute values, we may formalize this simple model as shown below. This model can handle queries like 'Find all terms belonging to a specific grammatical class' or vice versa Find all grammatical classes belonging to a specific term'.

Model I

term(TERM,DENOMINATION).
denomination(DENOMINATION,denomination,GRAMMAR).
grammar(GRAMMAR,grammar).

While such queries are within the scope of tree-like entry structures, queries like 'Find all terms which are assigned a legal domain which is different from that assigned to the text which contains the definition of the term.' may already be beyond the expressive power of tree-like entry structures and require the intervention of a student programmer. In our systems, such queries are very much within the system and do not require other tools than a simple term search. Of course, many more queries may be handled by this model.

Model II

term(TERM,DENOMINATION,DEFINITION,LEGAL_DOMAIN).
legal_domain(LEGAL_DOMAIN,legal_domain).
denomination(DENOMINATION,denomination,GRAMMAR).

grammar(GRAMMAR,grammar).
definition(DEFINITION,CORPUS_SEGMENT).
corpus_segment(CORPUS_SEGMENT,LEGAL_DOCUMENT).
legal_document(LEGAL_DOCUMENT,LEGAL_DOMAIN).

An almost complete sketch of the model designed for BISTRO is reproduced as Model III. A graphical representation of it in terms of arcs and nodes is shown in Figure 1.

Model III

term(TERM,DENOMINATION,DEFINITION,LEGAL_DOMAIN, CONTEXT,LEGAL_SYSTEM).
term_relation_type(TERM_RELATION_TYPE, term_relation_type).
term_relation(TERM_RELATION,TERM1, TERM_RELATION_TYPE,TERM2).
translation(TRANSLATION,TERM1, NORMATION,TERM2).
normation(NORMATION,normation).
legal_domain(LEGAL_DOMAIN,legal_domain).
legal_system(LEGAL_SYSTEM,legal_system).
legal_hierarchy(LEGAL_HIERARCHY,legal_hierarchy).
legal_quality(LEGAL_QUALITY,legal_quality).
author(AUTHOR,author).
publishing_house(PUBLISHING_HOUSE, publishing_house).
publishing_place(PUBLISHING_PLACE, publishing_place).
title(TITLE,title)
denomination(DENOMINATION,denomination, GRAMMAR,LANGUAGE).
grammar(GRAMMAR,grammar).
language(LANGUAGE,language).
definition(DEFINITION,CORPUS_SEGMENT).
context(CONTEXT,CORPUS_SEGMENT).
corpus_segment(CORPUS_SEGMENT,LEGAL_DOCUMENT, paragraph,corpus_paragraph).
bi-lingual_corpus_segment(BI-LING_CORPUS_SEGMENT,
LEGAL_DOCUMENT1,paragraph1,corpus_paragraph1,LEGAL_DOCUMENT2,
paragraph2,corpus_paragraph2).
tri-lingual_corpus_segment(TRI-LING_ CORPUS_SEGMENT,LEGAL_DOCUMENT1,paragraph1,
corpus_paragraph1, LEGAL_DOCUMENT2,paragraph2,corpus_paragraph2,
LEGAL_DOCUMENT3,paragraph3,corpus_paragraph3).
legal_document(LEGAL_DOCUMENT,TITLE,url,
publication_date,passing_date,LANGUAGE,LEGAL_DOMAIN, LEGAL_SYSTEM,
LEGAL_HIERARCHY,LEGAL_QUALITY,AUTHOR,PUBLISHING_HOUSE, PUBLISHING_PLACE).

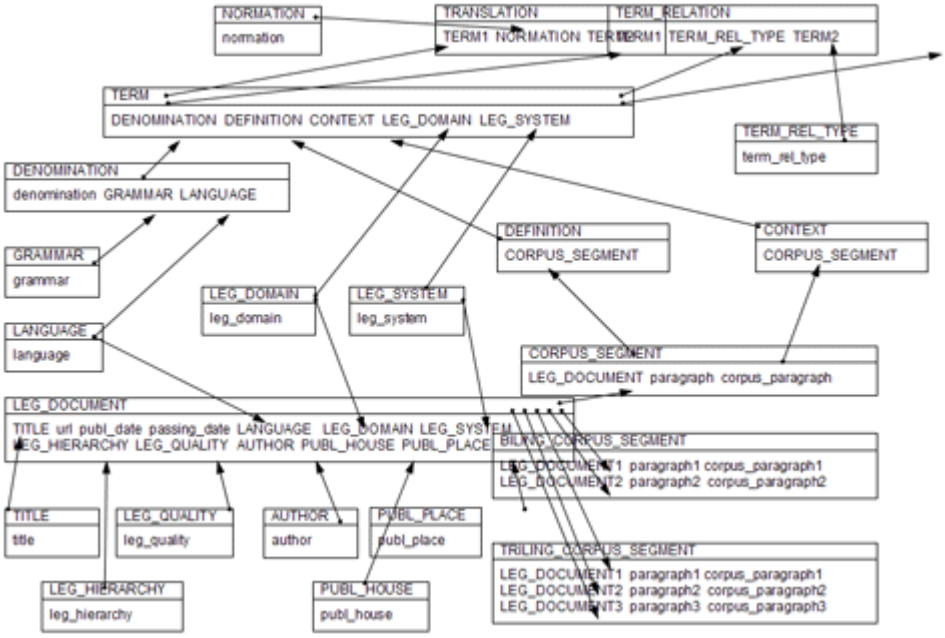


Figure 1: Modell III

Central to our model are terms. They are linked to a word or expression (DENOMINATION). This denomination bears grammatical information and language information. The term belongs to a legal system (e.g. Italy) and a legal domain (e.g. family law). These two attributes are also used to describe legal documents in a corpus. Legal documents are also described with respect to their legal hierarchy (international, national, regional, etc) and their legal quality (law, jurisdiction, treaty, text-book, etc). In addition, legal documents share a great number of bibliographic data, such as title, url, publication date etc. Segments of a legal document may serve as definition or as context for a term. The term and its context and definition should be consistent with respect to the language, the legal system and the legal domain.

Terms are linked to other terms via a relation like antonymy, synonymy, hyperonymy and meronymy on the one side and equivalence and partial equivalence on the other. While the first set of relations connects terms of the same legal system only (e.g. only Italian terms), the latter set is used to connect terms of different legal systems (e.g. Italian and Austrian terms, but also Austrian and German terms). As for the difficult case of legal systems including each other (EU includes Italy which includes the Autonomous Province of Bolzano) we had two options for coding the term relations. According to the first we assume an implicit equivalence and have no explicit connection of terms belonging to different hierarchies. The second option, which we finally followed, consists of marking the equivalence and non-equivalence of all terms explicitly. In fact, although as a general pattern terms from a lower hierarchy level are equivalent to the higher level, this may not be the case where the lower level enjoys a certain degree of autonomy. Terms of different languages and the same legal system (e.g. Swiss German and Swiss French) are also marked as equivalent or non-equivalent (**Figure 1**).

Without going too much into the details of this model, we notice a number of important differences between our model and tree-like entry structures. First, in our model we can cut out an arbitrarily large part of the network as answer to a query. Second, we can take any node of the network as starting point. We thus can do statements not only on terms, but on corpus segments, languages, definitions, normative statements, denominations, legal domains etc. Third, the combination of these two features allows for a flexible investigation and presentation of data.

For example, we can visualise and thus contrast the context citation of two terms which have a specific term relation, e.g. two synonyms, antonyms or hypo- and hyperonyms, compare the definitions of two equivalent terms in different legal systems, list the terms related to a legal domain, identify documents which support the normative statements etc, list false friends, and many more. All these specific queries are predefined in rules (e.g. two false friends have the same denomination but are not equivalent terms). The application of rules is triggered from hyperlinks (e.g. a "false friend"-button) and the output is rendered by specific style sheets, the "false friend"-style sheet.

One important rule identifies one term as a translation of the other. This is used for dictionary-like views on the term-bank. This rule selects all 'equivalent' terms belonging to different languages. It produces a set of descriptively adequate translations. To this set we join the normative translations as stipulated by normative bodies for the South-Tyrolean German and the two variants of Ladin, a Rhaeto-Romance language spoken in the Dolomites, in South-Tyrol.

Still other rules construct mono-, bi- and tri-lingual corpora, including the CATEX corpus of Italian legislation with its German translation (Gamper 1999) and the CLE corpus, an Italian, German and Ladin parallel corpus of regional and local legislation and administration (Streiter et al. 2004). Through the close linkage of terms and corpus, we do not only assist the descriptive and normative terminographical work, but prepare new perspectives on term-presentation. Thus, in near future we may have more than one contextual description of a term, each featuring a different aspect of the term or a dynamic linking of term and context, provided we will identify good heuristics for the ranking of contexts.

3 THE PROPER PLACE OF XML FOR TERM MANAGEMENT

Although XML and relational databases are equally suited to formalise the above model, there are a number of reasons to prefer relational databases, at least when data are written to or read from the hard disk.

Relational databases, be they commercial or free, safely handle simultaneous updates and retrievals, following the consistency requirements specified with the database. It is thus perfectly legitimate to

update a term, for which another author creates a definition, while term, context and definition are seen by an external user (cf. Ahmad, K. & Holmes-Higgin, P. 1996). XML as a text-based medium is organized in large files. They can be opened for writing by one author only at a time.

Validation of data is more immediate with relational databases than it can be with text-based XML. With XML, data are first entered and than validated against a DTD. Although DTD-aware XML-editors might help, the validation of data then is left over to a device at the periphery of the system. All updates which do not go through the editor remain unvalidated until they shall eventually be visualized.

The notion of transaction which assures completed inserts or updates in relational databases is basically absent in pure XML. It has to be added in the form of a commercial transaction server, which takes over the managing of data in a system as complex as a relational database. With relational databases however we have free reliable resources with additional features like user management, network support, and fast retrieval with configurable indices.

The place we assign to XML is another. We use it as middle-ware between the database and the output. Thus while our data are stored in a network of relations, a query returns a table. Reentrances in the network have been pruned off (they are not included in the answer) or have been dissolved by multiplication (e.g. each term has its proper grammar field although derived from one single node originally). The columns in a table can be shown in different orders. We call the leftmost column the TOPIC and it is the TOPIC which determines the sorting of the data. The second column is called the FOCUS and it serves as secondary sorting key.

RELATIONAL NETWORK

```
term(1,5,9,7).
term(2,5,10,7).
term(3,5,11,8).
term(4,5,12,8).
legal_system(7,'AT').
legal_system(8,'DE').
denomination(5,'Kind',6).
grammar(6,N_m).
definition(9,14).
definition(10,15).
definition(11,16).
definition(12,17).
corpus_segment(14,'Ein Kind ...').
corpus_segment(15,'Das Kind ...').
corpus_segment(16,'Ein Kind im ...').
corpus_segment(17,'Kinder im ...').
```

IS ORGANIZED IN A NUMBER OF TABLES:

term id	denomination	definition	legal_system
1	5	9	7
2	5	10	7
3	5	11	8
4	5	12	8

legal_system

id	object
7	AT
8	DE

denomination

id	object	grammar
5	Kind	6

etc...

RELATIONAL DATABASE OUTPUT:

denomination	legal_system	definition
Kind	AT	Ein Kind...
Kind	AT	Das Kind...
Kind	DE	Ein Kind im ...
Kind	DE	Kinder im ...

The tables returned by the relational database are converted into trees, collecting information belonging to the same TOPIC and FOCUS under the same node. An explicit annotation of the first and second column as topic and focus provides the further visualization of the data.

XML-STRUCTURE

```
<terms>
  <denomination function='theme'>

    <object>Kind</object>
    <legal_system function='focus'>
        <object>AT</object>
        <definition function='rheme'>
            <object>Ein Kind...</object>

        </legal_system>
        <definition function='rheme'>
            <object>Das Kind...</object>
        </legal_system>
    </legal_system>
    <legal_system function='focus'>

        <object>DE</object>
        <definition function='rheme'>
            <object>Ein Kind im ...</object>
        </legal_system>
        <definition function='rheme'>

            <object>Kinder im ...</object>
        </legal_system>
    </legal_system>
  </denomination>
</terms>
```

XHTML AFTER XSLT TRANSFORMATION

```
<h1>TERMS</h1>
 <table>
  <tr>
    <td class='highlight'>

      <h2>DENOMINATION</h2>
                <ul>
          <li>Kind</li>
                </ul>
    </td>

    <td class='focus'>
      <table>
        <tr>
          <td class='focus'>
            <h3>LEGAL SYSTEM</h3>
```

```
            <ul>
              <li>AT</li>
                    </ul>
                  </td>
          <td class='rheme'>
            <h3>DEFINITION</h3>

                    <ul>
              <li>Ein Kind ...</li>
              <li>Das Kind ...</li>
                    </ul>
          </td>

        </tr>
        <tr>
          <td class='focus'>
            <h3>LEGAL SYSTEM</h3>
                    <ul>

              <li>DE</li>
                    </ul>
                  </td>
          <td class='rheme'>
            <h3>DEFINITION</h3>

                    <ul>
              <li>Ein Kind im ...</li>
              <li>Kinder im ...</li>
                    </ul>
          </td>

        </tr>
      </table>
    </td>
  </tr>
</table>
```

Arbitrarily large parts of the network may be queried with pre-defined views which connect tables in a PROLOG-like syntax. A basic view, which features a term, a context and a definition with the bibliographic indications looks as follows:

```
view_term(TERM,DENOMINATION,LEGAL_DOMAIN,LEGAL_SYSTEM,
CORPUS_SEGMENT,TITLE,URL,PUBLICATION_DATE,PASSING_DATE,
PARAGRAPH,CORPUS_SEGMENT2,TITLE2,URL2,PUBLICATION_ DATE2,PASSING_DATE2,PARAGRAPH2):-

term(TERM,DENOMINATION,DEFINITION,LEGAL_DOMAIN, CONTEXT,LEGAL_SYSTEM),
definition(DEFINITION,CORPUS_SEGMENT),

corpus_segment(CORPUS_SEGMENT,LEGAL_DOCUMENT, PARAGRAPH,CORPUS_PARAGRAPH),

legal_document(LEGAL_DOCUMENT,TITLE,LEGAL_DOMAIN,
url,publication_date,passing_date,LANGUAGE,      LEGAL_SYSTEM,LEGAL_HIERARCHY,LEGAL_QUALITY,
AUTHOR,PUBLISHING_HOUSE,PUBLISHING_PLACE),
context(CONTEXT,CORPUS_SEGMENT2),

corpus_segment(CORPUS_SEGMENT2,LEGAL_DOCUMENT2, PARAGRAPH2,CORPUS_PARAGRAPH2),

legal_document(LEGAL_DOCUMENT2,LEGAL_DOMAIN2,
URL2,PUBLICATION_DATE2,PASSING_DATE2,LANGUAGE2,      LEGAL_SYSTEM2,LEGAL_HIERARCHY2,
LEGAL_QUALITY2, AUTHOR2,PUBLISHING_HOUSE2,PUBLISHING_PLACE2).
```

A classical terminological entry, composed of a term, its synonyms and equivalents is emulated as:

```
view_entry( ):-
view_term(TERM,DENOMINATION,LEGAL_DOMAIN, LEGAL_SYSTEM,CORPUS_SEGMENT,
TITLE,URL,PUBLICATION_DATE,PASSING_ DATE,PARAGRAPH,CORPUS_SEGMENT2,TITLE2,URL2,
PUBLICATION_DATE2, PASSING_DATE,PARAGRAPH2),
term_relation_types(TERM_RELATION_TYPE2,synonym),
term_relation(TERM_RELATION,TERM,TERM_RELATION_ TYPE2,TERM2),
view_term(TERM2,DENOMINATION2,LEGAL_DOMAIN2, LEGAL_SYSTEM2,
CORPUS_SEGMENT2,TITLE2,URL2, PUBLICATION_DATE2,PASSING_DATE2,PARAGRAPH2,
CORPUS_SEGMENT22,TITLE22,URL22,PUBLICATION_DATE22, PASSING_DATE22,PARAGRAPH22),
term_relation_types(TERM_RELATION_TYPE3,equivalent),
term_relation(TERM_RELATION,TERM,TERM_ RELATION_TYPE3,TERM3),
view_term(TERM3,DENOMINATION3,LEGAL_DOMAIN3,
LEGAL_SYSTEM3,CORPUS_SEGMENT3,TITLE3,URL3,
PUBLICATION_DATE3,PASSING_DATE3,PARAGRAPH3, CORPUS_SEGMENT32,TITLE32,URL32,
PUBLICATION_DATE32,PASSING_DATE32,PARAGRAPH32).
```

4 BISTRO'S TOOLS FOR TERMINOGRAPHY

BISTRO provides a number of term-tools to insert and update data. Term tools are program-modules, the power of which goes beyond that of deductive SQL-statements in a given data-model, however, similar to the SQL-statements they start from a set of entities of the data model and arrive (possibly with human intervention) at another set of entities. The difference actually resides in the type of reasoning involved. While SQL-statements are purely deductive, term-tools allow for knowledge acquisition through induction and abduction. The purpose of term tools is thus to extend the terminological knowledge base on the basis of the available data in the model, new external data, e.g. in URL-located documents and plausibility statements.

In traditional terminography humans initiate and undertake every step of knowledge extension. With electronic term tools knowledge extension becomes much more effective, more consistent and data-driven, but it needs to be well controlled as induction and abduction are not necessarily reliable reasoning modes.

We will present here term extraction, term recognition, text classification and term acquisition through agents.

1. Term extraction
Term extraction identifies terms in text. The quality of the chosen terms prejudices the quality of the produced terminology. While humans usually consider one text at a time, computers can process large amounts of text and add information on absolute and relative frequency, structural resemblances between already described and not yet described terms, and find variants. Our term extraction starts from terms in the data-base, distills models from the terms and applies those models to text. An in-depth description of this tool is provided in "Example-based Term Extraction for Minority Languages: A case-study on Ladin" (Streiter et al. under http://dev.eurac.edu:8080/ autoren/publs/termex5.pdf). The term candidates are then ranked. At this point the quality has to be controlled: Either terminographers assess the automatically produced list of term candidates in context, or other control mechanisms are installed, e.g. only the best term candidate is processed further.

2. Term annotation or term recognition
Term recognition starts from terms and texts and tries to identify texts as contexts of terms or terms as contained in texts. This is a useful application of the terminology, because users can immediately see where a text uses described terminology. They do not loose time triggering searches, sometimes on the wrong lemma and sometimes without results. On the other hand term recognition is also a tool for terminographers, because they can immediately perceive how many terms of the text are already described. If a text contains a lot of already described terms, it is a good text for intensifying the description of a subject field. If a text contains few previously described terms, it is rather a text for extending terminography to other subject fields (which might not be in the scope of the project).

Furthermore, term recognition provides a ranking criterion for term descriptive contexts: A context that contains a lot of described terms is often more precise and gives a more typical context, but its comprehension also demands more domain specific expertise.

3. Text classification

A text classification tool starts from texts and a grid of meta-tags, e.g. language, subject domain, contents and tries to establish the relation between text and grid. Such labeling is necessary for a weighted corpus, for a more targeted searching in texts and helps automatic tools to improve in quality. Our text classification approach and first results are described in Streiter et al. (2003) and Voltmer et al. (2003). For maintaining quality and have nevertheless very early computer assistance in the classification task, we start out with manual classification. As soon as some documents have been classified, the system uses these classified documents to propose a classification of a new document. Humans only need to confirm or modify. Manual classification of data however is not necessary when a term-bank is available. As shown in Voltmer (2005), when combining the contexts and definitions of terms into pseudo-corpora, almost perfect automatic classification results can be obtained based on these pseudo-corpora.

Nevertheless, the classification task in legal terminology is particularly difficult. Lawyers need to know if a text is a valid norm, if it is a law or a court decision and if it is emanated by the Austrian or the German legal system. Even experts could hardly rely merely on the text for such classifications, e.g. humans rely on the authority of an editor, and computers count e.g. on the reliability of official Internet sites. Text taken from the site http://www.bundesverfassungsgericht.de containing the string "entscheidungen" in the URL will most probably contain a valid court decision of the German legal system. For this reason, automatic classification tools in legal terminology must take into account not only the content of the document as input parameter, but those data which describe the publication of the document. In our experience there is no loss in quality when the classification relies on the classification of "external corpora" (here: official legal databases) rather than on the classification of a single text.

In accordance with copyright, these corpora can even be imported to build an internal corpus. Such a classification allows a meta-search on external resources: A search string is sent to the appropriate selection of Internet sites (e.g. an Italian string only to sites with text in Italian) and the results of all relevant sites are reported. A user can then avoid searching in sites where the string is not present or too frequent and he also gets a first feedback on the frequency of the search-string and whether it belongs rather to normative or judicial language.

4. Term acquisition through agents

Given that programs can classify texts from the Internet through text classification, given that another program can recognize the terminological value of a text through term recognition, and given that a term extraction program finds new term candidates, the tools 1, 2 and 3 combine to an agent. This agent starts from certain criteria (language, legal system etc.) and returns with appropriate texts and terms. Agents can even find contexts and definitions. A context search is a full text search restricted to a class of documents (e.g. of the desired language and subject domain). Term recognition on the results ranks them by relevance or even subject domain relevance.

A definition search (implemented also by GOOGLE: "http://www.google.com/search?q=define:XXX") over legal resources is a context search with further restrictions. The search string is included in definition-indicating context. The term "A" would be searched with "The definition of A is", "A is defined as", "A is a", "A is the", and so on. Such a definition search does not allow for automatic collection of definitions, but according to our experience it is an extremely valuable tool for terminographers, because they can start out from various definition-similar contexts. Their attention is immediately drawn to polysemy, competing definitions in different contexts (legal systems) and to the different starting points for the definition (inside subject domain, inside LSP, definition with common language).

The term-tools can operate on their own output. They can be used to intensify the description of one subject domain, to extend the description to other domains and even to control the quality of the existing terminology. The described tools can indicate the contexts which have a low occurrence of domain specific terms and propose contexts with a higher occurrence of LSP-terms.

5 TERM PRESENTATION

This section will offer an overview of how we designed the interface and term presentation in order to guarantee a user-friendly work-platform and an intuitive presentation of instruments, tools and data. In an idealized description we could characterize the interface of BISTRO as a browser to the described data model and its extension through term tools. The user thus can approach the same data from different points of view, arriving after each hyperlink at a new node in the data model from which the data can be continued to be explored, zoomed in or seen with a different focus. In order to reduce the unexpected complexity, some of the possible paths have been pruned off and some 'standard' views are provided, e.g. the entrance page. At the top bar BISTRO's first page offers a series of all the tools that might be used: term search, corpus search, term tools and book search. Before starting the search, the term search and its presentation offer the possibility to choose between searching through all the terms normed by a terminology commission (button TERKOM) or to see all the data (button BISTRO). The search can be done in all three languages Italian, German and Ladin (in two variants).

The search returns a hit list, containing the searched terms and similar terms. That means that looking up the Italian word "atto" results in a list featuring in first column the searched term “atto”, followed by compound terms like "atto amministrativo", "atto collegato" etc. The second column contains the equivalents in target languages, in this case, German and Ladin. Clicking here checks the back-translation for translation validation. The third column refers to to the legal domain of the term. Clicking here sets the focus on normative documents or the legal domain or allows to zoom in on terms from a specific normative document or legal domain. The last column contains a number of links which elaborate the term in various directions (zoom on one terminological entry or validation/exploration with WWW-search and corpus queries) (figure 2). The zoom to the terminological entry returns a description of a term in three parts: the first containing domain (e.g. administrative law) and sub domain, registration number and responsible terminographers, the second containing the term, its definition and context including the relative bibliographical data. The third part finally contains the normative data which specify the status of the term, e.g. normed, refused or waiting for decision and in case of normation the information on the normative document issued by the region Trentino-South Tyrol. The entry is hyperlinked according to our data model, e.g. providing links to the bibliographical data to every single context or definition.

(1)	**atto** N-m (giuridico)	Rechtshandlung N-f	law of obligations	* bistro
(2)	(documento)	Urkunde N-f	meta	* bistro
(3)		Handlung N-f	penal law	* bistro
(1)	atto avente forza di legge N-m	Akt mit Gesetzeskraft N-m	meta	* bistro
(2)	atto con forza di legge N-m	Akt mit Gesetzeskraft N-m	meta	* bistro
(3)	antefatto non punibile N-m	straflose Vortat N-f	penal law	* bistro
(4)	antefatto N-m	Vortat N-f	penal law	* bistro
(5)	contratto di lavoro N-m	Arbeitsvertrag N-m	labor law	* bistro
(6)	contratto individuale di lavoro N-m	Individualarbeitsvertrag N-m	labor law	* bistro
(7)	atto ablativo N-m	entziehender Verwaltungsakt N-m	admin.law	* bistro
(8)	atto ablatorio N-m	entziehender Verwaltungsakt N-m	admin.law	* bistro
(9)	atto a complessità esterna N-m	komplexer Verwaltungsakt unter Beteiligung verschiedener Körperschaften 0	admin.law	
(10)	atto a complessità ineguale N-m	komplexer Verwaltungsakt ungleichwertiger Willensträger 0	admin.law	
(11)	atto a complessità interna N-f	komplexer Verwaltungsakt unter Beteiligung verschiedener Organe ein und derselben Körperschaft 0	admin.law	* bistro
(12)	atto a forma scritta N-m	Verwaltungsakt in schriftlicher Form N-m	admin.law	* bistro
(13)	atto amministrativo N-m	Verwaltungsakt N-m	admin.law	* bistro
(14)	atto ampliativo N-m	begünstigender Verwaltungsakt N-m	admin.law	* bistro
(15)	atto automatico N-m	beherrschbare spontane Handlung 0	penal law	* bistro
(16)	atto collegato N-m (con altro atto)	in Beziehung stehender Verwaltungsakt 0 (mit einem anderen)	admin.law	
(17)	atto collegiale N-m	Kollegialakt N-m	admin.law	* bistro

Figure 2

The WWW-search of terms allows meta-search selected official internet sites on legal issues in Italy, Germany, Austria and Switzerland. Web-sites which have recently been used for the documentation of terms are preferred. Terminographers classify the sites to guide the WWW-search. By searching a term in CATEX (the bilingual Italian-German corpus) or in CLE (the trilingual Italian-German-Ladin corpus), BISTRO offers a special search mask that again permits the user to link his search, deciding to look up terms contained in documents edited before or after 2001, in specific types of documents or a specific legal system. The number of hits can be chosen, too. The searched terms are highlighted in the contexts

and variants found. Finally Bistro also contains a KWIC tool for all three languages (Italian, German and Ladin).

The idea behind BISTRO is thus to grant the user maximal autonomy with a proper perspective and focus. The approach could be termed equally 'constructivist' as users have access to the same term tools which terminologist used for the elaboration of the terminology. At the moment the user can decide to select between normed and not normed terms, choose the target language, the domain, to set a focus and to advance through the data through hyperlinks and term tools e.g. from term to context to bibliographical data, to similar bibliographical data, to associated corpora, to term extraction on the associated corpora and a KWIC of the extracted terms etc.

6 CONCLUSION

In this paper we have presented a new approach to managing terminological data, which is based on a structured database model. By describing the different subparts of the system, we have discussed its strengths in respect to features crucial to terminological systems like data consistency, powerful querying facilities and user adaptiveness. Further we have presented so called term tools for semi-automatic terminology work. In the end we have addressed the question of data representation and provided examples of how data is currently displayed.

In the BISTRO system all relevant information like term entries, definitions, bibliographical information and corpus segments are stored as a network of interlinked entities. The data kept in a relational database is displayed in the form of ordinary tables comprising different cutouts of information relevant to the user. The transformation from the data level to the graphical representation is completed by the help of XML.

Presenting our abstract data model we could show that a well designed net structure with a normalized set of atomic entities creates a framework for development and storage of a valid set of formalized, unambiguous and non redundant terminological data. By providing a facility for structured and modular data organization the proposed system overcomes a main shortcoming of previous models. As we emphasized before, modularity and coherency of data are a fundamental prerequisite for long term maintainability of data.

At the same time organizing data as a network enables us to autonomously access different parts of information content without the need to always consider a complete set of information. Addressing the question of flexible data access we demonstrated how our model is capable of resolving complex queries concerning relations between different parts of the network. We showed why powerful querying facilities are important for both efficient information access and user adapted data presentation.

Further we discussed the advantages of an architecture combining a relational database with XML representations over approaches solely based on XML. Relational databases are designed to ensure data validity at all times. Therefore one of its strengths is to provide mechanisms that handle simultaneous data access and modification automatically, whereas XML serves well for the flexible presentation of data to the user. We could show that the BISTRO system integrates these two techniques to fully exploit the benefits of each one. The BISTRO system is freely accessible online, where the user can choose between different search masks and data displays.

Overall we demonstrated that the proposed approach integrates different ideas and technologies to overcome well known problems that all terminological information systems have to struggle with. By doing that we could show the strengths of the BISTRO system in respect to the main requirements for a terminological system.

REFERENCES

AHMAD, K. & HOLMES-HIGGIN, P. (1996). Is your Terminology in Safe Hands? Data Analysis, Data Modelling and Term Banks, Terminology and Knowledge Engineering. In GALINSKI, C., SCHMITZ, K.D. (eds.) Proceedings of the 4th International Congress on Terminology and Knowledge Engineering, Vienna 1996.

GAMPER, J. (1999). Construction of a Parallel Text Corpus Encoding Primary Data. In: Academia, 18 1999. Bozen, 32-33.

POWERS, S. (2003). Practical RDF, Solving Problems with the Resource Description Framework, O'Reilly, Sebastopol CA.

STREITER, O. & VOLTMER, L. (2003). A Model for Dynamic Term Presentation. In: Terminologie et Intelligence Artificielle TIA-2003, Strasbourg 2003.

STREITER, O. & VOLTMER, L. (2003). Document Classification for Corpus-based Legal Terminology. In: VRABIE, G., Turi, J.G. (eds.) The theory and the practice of linguistic policies in the world. 24-26 May 2002, Iasi 2003.

STREITER, O., ZIELINSKI, D., TIES, I. & VOLTMER, L. (2003). Term Extraction for Ladin: An Example-based Approach. In: Actes de la Dixième Conférence Nationale sur le Traitemen Automatique des Langues Naturelles (TALN 2003), Batz-sur la Mer 2003.

STREITER, O., STUFLESSER, M. & TIES, I. (2004). CLE, an aligned Tri-lingual Ladin-Italian-German Corpus. Corpus Design and Interface. In: Workshop on "First Steps for Language Documentation of Minority Languages: Computational Linguistic Tools for Morphology, Lexicon and Corpus Compilation", Lisbon 2004.

STREITER, O., TIES, I., RALLI, N. & VOLTMER, L. (to appear). BISTRO, the online platform for terminology management: structuring terminology without entry structures. In: TEMMERMAN, R. & KNOPS, U. (eds.) Linguistica Antverpiensia, Antwerp.

STREITER, O., ZIELINSKI, D., TIES, I. & VOLTMER, L. (to appear). Example-based Term Extraction for Minority Languages: A case-study on Ladin. In: Soziolinguistica y Language Planning, Ortisei 2002.

VOLTMER, L., STREITER, O., ZIELINSKI, D. & TIES, I. (2003). Termextraktion durch Beispielterme Ansätze und Versuchsergebnisse für eine Minderheitensprache, Eurac Workingpaper October 2003.

VOLTMER, L. & STREITER, O., Klassifizierung von Korpora für die Rechtsterminologie, Eurac Workingpaper October 2003.

VOLTMER, L. (2005). Werkzeuge für Rechtsdatenbanken - Über computerlinguistische Verfahren zur Untersuchung, Speicherung und Kommunikation rechtlichen Wissens, Diss. LMU München 2005.

Birthe Toft
Institute of Business Communication and Information Science
University of Southern Denmark

ONTOLOGY IN SEMANTIC WEB TECHNOLOGIES – RE-INVENTING THE WHEEL?

Abstract

This contribution deals with the following issues related to ontology: the polysemic nature of the term and its application in philosophy of science, information technology, and terminology, as well as the viability of creating a common meta-ontology for terminology and information science, a meta-ontology which might build on Alwin Diemer's general ontology as described in Gerhard Budin's doctoral thesis.

INTRODUCTION

The question I have asked in the title is in my opinion an extremely relevant one. However, it has become more than just relevant and interesting to me because it is related to a discipline I have to teach in the very near future. It is one of several disciplines of a new master's programme in Web communication at the University of Southern Denmark, which is intended to provide students with 'relevant cross-disciplinary qualifications' in web-based communication, information structuring, and IT design strategies, and among the career opportunities suggested are those of information architect or knowledge manager.

The discipline in question is called Knowledge Modelling. According to the description, two of its main aims are to provide students with

1. Understanding of how **conceptual hierarchies** or **ontologies** can support knowledge structuring and knowledge sharing
2. Competence in applying **terminological** and other methods and tools appropriate for semantic **modelling** and representation of knowledge

Of course the coupling of conceptual hierarchies and ontologies will not come as a surprise to this target group. But what is ontology and ontologies, both from a terminological and from an information technology point of view, and is it possible to establish a common theoretical basis for the creation of conceptual hierarchies and ontologies without, as it says in my title, re-inventing the wheel?

ONTOLOGY AS A BRANCH OF PHILOSOPHY OF SCIENCE

The etymology of the word ontology is Greek. According to Sowa (2000: 55 – 57), the first element **tó on**: (what is, Being) stems from Aristotle, the second one from Heraclitus: **lógos** (word, reason, principle, plan, doctrine). In spite of the ambiguity of the latter word, there seems to be general agreement that ontology can be rendered **the doctrine of what exists**.

Today the investigations of Aristotle (c. 384 – 322 b.C.) and of Porphyry (c. 234 – 305 a.C.) would be ranked under the label of ontology, but according to Robering (2004: 8), ontology as a branch of philosophy is an invention dating from the Enlightenment, defined by the German philosopher Christian Wolff (1679 – 1754) as the **general science of being**.

After a long period of neglect during the reign of Empiricism and Positivism, ontology was 'rehabilitated' a few decennia ago.

TERMINOLOGY AND ONTOLOGY

Wüster's famous 'Word model' illustrates, inter alia, the relationship between concepts and individual entities in extra-linguistic reality (Wüster 1959: 303), and the title of his programmatic speech (Wüster 1974), held in May 1972 at the University of Vienna, comprises the following definition of terminology:

> a field in the borderland of linguistics, logic, **ontology**, information science, and the subject fields

However, Wüster found that terminologists need not concern themselves with ontology unless in so far as it is necessary in order to ascertain what conceptual relations are possible (Wüster 1974: 360 – 61). The necessity, he found, arose whenever logic did not suffice to describe all conceptual relations; some relations based on contiguity in space and/or time had to be defined using ontological categories.

To sum up, in Wüster's time ontology was viewed simply as a branch of philosophy which must be 'consulted' because logic does not suffice to explain the nature of all conceptual relations, whereas in 2004 it is viewed by most as an information technology engineering field applying concepts and methods extremely similar to those of terminology.

In Budin's words, terminologies and ontologies are the intellectual (conceptual) infrastructures of content (Budin 2002: 100), and terminology theory has become a major foundation of recent ontology engineering and provides a solid basis for designing knowledge organization systems for the Semantic Web (Budin 2003: 78). I shall return to Budin's ideas of ontological organisation below.

From the beginning, terminology followed a formal course; in Alexeeva's words, it came under the influence of logic (Alexeeva 2003: 66ff). I also agree with her that terminology has now taken a philosophical turn in that formal and logical aspects of terminology are gradually substituted by theoretical and cognitive ones, and that we need a philosophy of terminology.

In fact, I suggest that we divide a new, scientifically updated discipline of terminology into the following subdisciplines or components:

(1) A **philosophy of science and language component**, including ontology, epistemology, conceptology, and sign theory. Combining philosophy of science with philosophy of language makes perfect sense, not least in connection with terminology

(2) A metascientific component dealing with relationships among sciences and between terminology and science

(3) A linguistic component

(4) A **pragmatic component** dealing with terminography, modelling, and representation of conceptual structures in various media

Of the above components, (1) and (4) comprise issues of clear relevance to information technology. In what follows, I shall concentrate on the elements of (1), which is a prerequisite to seriously dealing with the issues in (4) — a fact which is often neglected, particularly in information technology.

ONTOLOGY IN INFORMATION TECHNOLOGY

Today, every discipline dealing with the management of information or content of some kind or other seems to be searching for the grand final ontological model to build on.

From an information technology perspective, the different meanings of the term ontology may be grouped under two headings, which are named as follows by Daconta et al. in their textbook on the Semantic Web (2003: 186):

1) **BIG O**: A branch of philosophy (of science)

2) **Little o**:
a) an engineering discipline, also called ontological engineering
b)engineering applications, i.e. conceptualizations establishing joint terminologies between the members of specific communities of interest

As will become clear from the section below, I should like to add one more meaning to the list, viz. **meta-ontology**.

Engineering applications, corresponding to meaning 2 b) above, are arranged by Daconta et al. (2003: 157) in an **ontology** spectrum ranging from the semantically weak to the semantically strong applications:

- **Taxonomies** represent knowledge with minimal hierarchic or parent/child structure, if any
- **Thesauri** comprise relations among words and synonyms
- **Conceptual models** comprise more complex knowledge
- **Logical theory** represents knowledge in a rich, complex, consistent, and meaningful way

Thus the generic term ontology is used to cover all the conceptualisations and corresponding applications above; however, the term ontology is also often used about a particular subtype of ontology represented in a so-called web ontology language; in the ontology spectrum, this subtype can be found at the semantically rich end (Daconta et al. 2003: 181).

THE NEED FOR A COMMON META-ONTOLOGY

As mentioned above, one meaning of the term ontology needs to be added to the list in between Big O and little o ontology, viz. meta-ontology. There is probably no agreement whatever as to the meaning of that term, but at least a useful information technology definition can be found in Cappelli et al. (2004: 86):

> A methodology (of ontology design) based on formal philosophical notions general enough to be used independently of a particular domain

I prefer the definition below, which explicitly mentions the purely theoretical aspect:

> The theory behind and the methodology of ontology design

At a high level of abstraction, it must be possible to create a common meta-ontology for terminology and information technology. And instead of re-inventing the wheel, why not exploit existing approaches, one of them described nearly 10 years ago by Gerhard Budin?

In 1996, Gerhard Budin introduced his theoretical WIKO meta-model to describe the dynamics and complexity of scientific information and communication processes, based on the hypothesis that the domains of information science, LSP, applied linguistics, and the discipline of terminology call for a common meta-theoretical approach.

He used the concept of **organisation** as a complex, dynamic basic principle capable of integrating knowledge, information, communication, and terminology and showed that integration could take place not only at a metascientific level, but also at the concrete modelling level.

The concept of **ontological organisation** on which Budin based his model originates from Alwin Diemer's general ontology, described below.

ALWIN DIEMER'S GENERAL ONTOLOGY

Diemer's general ontology consists in a typology of objects. An English version of Budin's table showing this typology can be found in table 1 (Budin 1996: 28). I agree with Budin that it is an astonishingly flexible and generally applicable instrument for differentiating ontical phenomena.

Table 1

ALWIN DIEMER'S OBJECT TYPOLOGY	ELEMENTARY ENTITIES				INTENTIONAL ENTITIES			
	SIMPLEX		COMPLEX		SIMPLEX		COMPLEX	
ELEMENTALS	ELEMENT	STATIC: thing, property	WHOLE / MOLECULAR (STATIC/DYNAMIC): set, aggregate, collective, system, organism		INTEN-TIONAL ELEMENT	STATIC: sign, symbol	INTEN-TIONAL WHOLE	STATIC: product of art, science, etc.
		DYNAMIC: event				DYNAMIC: planned act or event		DYNAMIC: meaning system, e.g. society
INTERALS	INTERAL	STATIC: relation, function	RELATI-ONSHIP	STATIC: structure organisation	INTEN-TIONAL RELATION	STATIC: meaning, reference	INTEN-TIONAL COM-PLEX RELA-TION	STATIC: meaning complex, context
		DYNAMIC: process		DYNAMIC: coupling		DYNAMIC: action		DYNAMIC: communication, interaction
TOTALS	TOTAL (field)		ENVIRONMENT		WORLD			

It is especially useful in a systemic approach to modelling since it is based on the principle of the common source of the three basic phenomena: element, **relation**, and **totality** (Diemer 1959: 72). No doubt Budin is right that this **triadic** principle allows the integration of three normally incompatible ontological/epistemological positions: Atomism, Relationism, and Holism (Budin 1996: 27).

Diemer's triadic approach is only one among several similar **trichotomies** described by Sowa (2000: 58 – 67), i.e. those of Kant, Hegel, Peirce, Husserl, Whitehead, and Heidegger. Charles S. Peirce defines three modes of being: firstness (unstructured or monadic being), secondness (binary, relational being), and thirdness (complex, structured being) (Peirce 1903: 1.23-6). According to Sowa, it is by far the most appealing and useful among the triadic approaches; although strongly grounded in logic, it avoided the split which later developed between logic-based analytic philosophy on the one hand and phenomenology and existentialism on the other (Sowa 2000: 66).

Whatever approach is chosen, what is important to terminologists and others is to agree on truly basic principles of ontological organisation before turning to methodological questions. In my opinion, Budin's interpretation of Diemer's general ontology enables us to do just that.

BIBLIOGRAPHY

ALEXEEVA, L. (2003): Interaction between terminology and philosophy. In: IITF Journal, Vol. 14 (2003). p. 64 — 73.

BUDIN, G. (1996): Wissensorganisation und Terminologie. Die Komplexität und Dynamik wissenschaftlicher Informations- und Kommunikationsprozesse. (Universität Wien: Habilitationsschrift). Tübingen: Gunter Narr Verlag.

BUDIN, G. (2002): Computational Terminology for Transcultural Content Management. In: Proceedings of TAMA '02, p. 95 – 106.

BUDIN, G. (2003): Prospects of a Philosophy of Terminology. In: IITF Journal, Vol. 14 (2003). p. 74 — 83.

CAPPELLI, A., GIOVANNETTEI, E., MICHELASSI, P. (2004): Ontological Knowledge and Language in Modelling Classical Architectonic Structures. In: Oltramari et al. (eds.): Proceedings of the LREC 2004 satellite workshop 'Ontolex 2004 – Ontologies and Lexical Resources in Distributed Environments. p. 85 — 89. Lisbon: ELRA.

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. Wiley Publishing Inc.

DIEMER, A. (1959): Einführung in die Ontologie. Verlag Anton Hain K.G.

ROBERING, K. (2004): Objekte als Gegenstände des Nachdenkens - Einführung in die Seminarlitteratur. University of Southern Denmark.

PEIRCE, C. S. (1903): Collected Papers (excerpts). In: Buchler, J. (ed.): Philosophical Writings of Peirce. New York: Dover Publishing 1955.

SOWA, J.F. (2000): Knowledge Representation; Logical, Philosophical, and Computational Foundations. Brooks/Cole: Thomson Learning.

Wüster, E. (1959): Das Worten der Welt; schaubildlich und terminologisch dargestellt. In: Laurén & Picht (eds.): Ausgewählte Texte zur Terminologie. Wien: TermNet, 1993, p. 302 – 330.

WÜSTER, E. (1974): Die Allgemeine Terminologielehre - ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sachwissenschaften. In: Linguistics, no. 119, reprinted in Laurén, C., Picht, H. (eds.): Ausgewählte Texte zur Terminologie, p. 331 - 374. Wien: Infoterm 1993.