Multilingual Terminology Work in Theory – and in Practice

Erdman Thomsen, Hanne; Madsen, Bodil Nistrup; Lassen, Tine

Document Version
Final published version

Published in:
Multilingualism in Specialized Communication: Challenges and Opportunities in the Digital Age

Publication date:
2016

Multilingual terminology work in theory – and in practice

Thomsen, Hanne Erdman; Madsen, Bodil Nistrup; Lassen, Tine
Department of International Business Communication, Copenhagen Business School
Denmark

Correspondance to: het.ibc@cbs.dk

Abstract. In theory, multilingual terminology work is done by creating concept diagrams in each of the languages and comparing them to establish equivalences between concepts in the two languages. In practice, however, various terminology management systems (TMS) are used, end these systems hardly ever support the ideal working method. First of all, only very few integrate adequate tools for modelling concept systems. Second, the data structure and the user interface do not support the process of linking entries in two languages. Concerning the data structure, the understanding of "concept oriented" plays a major role. In many cases the concept is perceived as a unit at the interlingual level, and in the data structure an entry corresponds to one concept with terms from several languages connected. In other cases, the concept is seen as language-specific, and in the data structure an entry contains equivalent concepts from the languages under consideration. In this paper, we illustrate the clash between theory and practice with an example, and outline the requirements for an optimal data structure and user interface that would allow theory and practice to meet.

Keywords. Terminology work, terminology management system, TMS, data structure, translation equivalents, concept orientation

1. Introduction

In the late 1990's, practical multilingual terminology work led us to propose a new structure for handling equivalence (Hull, Madsen, and Thomsen 1998; Madsen and Thomsen 1998). In this paper, we introduce this proposal again, this time based on our work in the research project DanTermBank (Lassen et al. 2013), a project creating foundations for a national term bank, including development of tools for automating terminology work, a revision of data categories for terminology databases in general (Madsen et al. 2013) and a proposal for a new structure for termbases.

We briefly introduce terminology work in theory, and show that state-of-the-art TMSs do not support the theory in practical work. Based on a concrete example of English-Danish terminology work, we repropose a revision of the data structure in termbases and sketch a user interface that will support practical terminology work procedures in accordance with the theory.

2. Multilingual terminology work in theory

Terminological theory recommends that multilingual terminology work is carried out by analyzing the terminology of one language at a time, and finding equivalents in a second step. This is done in order to account properly for differences in the conceptualizations of two or more cultures, were there may be cases of partial equivalence or gaps in one language compared to another. (ISO 860 :2007) is the only standard addressing multicultural terminology work (harmonization of concepts and terms), and also here the above method is prescribed: The first stage after deciding the scope of the work, is comparison of concept systems, which are assumed to exist before the harmonization starts.

The analysis of the terminology in a given domain comprises the elaboration of concept systems and registration of all synonymous terms for each concept, definitions and other relevant information (see for instance ISO 704: 2009). In the DanTermBank project, concept systems take the form of terminological ontologies, which means that they also include characteristics and subdivision criteria, as illustrated in Fig. 1. Characteristics are represented as attribute-value pairs below the concepts, and subdivision criteria are written in white boxes which span over the relations leading to the concepts they differentiate. Terminological ontologies are described in more detail in (Madsen 1998; Thomsen 1998; Madsen, Thomsen, and Vikner 2004).

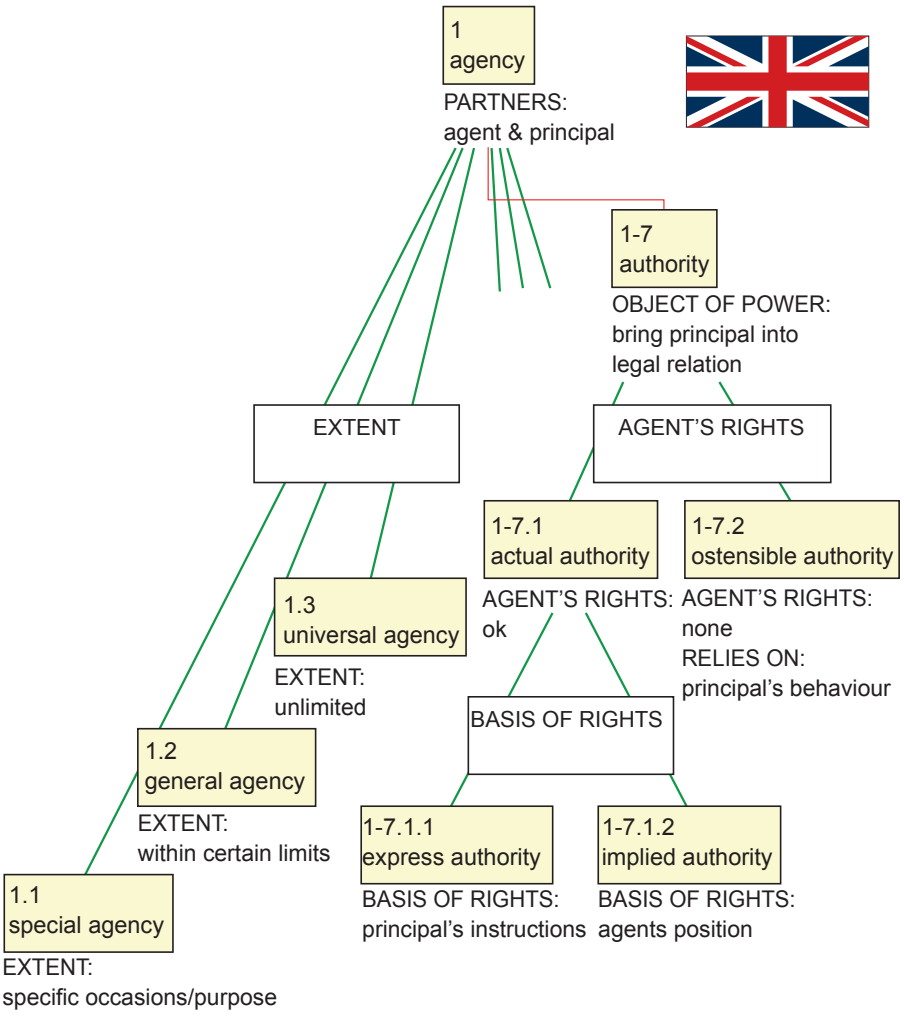

Figure 1: Terminological ontology of selected concepts from the domain of agency in the UK.

3. Multilingual terminology work in practice

3.1. Using a TMS

In practice, most terminologists use a terminology management system (TMS). Generally, this makes it difficult to carry out the work in accordance with terminological theory as we will illustrate below.

In a TMS, terminological data are stored in a concept oriented structure, which means that the concept is the central unit, and for each concept, information on several languages is registered, e.g. definition and all more or less synonymous terms. This structure is also prescribed in (ISO 30042 :2008) and is depicted in Fig. 2.

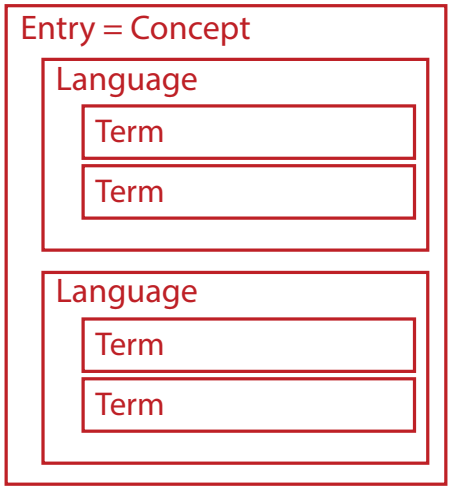

Figure 2: Concept oriented structure of entries according to TBX

The first difficulty is that, with a few exceptions, these systems do not support concept system development. This means that concept systems have to be drawn with other tools, separate from the registration of information on each concept, so information from the concept systems is not directly integrated into the TMS. The terminological ontology in Fig. 1 is developed in the TMS i-Term® (Madsen, Thomsen, and Wenzel 2006; DANTERMcentret n.d.), developed at the DANTERMcentret at Copenhagen Business School. In this tool, the terminological entries are directly accessible from the concept system and vice versa, and information generated in the diagram is also presented in the entry, as shown in Fig. 3, where information on characteristics and related concepts come from the concept system.

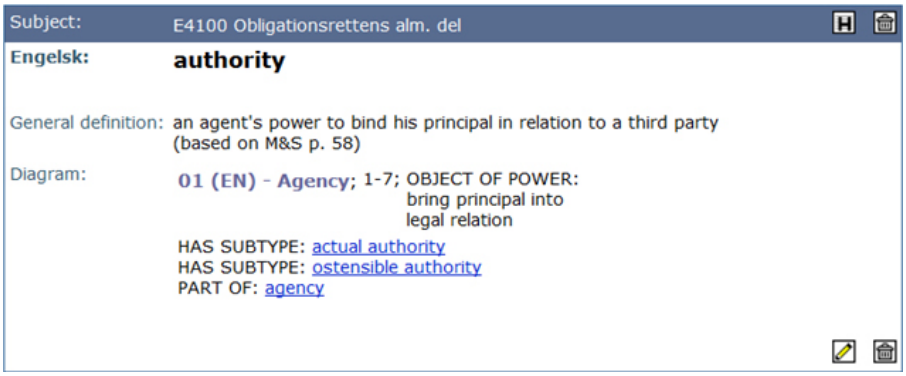

Figure 3: Terminological entry in the TMS i-Term®

Another difficulty is that systems do not support the recommended procedure for multilingual work. TMSs are constructed to **store the result** of terminology work, which was a big step forward when the first systems were developed, but today focus should be on adjusting the systems to also **support the terminological working process**. Current systems do not allow the users to analyze the terminology in one language at a time and then combine the equivalent entries afterwards. Instead, information on concepts in a second language must be entered into the entries of their equivalents in the first language.

In cases where concepts in two languages are equivalent, the basic structure of termbases according to TBX does not exclude system support of the recommended working method, but it requires that systems include a user interface that will move the language section of one entry into another entry, as illustrated in Fig. 4, when the user confirms the equivalence.

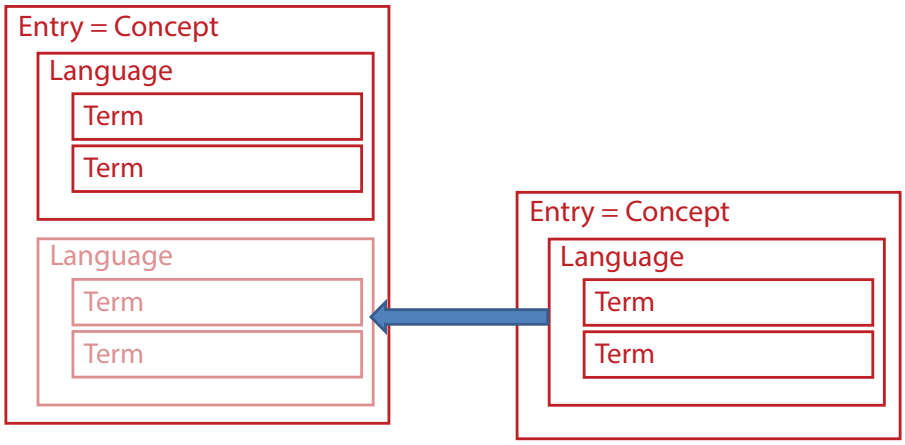

Figure 4: Moving language section from one entry into another

3.2. Cultural differences

In many areas, however, there is not a one-to-one correspondence between concepts in two languages (or cultures). One such area is agency in the case of UK English and Danish. In Fig. 5, the terminological ontology from Fig. 1 is compared to a corresponding ontology of Danish concepts pertaining to agency in Denmark. In this case, there are two examples where one concept in one language does not correspond to only one concept in the other language.

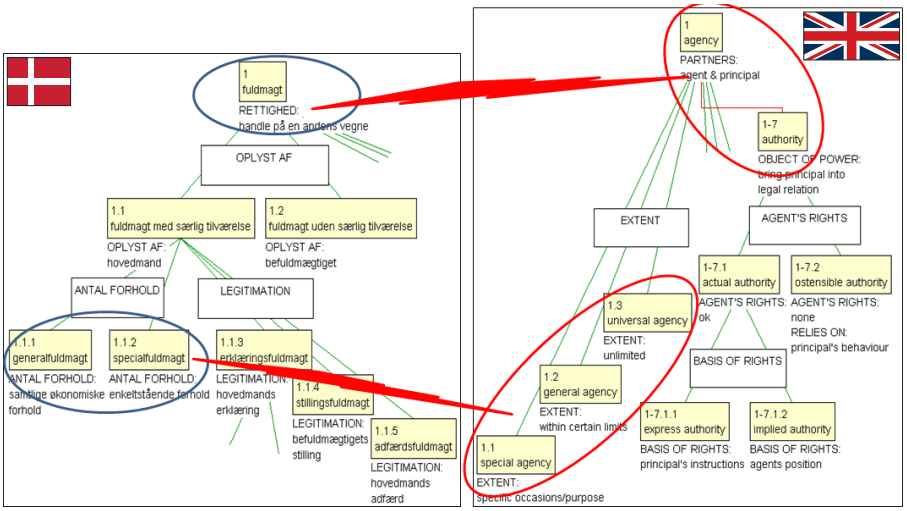

Figure 5: Terminological ontologies of selected concepts from the domain of agencey in the UK and in Denmark

The Danish concept *fuldmagt* corresponds to both *agency* and *authority* in the UK. In Fig. 6, the entries corresponding to these three concepts are shown, with arrows indicating the equivalences.

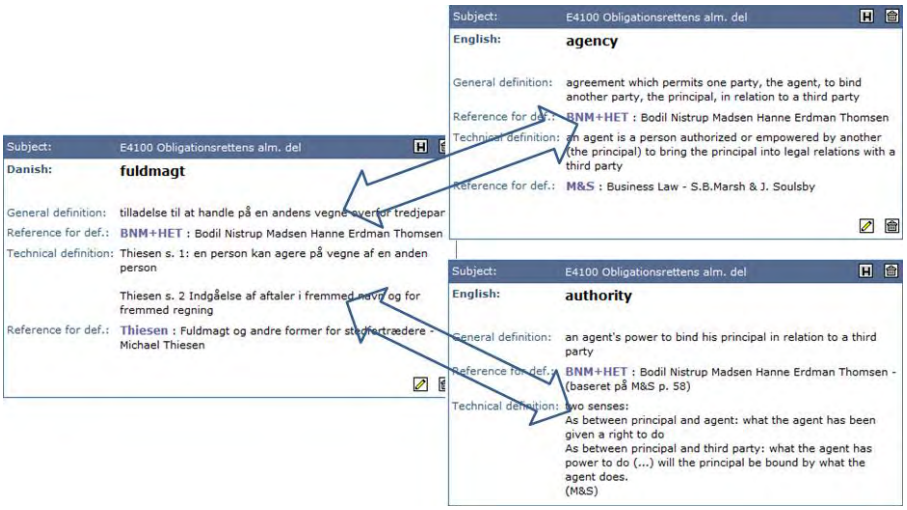

Figure 6: Entries for one Danish concept and the two English equivalents

The procedure described above of copying the language sections of the English equivalents into the Danish entry would result in two English language sections in one entry as illustrated in Fig. 7. This solution is not satisfactory, as it would indicate that the two English concepts are equivalent, i.e. synonymous, which they are not.

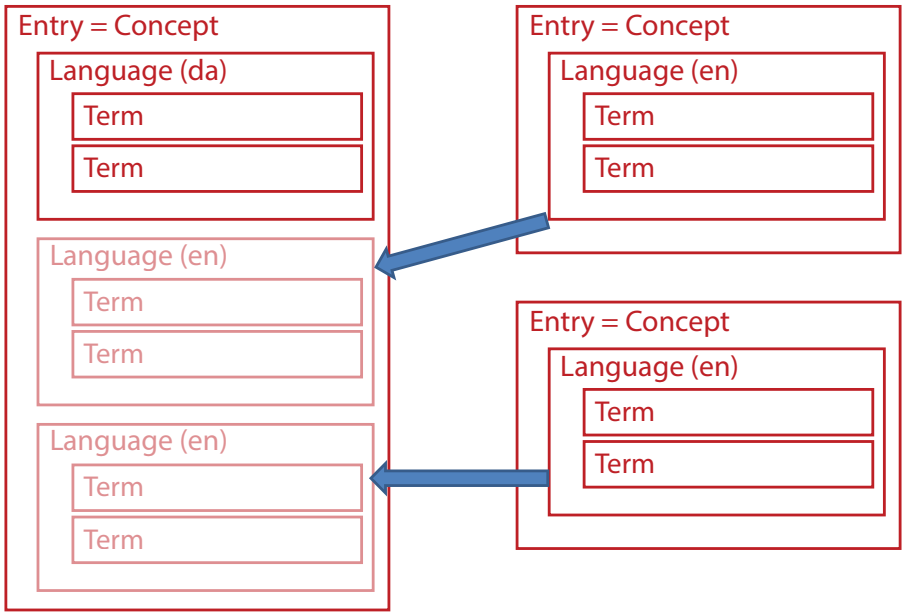

Figure 7: Two English language sections moved into one entry containing a Danish language section

In practice, the current structure of TMSs is thus not suitable for handling such cases of one-to-many equivalence. Instead, terminologists have to create doublettes of the entry in the one language, here the Danish entry for *fuldmagt*, and move one English equivalent language section into each of the entries containing identical Danish language sections, as it is done in Fig. 8.

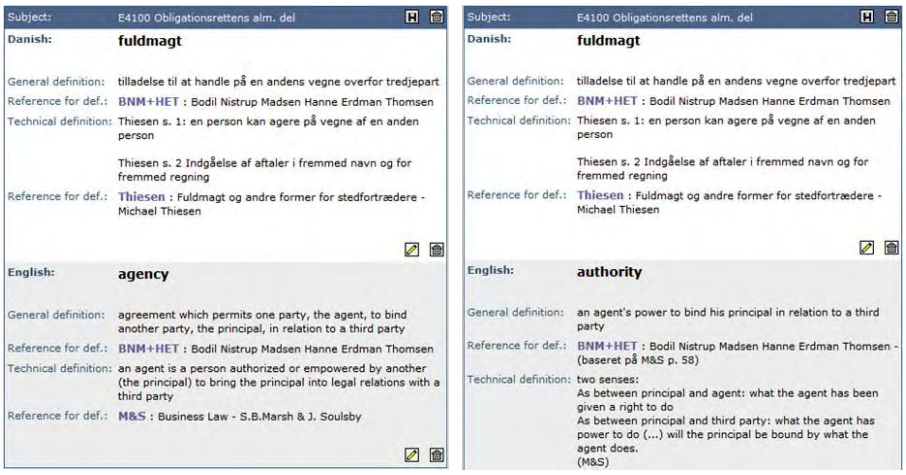

Figure 8: Danish doublettes

This is common practice, but not a satisfactory solution. When searching the termbase in Danish, *fuldmagt* will return two search results, indicating that there are two homographs, which is not the case. Furthermore, updating the termbase becomes complicated, since changes to the Danish concept *fuldmagt* requires the change to be duplicated, and this is contrary to the idea of a database: database integrity is compromised. Finally, if the termbase comprises more than two languages, which is often the case, duplications may increase even further.

4. Revising the data structure

In order to handle one-to-many equivalence, the structure of termbases must be revised. The structure needed, must be able to handle the situation illustrated by the example in Fig. 6. Instead of combining more languages in one entry, we propose to have separate entries for each language section and introduce equivalence as a relation between entries, as shown in Fig. 9.

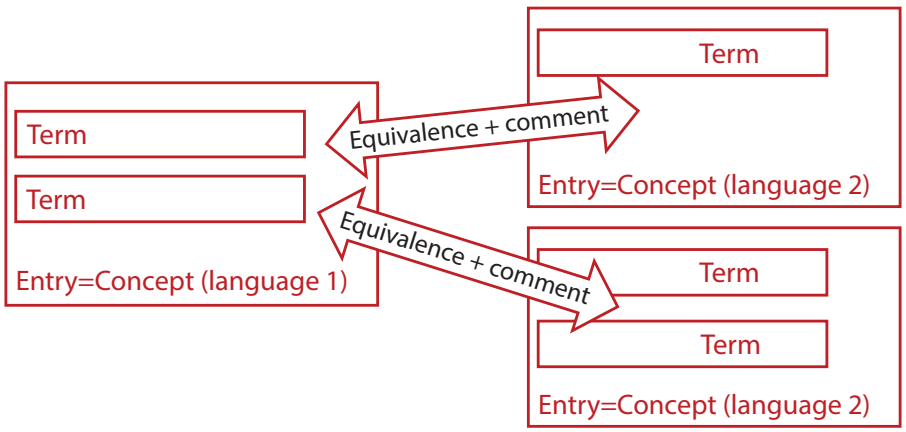

Figure 9: Proposed new structure for termbases

The original structure in Fig. 2 indicates that concepts are cross-lingual, whereas the structure in Fig. 9 implies that concepts are language-specific. As the example in Fig. 5 shows, at least some concepts are language-specific. On the other hand, within some domains, there are cases where concepts may be viewed as cross-lingual. The proposed structure in Fig. 9 can handle practical terminology management in both of these cases in a satisfactory way, whereas the structure in Fig. 2 covers only the fully cross-lingual case.

We first proposed the structure in Fig. 9 in (Hull, Madsen, and Thomsen 1998) and (Madsen and Thomsen 1998), and in (Thomsen 2016) it was reintroduced. In the DanTermBank project mentioned above, it was decided to use this structure in a future national term bank. We recommend that this structure is also integrated in future revisions of TBX (ISO 30042 :2008), TMS (ISO 26162 :2012) and other ISO standards on terminology databases.

5. Supporting the working process – from theory to practice

A further advantage of the proposed structur, is that it will also make it easier to develop tools that support the terminological working process as prescribed by theory and standards. In a termbase with the structure described in Fig. 9, the terminology of two (or more) languages can be analysed and registered separately, even by separate terminologists, and in a second stage equivalences can be found and registered without having to move content from one entry to another.

In Fig. 10, we show how a user interface for establishing equivalences relations between entries might be constructed. The terminologist user selects the relevant terminological ontologies, chooses concepts to be linked, and adds comments if appropriate. The entries created for each language remain separate entries, but the equivalence relation ensures that end users will be presented with the correct equivalents.

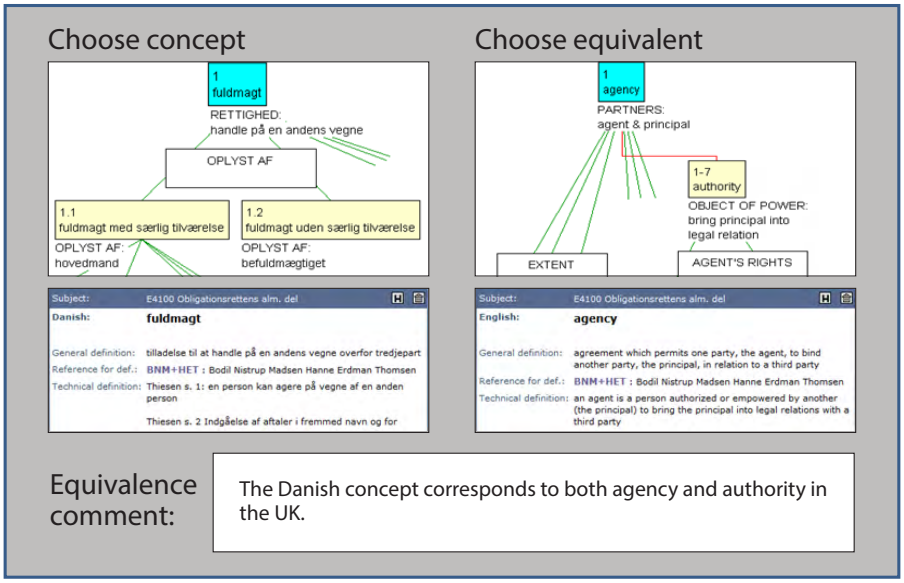

Figure 10: Mock-up user interface for establishing equivalence relations

6. Concluding remarks

Experience from practical multilingual terminology work has led us to propose a new termbase data structure for handling equivalence, a proposal that we are now putting forward again. The proposed structure, where equivalence relations are introduced between pairs of equivalent concepts, instead of the combination of equivalent concepts in the same entry, reflects more precisely what equivalence is, and makes it possible to register equivalence between one concept in one language and two or more concepts in another without having to compromise data base integrity through the introduction of doublettes. Moreover, it also enables terminologists to work in the prescribed manner, i.e. to register concepts and develop concepts systems for one language at a time and then, in a second step, find equivalence relations between concepts in the two languages. We therefore urge strongly that this change in structure is also introduced in standards for terminology databases and for the exchange of terminological data.

7. Acknowledgements

Work in the danTermBank project was supported by the VELUX foundation.

8. References

DANTERMcentret. *Knowledge Managamenet on the Internet with the i-Term Suite.* DANTERMcentret, accessed 01/27, 2015, http://www.iterm.dk/.

Hull, Anthony, Bodil Nistrup Madsen, & Hanne Erdman Thomsen (1998). DANTERMCBS for Everyone. In *TAMA '98 - Terminology in Advanced Microcomputer Applications. Proceedings of the 4th TermNet Symposium: Tools for Multilingual Communication*, 67-85. Wien: TermNet.

ISO 26162:2012. *ISO 26162: Systems to Manage Terminology, Knowledge and Content -- Design, Implementation and Maintenance of Terminology Management Systems.* International Standard., edited by TC 37/SC 3. Switzerland: International Organization for Standardization.

ISO 30042:2008. *ISO 30042: Systems to Manage Terminology, Knowledge and Content -- TermBase eXchange (TBX).* International Standard., edited by TC 37/SC 3. Switzerland: International Organization for Standardization.

ISO 704:2009. *ISO 704: Terminology Work -- Principles and Methods*, edited by ISO TC 37/SC 1/WG 1. Switzerland: International Organisation for Standardization, ISO.

ISO 860:2007. *ISO 860: Terminology Work - Harmonization of Concepts and Terms.* International Standard., edited by TC 37/SC 1/WG 2. Switzerland: International Organization for Standardization.

Lassen, Tine, Bodil Nistrup Madsen, Louise Pram Nielsen, & Hanne Erdman Thomsen (2013). The DanTermBank Project: Automatic Knowledge Extraction and Knowledge Structuring for a National Term Bank. Oslo 11-12 October 2012, Språkrådet, .

Madsen, Bodil Nistrup (1998). Typed Feature Structures for Terminology Work - Part I. In Lita Lundquist, Heribert Picht & Jacques Qvistgaard (eds.), *LSP - Identity and Interface - Research, Knowledge and Society. Proceedings of the 11th European Symposium on Language for Special Purposes*, 339-348: Copenhagen Business School.

Madsen, Bodil Nistrup & Hanne Erdman Thomsen (1998). The DANTERMCBS Database. In *LAMBDA*. Vol. 25, 143-179: Handelshøjskolen i København.

Madsen, Bodil Nistrup, Hanne Erdman Thomsen, Tine Lassen, Louise Pram Nielsen, Anna Elisabeth Odgaard, & Pia Lyngby Hoffmann (2013). Towards a Taxonomy of Terminological Data Categories. *eDITion - Fachzeitschrift Für Terminologie* 9 (1): 18-24.

Madsen, Bodil Nistrup, Hanne Erdman Thomsen, & Carl Vikner (2004). Principles of a System for Terminological Concept Modelling. In Maria Teresa Lino, Maria Fransisca Xavier, Fé1tima Ferreira, Rute Costa & Raquel Silva (eds.) *Proceedings of the Fourth International Conference on Language Resources and Evaluation*, 15-19: ELRA.

Madsen, Bodil Nistrup, Hanne Erdman Thomsen, & Annemette Wenzel (2006). I-Term for NORDTERM. In Rute Costa and others (eds.), *Terminology Design. Quality Criteria and Evaluation Methods (TermEval)*. ELRA: European Language Resources Association.

Thomsen, Hanne Erdman (1998). Typed Feature Structures for Terminology Work - Part II. In Lita Lundquist, Heribert Picht & Jacques Qvistgaard (eds.), *LSP - Identity and Interface - Research, Knowledge and Society. Proceedings of the 11th European Symposium on Language for Special Purposes*, 349-359: Copenhagen Business School.

Thomsen, Hanne Erdman (2016). Visualizing Culture-Specific Conceptualizations - a Tool and a Method for Concept Clarification and Intercultural Terminology Management. In *Terminologie & Kultur. DTT-Tagungsakte 2016*. Mannheim.