LSP Journal

Language for Special Purposes
Professional Communication
Knowledge Management
Cognition

Vol. 2 No. 2 (2011)

Table of contents:

Editorial

Zoom-in on… will be a recurring suite of feature articles on regional research centers active in the fields of LSP, professional communication, knowledge management and cognition. The objective is to show how state-of-the-art research and field work in businesses and organizations benefit from each other. The first zoom-in in this issue is on the backyard of this Journal, i.e. the Copenhagen Business School where the DANTERMcentre is hosted. Ongoing research and projects involving knowledge acquisition, validation and dissemination with a view to user-group oriented knowledge profiles are described, and some recent developments in practice-oriented ontology building projects with Danish authorities are featured. Given the interest these projects have raised, hope is that research centers and authorities in other countries could find some inspiration. We will gladly welcome proposals for centers volunteering to be the next ones to be zoomed in on.

Readers have inquired on several occasions whether it would be possible to include a list of relevant conferences in the LSP Journal. At present, it is our policy not to take on this demanding task. The LSP Journal is published regularly twice a year, and a list of conferences which would need constant maintenance and updating is not compatible with the publication of an electronic journal. However, if a report like the one presented in this issue by Sandra Campagna and Elizabeth Rowley-Jolivet in connection with a conference announcement, we will of course include a link to the event. For a portal to relevant conferences and courses, we would like to refer you to Helga Niska's well-maintained list of conferences within translation, interpreting, LSP, and terminology: http://www2.tolk.su.se/konfindx.html.

This issue also presents a call for papers for the journal. We encourage authors and readers to disseminate the call and at the same time to invite interested colleagues to subscribe to the journal.

Henrik Selsøe Sørensen
Editor in chief

http://uk.cbs.dk/staff/henrik_selsoe_soerensen

Suffixes in word-formation processes in scientific English

Begoña Montero-Fleta
Universidad Politécnica de Valencia
Valencia, Spain
bmontero@idm.upv.es

Abstract

Scholars have stated the particularities of the language used in specialized discourse but little attention has been so far paid to the role derivational morphology may play in register variation. The present research makes a contribution to the study of word-formation in scientific registers by means of a corpus-based approach to the productivity of 14 suffixes in two scientific English registers, i.e., computer science and medicine. In order to empirically examine the productivity of the suffixes in each register, types, tokens and hapaxes ratio were used. Results obtained were then contrasted with the presence of the same suffixes in the written language wordlist of the *British National Corpus* (BNC). The study shows that suffixes are a productive word-formation resource in scientific registers and that their productivity differs in the registers under study. Findings ranked higher productivity of abstract noun-forming suffixes such as *-ity, -ion* and *-ness* in scientific registers than in the BNC. The suffix *–ize* reached values in the scientific corpora highly over the ranking drawn from the BNC. On the contrary, the BNC yielded an outstanding productivity rate of *–free* and *-like,* suffixes which proved to be fully unproductive in the scientific registers under study.

1 Introduction

Scientific English can be approached from different angles taking into account its specific vocabulary and the framework of the syntactic structures most commonly used. Specific disciplines often develop their own patterns of discourse which do not always coincide with those of general English; thus, an issue of current interest for researchers in applied linguistics is determining how the frequency of use of particular discourse-level patterns may be characteristic of certain disciplines.

To name new concepts and refer to issues of their field of studies, specialized registers extend their vocabulary mainly by borrowing words from another language but also by using their own linguistic resources in compounding and derivation processes**.** However, word-formation and the factors which govern the acceptance of new formations into the language are generally taken very much for granted by the average speaker (Adams, 1973). The ways in which new words are formed have long been discussed in classical literature from a theoretical perspective (see e.g., Adams, 1973, Bauer, 1983 or Marchand 1969, to name a

few). Halliday (1978: 195) identifies seven strategies commonly used in the formation of specialized terminology:

- Reinterpretation of existing words
- Creation of new words from a native word stock
- Borrowing of words from foreign languages
- Calquing
- Invention of totally new words
- Creation of locutions
- Creation of new words from a non-native word stock (cf. Moskowitch, 2010).

Halliday's second strategy, creating new words from a native word stock, may imply different word-formation processes making use of compounding, conversion, back formation or derivation. Derivation is yielded by means of affixation, i.e., adding a derivational affix to a word by means of prefixation, suffixation or infixation. Prefixes are attached to the front of a base, suffixes to the end of a base and infixes are inserted inside a root. The present research will focus on a productive process of word-formation making use of derivation, i.e., the use of suffixes, which allows the creation of new words enlarging or changing the structure of the headwords, and will analyse its productivity in specialized registers.

The notion of register refers to the fact that the language we speak or write changes according to the type of situation. In this context, the concept of register comes under the larger concept of language variation in applied linguistics (Ghadessy, 1993). Some concepts have to be clarified to predict the linguistic characteristics of a situational context, following Halliday (1982): discourse field (institutional framework in which language is used, which includes the topic dealt with), discourse tenor (relationships within participants), and discourse mode (communication channel). Differences regarding the discourse field, discourse tenor and discourse mode will produce different varieties of language that can differ from other forms of the language systematically and coherently. Studies in register variation have found evident differences among specialized registers both from a morphological and syntactic point of view, and have stated that the clustering of such properties can even be used in defining a certain type of discourse (cf. Biber, 1995). The present research will focus on the use of suffixes in scientific English registers represented by the discourse field of medicine and computer science. Specialized manuals will be the discourse tenor in which the productivity of the word formation process will be studied.

Regarding the registers under study, Salager-Meyer & Alcaraz Ariza have carried out research on different medical genres (see, e.g., Salager-Meyer & Alcaraz Ariza, 2001, Alcaraz Ariza & Salager-Meyer, 2002, among others). A considerable body of literature has debated medical discourse (Taavitsainen, 2004), its lexicon (Norri 1998, 2004) or have analyzed it in studies on code-switching (Pahta, 2004). As for computer science registers, a corpus-based lexical analysis has been attempted by James (1994). Plag, Dalton-Puffer & Baayen (1999) claimed that little attention has been paid to the role derivational morphology may play in register variation, and studied the productivity of suffixes in different types of discourse in the British National Corpus. Apart from these studies, there is so far not much research on the productivity of suffixes in specialized genres, and, besides, many word counts performed are based on diachronically and dictionary-based studies. However, approaches in which the dictionary is used to make affix counts are not altogether satisfactory and, as Booij (2007: 65) points out, a dictionary is not useful for these studies because it only registers the words

which have become established words after some time, making a corpus a better source of information than a dictionary.

The justification for the present research lies, thus, in the existing need of research on the productivity of suffixes as a word-formation process in specialized registers. By productivity we mean morphological productivity, defined by Bolinger (1948: 18), as "the statistically determinable readiness with which an element enters into new combinations". The term 'productive' is used to describe a pattern, meaning that "when occasion demands, the pattern may be used as a model for new items" (Adams 1973: 197). Productivity is, thus, the capacity of a word element to produce new words (Plag, Dalton-Puffer & Baayen, 1999). As a language evolves, different suffixes may rise in productivity and later fall in productivity, or may be predominantly used in certain discourse fields; the present study will assess the rate of use of suffixes across registers.

2 Some preliminary remarks on productivity

As opposed to free morphemes, affixes are bound morphemes, i.e. morphemes that cannot be used on their own but must be attached to another word. There has been scholarly attention over the past decades to identify constraints on word- formation that cause one affix to be less productive than another, e.g. Anshen & Aronoff (1989) Aronoff (1976) or Plag, Dalton-Puffer & Baayen (1999). The linguistic factors which affect the productivity of word-formation rules were addressed in studies by Fabb (1988), Marle (1986) or Rainer (2005). Stein (1977) offered contributions on affix ordering. The influence of suffixation on stress was studied by Bauer (1983). Hill (1974) and Levi (1973) addressed the problems of distinguishing suffixes and the specifications of their meanings. Quantitative measures of productivity were discussed by Aronoff (1976), Baayen (1991), Baayen & Lieber (1991), Baayen & Renouf (1996), Bauer (2001), Booij (2000), Nishimoto (2004), Plag (1996, 2003), Plag, Dalton-Puffer & Baayen (1999) or Marle (1986). From the different productivity measures proposed in these studies, the present study follows Baayen's (1991) productivity measures.

Often confused terms in productivity measurements such as type, token and hapax, merit a clarification before presenting the methodology used in the study:

- Type frequency is the number of different words, i.e., word types with a given suffix.

- Token frequency is the total frequency of use of all the words of that particular type. For example, if 500 different nouns in –ity were found in a corpus, the type frequency of this noun-forming suffix would be 500. The token frequency could be much higher than the type, as it counts the total number of times of its occurrence, which includes all repetitions of the same item.

- Hapax legomena or hapaxes are words that occur only once in a large text corpus. Their role is significant in the determination of productivity.

3 Purpose

English has a remarkably small inventory of affixes in comparison with languages such as Spanish. The aim of the present study was to assess and compare the productivity of 14 English suffixes in two different scientific registers, medicine and computer science in comparison with their productivity in the *British National Corpus* (BNC). The suffixes analyzed were those studied by Plag, Dalton-Puffer & Baayen (1999) in their contrastive study of the suffixes in the written and spoken wordlists of the BNC. Insofar as

nominalization is a common word-formation resource, the main criterion for choosing the suffixes for the study was the need to complement Biber's (1995) research on register variation through nominalization by the use of appropriate suffixes (e.g., *-ness -ity*, *-ance/-ence* or *–dom*) with other derivational patterns performing different morphosyntactic and morphosemantic functions. The suffixes analyzed can be grouped into different types according to the lexical items they generate:

- Suffixes forming abstract nouns: *-ity, -ness, -ion;*
- Suffixes forming participant nouns: *- er, -ist;*
- Suffixes forming measure partitive nouns: *-ful;*
- Suffixes forming derived verbs: *-ize;* and
- Suffixes forming derived adjectives: *-able, -free, -ful, -ish, -less, -like, -wise.*

4 Hypothesis

Our research will attempt to answer the following research questions:

1. Are there more productive suffixes in medicine registers than in computer science registers?

2. Is there a difference in the behaviour of these suffixes in the specialized registers of medicine and computer science compared to a wider, more general corpus, i.e., the *British National Corpus* (BNC)?

The analysis of the suffixes used will show their productivity in the registers under study. The hypothesis is that since derivational suffixes are very frequently used in word-formation processes in formal settings, their presence in the specialized discourse of medicine and computer science will be higher than in the BNC. This paper will thus observe to what extent scientific registers have their own word-formation processes reflecting specificity of the field, and whether differences in use can be observed from a more general corpus.

5 Methodology

5.1 The corpus

The study was based on the contrastive analysis of three different corpora compiled from the following sources:

- Miller, R.D. (ed.) (2000): *Anesthesia.* 5th ed. Philadelphia: Churchill Livingstone, Inc., chosen as representative of the language of medicine totalling 1.169.749 words. This specialized manual deals with anesthesiology, the branch of medicine that is concerned with the study and practice of anesthesia which requires competency in general medicine, a broad understanding of surgical procedures, and a comprehensive knowledge of clinical obstetrics, chest medicine, neurology, pediatrics, pharmacology, biochemistry, cardiology, and cardiac and respiratory physiology.

  The corpus obtained from Miller (2000) will, due to space constraints, sometimes be referred to as MC (medicine corpus) throughout the paper.

  Grama, A. et al. (2003): *Introduction to Parallel Computing*, 2nd ed. Boston: Addison Wesley, taken as a representative language sample of computer science, with 358.509 words. The manual presents new developments in parallel-computing and discusses

topics such as parallel architectures, designing and analyzing parallel algorithms, and programming techniques. Emerging areas such as computational biology and nanotechnology have implications for algorithms and systems development, while changes in architectures, programming models and applications have implications on how parallel platforms are made available to users in the form of grid-based services.

- In this study, a preliminary estimation of the productivity of the corpus suggested some changes on the size of the samples, as both scientific corpora chosen differed in length. As productivity may be dependent on the size of the corpus (Baayen, 1993), it was thought more convenient to compare equal-sized scientific samples. In the process of corpus compilation, Grama et al.'s (2003) was complemented with another source so as to equal the number of words of running text provided by the MC. A number of chapters from Tucker, A.B. (ed) (2004): *Computer Science Handbook* (2nd. ed). Brunswick Maine: Bawdain College, were randomly added to adjust the size requirements. Tucker (2004) offers a rich collection of theory and practice that fully characterizes the current state of the field of computer science. The number of words of running text of the new corpus was now estimated to be sufficient to obtain preliminary results of the behaviour of suffixes. The corpus obtained from Grama et al. (2003) and Tucker (2004) will be henceforth abbreviated as CSC (computer science corpus).

The analyses of these two corpora of scientific registers were compared with the productivity of the same suffixes in a more general sample, *The British National Corpus* (World Edition) (2000): Oxford: Oxford University Computing Services. The British National Corpus (BNC) contains 100 million word tokens from samples of written and spoken language drawn from a wide range of sources, designed to represent a wide cross-section of British English from the latter part of the 20th century. The written part of the BNC (90%) includes, for example, extracts from regional and national newspapers, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published and unpublished letters and memoranda, school and university essays, among many other kinds of text. The wordlist that corresponds to written sources of this corpus was the basis for the contrastive study of the productivity of suffixes with the scientific corpus carried out in the present study.

5.2 Procedure
The register analysis undertaken has benefited from advances in computational linguistics. In the present research, after converting the PDF files containing Miller's (2000) and Grama et al.'s (2003) and Tucker (2004) corpora into plain text, *WordSmith Tools* (Scott, 1996) generated wordlists based on text analysis, and was instructed to find out the words containing the suffixes required. But in spite of the advantages of this software for extracting the suffixes from the corpus, the wordlists obtained were not totally reliable. The result of the process produced raw frequency data, that is, a long list of words with suffixes which had to be cleaned.

The main method used in this study for measuring productivity rates was the estimation of Baayen's (1992) productivity index. Baayen measures productivity rates (P) by calculating the ratio of hapax legomena to tokens for a given affix by using the following formula:

$$P = n_1/N$$

In Baayen's formula, n_1 is the number of hapax legomena with a given affix, and N is the number of tokens with the same affix.

6 Results and discussion

As mentioned above, preliminary results of the electronic processing of the corpus required a manual refinement of the wordlist obtained. This process discarded a large number of word counts which did not contain the expected suffixes, as many suffixes listed were, instead, part of the lexeme or even the proper lexeme: e.g. *cable*, *table* and the adjective *able* were included by *WordSmith* as items containing the suffix *–able*. The corpus was then manually analysed so as to eliminate entries which shared the string of letters of a suffix but were not the required morpheme. To clean the wordlist, one-syllable words as well as proper nouns were eliminated. Context as well as the *Oxford English Dictionary* were used to find out if the word under study contained the required suffix, a fact which was easier to tell with some suffixes than with others, with *-er* and *–ist* being the most problematic. As an example, from a total of 2258 types recorded in respect of *-er* in the CSC, only 179 were admitted as valid types. In contast, fewer inconsistencies were observed in other *WordSmith* suffix counts, such as, e.g., *-ion, -ity* or *–ness*. Table 1 exhibits the total types found for each suffix in the MC corpora, so as to show the reader the rate of suffixes rejected:

Medicine corpus (MC)		
Suffix	Total types	Valid types
-able	235	228
-er	2258	179
-free	2	0
-ful	36	36
-ion	1096	1058
-ish	49	28
-ist	109	47
-ity	312	309
-ize	78	76
-less	44	41
-like	5	2
-ness	114	113
-wise	10	8

Table 1. From raw frequency data to clean data. Valid types in the medicine corpus (MC).

Even the task of identifying words with capital letters as proper nouns was not straightforward, as wordlists compiled by means of this software used all initial capital letters. *WordSmith Tools* isolated the items containing suffixes in the context in which they were used, a task which helped to identify the words, e.g., *Bender, Berger* or *Benzer* as proper nouns, as headwords to which the *–er* suffix had been added or as neologisms. It was frequently problematic to decide if some words should be included or excluded as examples of suffixation.

The preparation of the corpus was followed by the estimation of tokens and hapaxes. A two stage analysis was then accomplished. In the first stage, the productivity of both corpora was

calculated by dividing the hapaxes by the number of tokens. Table 2 illustrates the rates obtained:

Suffix	Productivity (P) MC	Productivity (P) CSC
-able	0,01218658	0,009449466
-er	0,01649359	0,011346445
-free	0	1*
-ful	0,00681115	0,011235955
-ion	0,00341802	0,005245546
-ish	0,02325581	0,058823529
-ist	0,01481481	0
-ity	0,00762443	0,020016681
-ize	0,04004711	0,046728972
-less	0,03370787	1*
-like	1*	0
-ness	0,02889825	0,120000000
-wise	0,01063830	0,012658228

Table 2. Productivity of suffixes in medicine corpus (MC) and computer science corpus (CSC)
*An asterisk in the grid indicates that the presence of the corresponding suffix in the corpus reaches such low occurrences that productivity measuring was discarded.

From the data obtained the following can be drawn when comparing both corpora:

- The high values of productivity in the case of *-free* or *-like* in both the MC and the CSC can be biased by the small number of tokens (*-free* accounts for 0 tokens in the MC and 2 in the CSC, and *–like* records 2 tokens in MC and 0 in the CSC). The productivity values obtained are misleading and thus cannot be accepted as true indicators of productivity in medicine and computer science registers.

- The verb forming suffix *-ize* generates high values of productivity in both registers.

- The suffixes *-ity,* and *-ion* achieve high values of productivity in the MC and the CSC. These results are not surprising as these suffixes are often used in scientific texts to encode field or domain specific concepts (Plag, Dalton-Puffer & Baayen, 1999). The suffix *-ful* is also productive in both registers as an adjective-forming suffix.

- The suffixes *–ist* and *-less* are fully productive in the MC but unproductive in the CSC.

From the above mentioned data and in answer to our first research question, if there were suffixes more frequently used in medicine registers than in computer science registers, this study provides preliminary results that confirm the similarity of both registers regarding the productivity of the suffixes *–ize* forming derived verbs, and *-ity* and *–ion* forming abstract nouns. Suffixes forming adjectives from verbs, *-able* and *-ful* were also moderately productive in both registers. However, no occurrences were encountered of *-ful* forming partitive nouns in the analysis. The suffixes *-free* and *-like* do not account for word-formation processes in the scientific registers under study. As for the differences observed, some suffixes have to be

highlighted as their productivity stands out in the MC corpus and not in the CSC, i.e, *-ist*, a suffix which forms nouns from adjectives, and *-less*, a suffix which forms adjectives from nouns and also from verbs.

A second stage of the process implied the comparison between the productivity values obtained from the analysis of the scientific registers with the productivity rates of the suffixes in the British National Corpus. Table 3 exhibits the different productivity data of the suffixes across the corpora investigated.

Productivity index (P)		
Suffix	**MC**	**CSC**	**BNC**
_able	0,012186581	0,009449466	0,002211524
_er	0,016493586	0,011346445	0,019466634
_free	*	*	0,103613409
_ful	0,006811146	0,011235955	0,001025885
_ion	0,003418018	0,005245546	0,000382729
_ish	0,023255814	0,058823529	0,033828276
_ist	0,014814815	*	0,003582162
_ity	0,007624432	0,020016681	0,000917291
_ize	0,040047114	0,046728972	0,002109537
_less	0,033707865	*	0,009597742
_like	*	*	0,110412371
_ness	0,028898254	0,12	0,008816627
_wise	0,010638298	0,012658228	0,06121473

Table 3. Productivity index: medicine corpus (MC), computer science corpus (CSC) and British National Corpus (BNC).

Given the second research question motivating this study, if there were a difference in the productivity of suffixes in the specialized registers of medicine and computer science from the productivity in a more general corpus, the BNC, the following can be inferred:

- The BNC shows an outstanding productivity rate of the suffixes forming derived adjectives *-free, -like.* This productivity contrasts with the extremely low use of these suffixes in the scientific registers under study. From the other suffixes generating adjectives in the study, *-ish* offers similar values in the specialized registers to the ones of the BNC, although the CSC yields slightly higher counts; and *-able,* a suffix of Romance origin, derived from transitive verbs, recorded more counts in the MC and the CSC than in the BNC.

- The agentive deverbal *-er* forms participant nouns and accounts for very similar productivity in both the MC and the CSC registers and the data obtained from the BNC, although slightly lower in the latter. The overall results for the suffix *-er* seem to be plausible, since it has similar values in all the corpora attached to virtually any semantically appropriate verb.

- The suffixes *-ist,* forming nouns from verbs*,* and *–less,* added to nouns to form adjectives, record a higher productivity in the MC than in the BNC.

11

- What is outstanding is the behaviour of *-ity* and *-ion,* abstract noun-forming suffixes mainly attached to words of classical origin. These suffixes record much higher productivity values in the MC and the CSC than in the BNC. This high frequency of use is also shown in the CSC by *-ness,* a suffix added to adjectives to form abstract nouns.

- The suffix *–ize,* forming verbs from nouns and adjectives frequently of neoclassical origin, is fully productive in the MC and the CSC: values are higher than in the BNC.

Figure 1 displays the differences observed in the behaviour of the suffixes in the different corpora more visually:
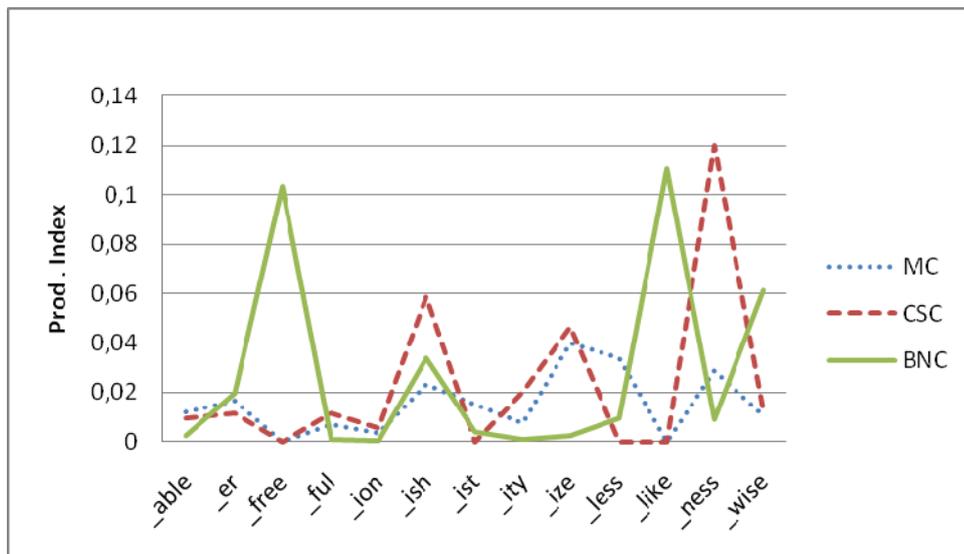

Figure 1. Productivity index: medicine corpus (MC), computer science corpus (CSC) and British National Corpus (BNC).

6 Conclusion

The quantitative description of the productivity of suffixes can be interesting in the study of specialized discourse to reveal their morphological differences, and as some scholars have pointed out, the productivity ranking obtained with a corpus-based productivity measure will be useful in many forms of linguistic research, not necessarily limited to the study of word-formation.

On the theoretical level, the foregoing study makes a contribution to the study of word-formation in scientific registers and describes differences among different fields of study, a fact which turns out to be even more outstanding when findings are compared with the BNC. The hypothesis on the higher prevalence of suffixes in scientific registers has been partly confirmed. Our results confirm that there is a difference in the behaviour of some suffixes across scientific registers, and confirm that morphological productivity is subject to register variation. A higher productivity of certain suffixes in scientific registers than in the BNC has been revealed. However, scientific registers have also recorded a total absence of some suffixes which were fully productive in the BNC.

Although some preliminary conclusions have been drawn, these are not definite. Further studies are required to confirm these findings across larger corpora. The intention here was to

consider patterns in the behaviour of the suffixes rather than to determine definite measures for individual affixes. Future studies on morphological derivation by the author will also strive to assess the characteristics of individual suffixes with higher productivity in scientific registers from the point of view of their morphological and syntactic level and will also consider the semantic value of the derived form.

7 References

Adams, V. (1973): An introduction to modern English word-formation. London: Longman.

Alcaraz Ariza, M.A. & Salager-Meyer, F. (2002): Género y crítica en la prosa médica escrita en español: Función comunicativa y relación de poder. *Hermes, Journal of Linguistics*, 29: 163-186.

Anshen, F. & Aronoff, M. (1989): Morphological productivity, word frequency and the OED. In R. Fasold & D. Schriffin (eds). *Language Change and Variation* (pp.197-202). Amsterdam/Philadelphia: John Benjamins.

Aronoff, M. (1976): *Word formation in generative grammar*. Cambridge, Mass: MIT Press.

Aronoff, M., & Fudeman, K. (2005): *What is morphology?* Oxford: Blackwell Publishing.

Baayen, H.R. (1991): Quantitative aspects of morphological productivity. In G. Booij & J. van Marle (eds.). *Yearbook of Morphology* ,1991 (pp. 109-49). Kluwer Academic Publishers, Dordrecht, 181-208.

Baayen, H.R. (1993). On frequency, transparency and productivity. In G. Booij & J. van Marle (eds.). *Yearbook of Morphology,* 1992 (181-208). Dordrecht: Kluwer Academic Publishers.

Baayen, H.R. & Lieber, R. (1991): Productivity and English word-formation: A corpus-based study. *Linguistics*, 29: 801-43.

Baayen, H.R & Renouf, A. (1996): Chronicling The Times: Productive lexical innovations in an English newspaper. *Language,* 72(1): 69-96.

Bauer, L. (1983): *English word- formation*. Cambridge: Cambridge University Press.

Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. New York: Cambridge University Press.

Bolinger, D. L. (1948): On defining the morpheme. *Word* 4: 18–23

British National Corpus (World Edition). CD-ROM. (2000). Oxford University Computing Services.

British National Corpus. http://www.natcorp.ox.ac.uk/corpus/index.xml.

Booij, G.E. (2000): The phonology-morphology interface. In L. Cheng & R. Sybesma (eds). *The first international state of the article book*. Berlin: Mouton de Gruyter, 287-306.

Booij, G.E. (2007): *The grammar of words. An introduction to morphology*. Oxford: Oxford University Press.

Fabb, N. (1988): English suffixation is constrained only by selectional restrictions. *Natural Language and Linguistic Theory,* 6: 527-539.

Grama, A., Grupta, A., Karypis, G. & Kumar, V. (2003): *Introduction to parallel computing*, 2nd ed. Boston: Addison Wesley.

Ghadessy, M. (ed.) (1993): *Register analysis: Theory and practice.* London: Pinter Publishers.

Halliday, M. A. K. (ed.) (1978): *Language as social semiotic*. London: Edward Arnold.

Hill, A.A. (1974): Word stress and the suffix –ic. *Journal of English Linguistics,* 8: 6-20.

James. G. (1994): *English in computer science: a corpus-based lexical analysis*. Hong Kong: Longman.

Levi, H.J.N. (1973): Where do all those adjectives come from? *Papers from the Regional Meetings of the Chicago Linguistic Society,* 9: 332-345.

Marchand, H. (1969): *The categories and types of present day English word-formation: a synchronic-diachronic approach.* Munich: Beck.

Marle, J. van (1986): The domain hypothesis: The study of rival morphological processes. *Linguistics,* 24: 601-627.

Miller, R.D. (ed.) (2000): *Anesthesia.* 5th ed. Philadelphia: Churchill Livingstone, Inc.

Moskowitch, I. (2010): Morphologically complex nouns in English scientific texts after Empiricism. *Linguistik Online* 43(3). http://www.linguistik-online.de/43_10/moskowich.html. Last visited September 2011.

Nishimoto, E. (2004): Defining new words in corpus data: Productivity of English suffixes in the British National Corpus. http://www.cogsci.northwestern.edu/cogsci2004/papers/paper505.pdf. Last visited June 2011.

Norri, J. (1998): Names of body parts in English, 1400–1550. *Annales Academiae Scientiarum Fennicae, Humaniora* 291. Helsinki.

Norri, J. (2004): Entrances and exits in English medical vocabulary, 1400–1550" In I. Taavitsainen & P. Pahta (eds.) *Medical and Scientific Writing in Late Medieval English* (pp. 100–143). Cambridge: Cambridge University Press.

Pahta, P. (2004): Code-switching in medieval medical writing. In I. Taavitsainen & P. Pahta (eds). *Medical and Scientific Writing in Late Medieval English* (pp. 73–99). Cambridge: Cambridge University Press.

Plag, I. (1996): Selectional restrictions in English suffixation revisited. *Linguistics,* 34: 769-798.

Plag, I. (1999): *Morphological productivity: Structural constraints in English derivation.* Berlin: Mouton de Gruyter.

Plag, I. (2003): *Word - formation in English.* Cambridge: Cambridge University Press.

Plag, I., Dalton-Puffer, C. & Baayen. H. (1999): Morphological productivity across speech and writing. *English Language and Linguistics,* 33: 209-228.

Rainer, F. (2005): *Constraints on productivity.* In P. Stekauer & R. Lieber (eds.), *Handbook of Word-Formation* (pp. 335-52). Dordrecht: Springer.

Salager-Meyer, F. & Alcaraz Ariza, M.A. (2001): Lo cortés no quita lo valiente: la retórica de la discrepancia en el discurso médico escrito en español (1880-1899). In J.C. Palmer, S. Posteguillo & I. Fortanet (eds.). *Discourse analysis and terminology in Languages for Specific Purposes* (pp. 15–24). Collecció Estudis Filològics 5. Castellón: Publicaciones de la Universitat Jaume I, 15-24.

Scott, M. (1996): *WordSmith tools.* Oxford University Press: Oxford.

Stein, G. (1977): The place of word-formation in linguistic description. In H.E. Brekle & D. Kastovsky (eds.). *Perspektiven der Wortbildungsforschung. Beiträge zum Wuppertaler Wortbildungskolloquium vom 9.-10. Juli 1976.* Wuppertaler Schriftenreihe Linguistik 1. (pp. 219-235). Bonn: Bouvier Verlag Herbert Grundmann.

Taavitsainen, I. & Pahta, P. (eds.) (2004): *Medical and scientific writing in late medieval English.* Cambridge: Cambridge University Press.

Tucker, A.B. (ed) (2004): *Computer Science Handbook* (2nd ed). Brunswick Maine: Bawdain College.

***

Zoom-in on

DANTERMcentre

&

The Research Group Terminology, Lexicography, and Professional Communication
Copenhagen Business School
Denmark

Bodil Nistrup Madsen, Hanne Erdman Thomsen & Pia Hoffmann
Copenhagen Business School
Copenhagen, Denmark

The DANTERMcentre is the centre for terminology at Copenhagen Business School (CBS) and one of the country's leading consultancy and development centres in the fields of language technology, terminology and concept organization.

The Centre is part of The Department of International Language Studies and Computational Linguistics at CBS and the employees all have a background in IT, terminology or computational linguistics.

It is the aim of the DANTERMcentre to:

- disseminate knowledge of language technology and terminology
- develop language technology tools
- contribute to the promotion of concept organization in the construction of IT systems
- assist public authorities and others in achieving concept clarification
- create a wide national and international network within terminology and language technology

The language technology products developed by the DANTERMcentre include bespoke term and knowledge bases such as i-Term, a tool consisting of a concept-oriented term base, a graphical concept-modeling tool and an optional search tool.

The primary areas of the **Research Group Terminology, Lexicography, and Professional Communication** are ontology, terminology, lexicography and LSP communication. The research of the group includes development of new theories on concepts and conceptual analysis in connection with the improvement of LSP communication. It is the aim of the group to contribute actively to the development of value-adding terminology and knowledge databases with a view to supporting business enterprises and organisations in their internal and external communication. In addition, the group applies theories on naming and concept construction to innovative business and product development.

The research and development project CAOS – Computer-Aided Ontology Structuring – was carried out in the period from 1998-2006. The project received funding from the Danish Research Council for the Humanities from 1998-2001. The aim of this project was to develop a computer system designed to enable semi-automatic construction of concept systems, or ontologies.

The research carried out in the DanTermBank project is a prerequisite for establishing a national Danish terminology and knowledge bank which can ensure development and quality of Danish LSP. The aim of the project is to develop methods for automatic knowledge extraction, automatic construction and updating of ontologies. In the project, methods will be developed for automatic merging of terminological data from various existing sources, as well as methods for target group oriented knowledge dissemination. The project is supported by VELUX FONDEN, http://veluxfonden.dk/, in the period from 2011 to 2014.

Read more about some of the ongoing projects of the DANTERMcentre and the research group in the following three papers and on the centre website www.danterm.dk.

Zoom-in on

Why knowledge modeling
is important for business and
for a Danish terminology and knowledge bank

Bodil Nistrup Madsen
DANTERMcentre
Copenhagen Business School, Denmark
bnm.danterm@cbs.dk

Abstract

Businesses and organizations, including public authorities, have a growing need for organizing and handling large amounts of data. In order to manage complex knowledge, knowledge must be modeled and structured. One very powerful method used for structuring knowledge is the use of ontologies. Businesses and organizations need efficient tools for building domain-specific ontologies and systems for managing knowledge. Most medium-sized and large Danish businesses and organizations operate in a bilingual or multilingual environment, where knowledge is transferred and stored in Danish and/or English, and possibly other languages, so information must be freely retrievable and communicable in several languages. In this paper I will give examples of why knowledge modeling is important for businesses, and why knowledge modeling is a central part of the DanTermBank project, the aim of which is to lay the foundation for a national terminology and knowledge bank. Furthermore, I will briefly introduce our plans for teaching within the field of knowledge modeling at Copenhagen Business School, CBS.

Introduction

In order to manage vast quantities of digitized information, businesses and public authorities need to structure and manage their knowledge so that the right persons have access to the right information at the right time. This is posing an ever-increasing challenge in an age where the flow and complexity of information are growing rapidly, and where the need for quick, efficient and reliable communication is essential. The challenge can only be met if we are capable of structuring knowledge appropriately and if we can navigate freely in this knowledge.

In this paper, I will first give examples of activities in businesses and organizations where knowledge modeling is needed and examples of research which may contribute to the fulfillment of these needs. Then, I will introduce the concept of terminological ontologies and give examples of ontology work already carried out in the Danish authorities as well as examples of relevant education initiatives.

Finally, I will give an introduction to the plans and ongoing work in the DanTermBank project, which runs from 2011-2014 at the Department of International Language Studies and Computational Linguistics, ISV, CBS. The project is supported by the VELUX Foundation, www.veluxfonden.dk.

If a terminology and knowledge bank does not contain a sufficient number of terms, users will not feel encouraged to use it, and on the other hand, users will be frustrated if it contains a large amount of terms with only little or poor quality information. Therefore, it is necessary to use automatic procedures in order to extract and systematize information about terms. In this project, we will develop methods for automatic knowledge extraction and automatic construction and updating of ontologies. The project also aims at developing methods for automatic merging of terminological data from various existing sources as well as methods for target group-oriented knowledge dissemination. Terminological ontologies will play an important role in all the subprojects (see below).

Business activities

Figure 1 presents keywords related to activities of businesses and public authorities.
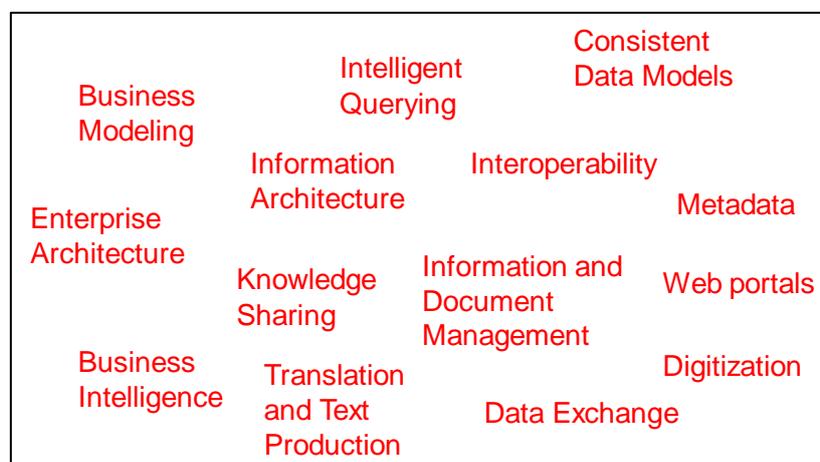

Figure 1. Examples of issues that are relevant for businesses and public authorities

- Descriptions of core aspects of a business: business modeling (Enterprise Architecture and Business Intelligence)

- IT-related issues: information architecture, consistent data models, interoperability of knowledge resources and data exchange

- Knowledge dissemination: digitization (web portals, metadata)

- Activities which support business processes: intelligent querying, knowledge sharing, information and document management, translation and text production.

As a basis for the successful conduct of central business activities, there is a need for concept clarification and knowledge modeling.

Humanistic research

Figure 2 presents examples related to humanistic research which can support business activities.
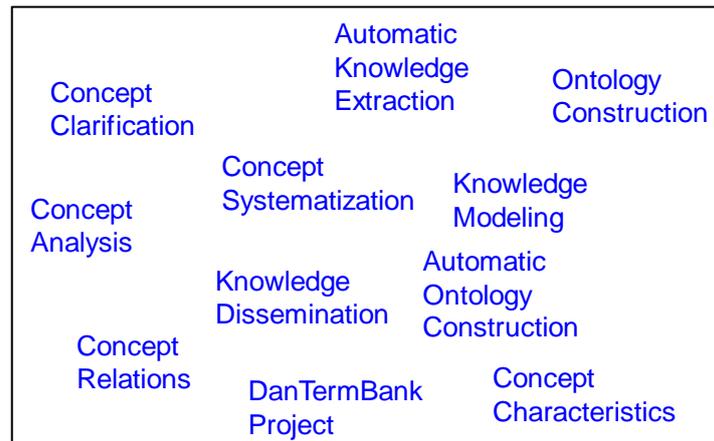
Figure 2. Examples of humanistic research which can support business activities

- Core research issues: concept clarification, concept analysis (concept relations and concept characteristics)

- Methods to organize knowledge: knowledge modeling, concept systematization, ontology construction

- Research issues of the DanTermBank project: automatic knowledge extraction, automatic ontology construction, knowledge dissemination.

What is an ontology?

Figure 3 presents a simplified example of a terminological ontology for taxes.

Figure 3. Example of a terminological ontology for Danish tax collection

The ontology in this figure has been created using the ontology modeling module i-Model of the terminology and knowledge management system i-Term ®, developed by the DANTERM Centre at CBS, www.i-Term.dk.

As an introduction, the elements of a terminological ontology will be briefly explained.
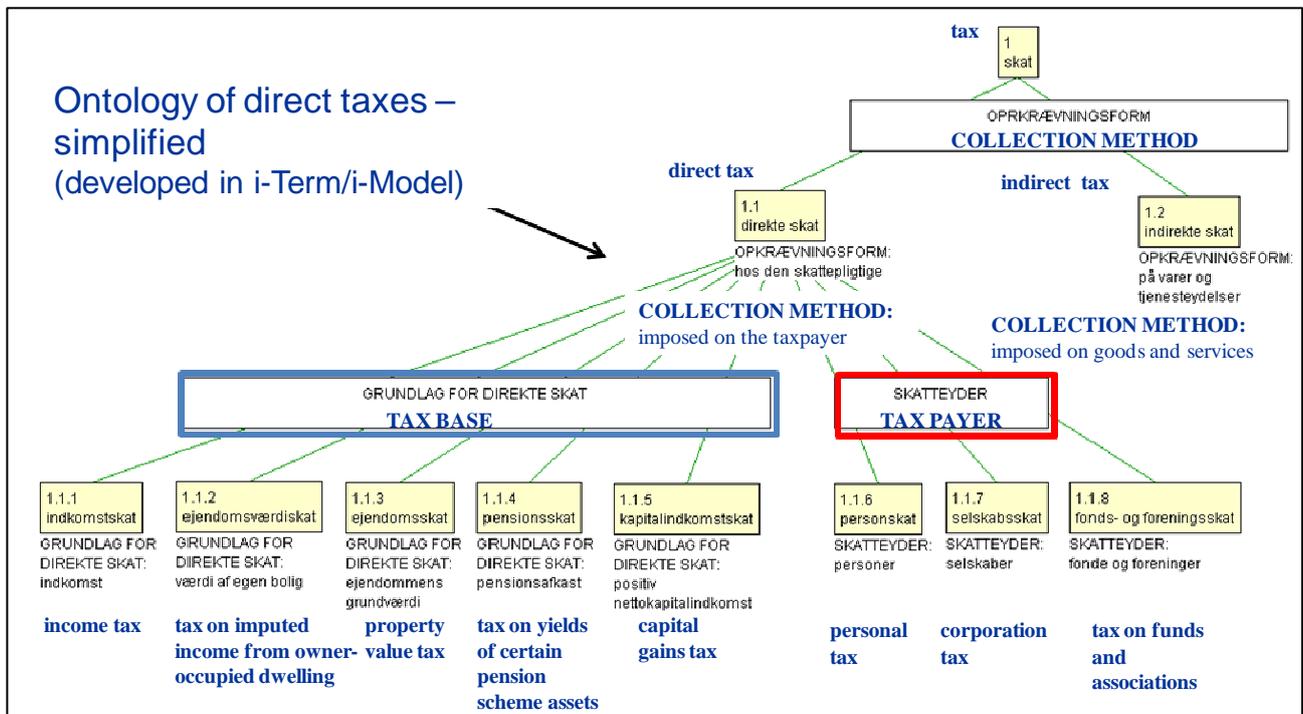
The ontology reflects the Danish taxation system. English glosses are added, but it should be emphasized that if we develop an ontology to reflect the English taxation system, it will differ from the Danish one. It is our intention to develop both a Danish and an English ontology In the DanTermBank project, and these will form the basis for clarification of equivalence (or non-equivalence) between the two languages, that will be reflected in the entries of the terminology and knowledge bank.

Concepts are represented by means of yellow boxes and type relations by means of green lines. The concepts *direkte skat* (direct tax) and *indirekte skat* (indirect tax) differ with respect to OPKRÆVNINGSFORM (COLLECTION METHOD). This is what we call a subdivision criterion. Subdivision criteria are shown in white boxes. Subdivision criteria group concepts according to the characteristics which are given in the form of feature specifications below the concepts – see *direkte skat* (direct tax) and *indirekte skat* (indirect tax): *OPKRÆVNINGSFORM: hos den skattepligtige* and *OPKRÆVNINGSFORM: på varer og tjenesteydelser* (COLLECTION METHOD: imposed on the taxpayer and COLLECTION METHOD: imposed on goods and services).

The subdivision criteria help the user to understand the meaning of the concepts, give a good overview and help the terminologist in writing consistent definitions. The definition of a concept is given by means of the position in the ontology and the characteristics. Definitions may be created on the basis of this information.

In Figure 4, the ontology from Figure 3 is compared with an extract of a classification system from a catalogue of public services, the class *Taxes on income and wealth* at the bottom of Figure 4.

The classification system is not based on the ontology. Some types of taxes are missing and the order of taxes in the classification system is not intuitive compared to the ontology: the classes corresponding to the sub-concepts in the ontology are not grouped according to the two aspects (subdivision criteria): *GRUNDLAG FOR DIREKTE SKAT* (TAX BASE) and *SKATTEYDER* (TAX PAYER). It is not clear whether *property taxation* covers both *ejendomsværdiskat* (tax on imputed income from owner-occupied dwelling) and *ejendomsskat* (property value tax), and *indkomstskat* (income tax) is missing.

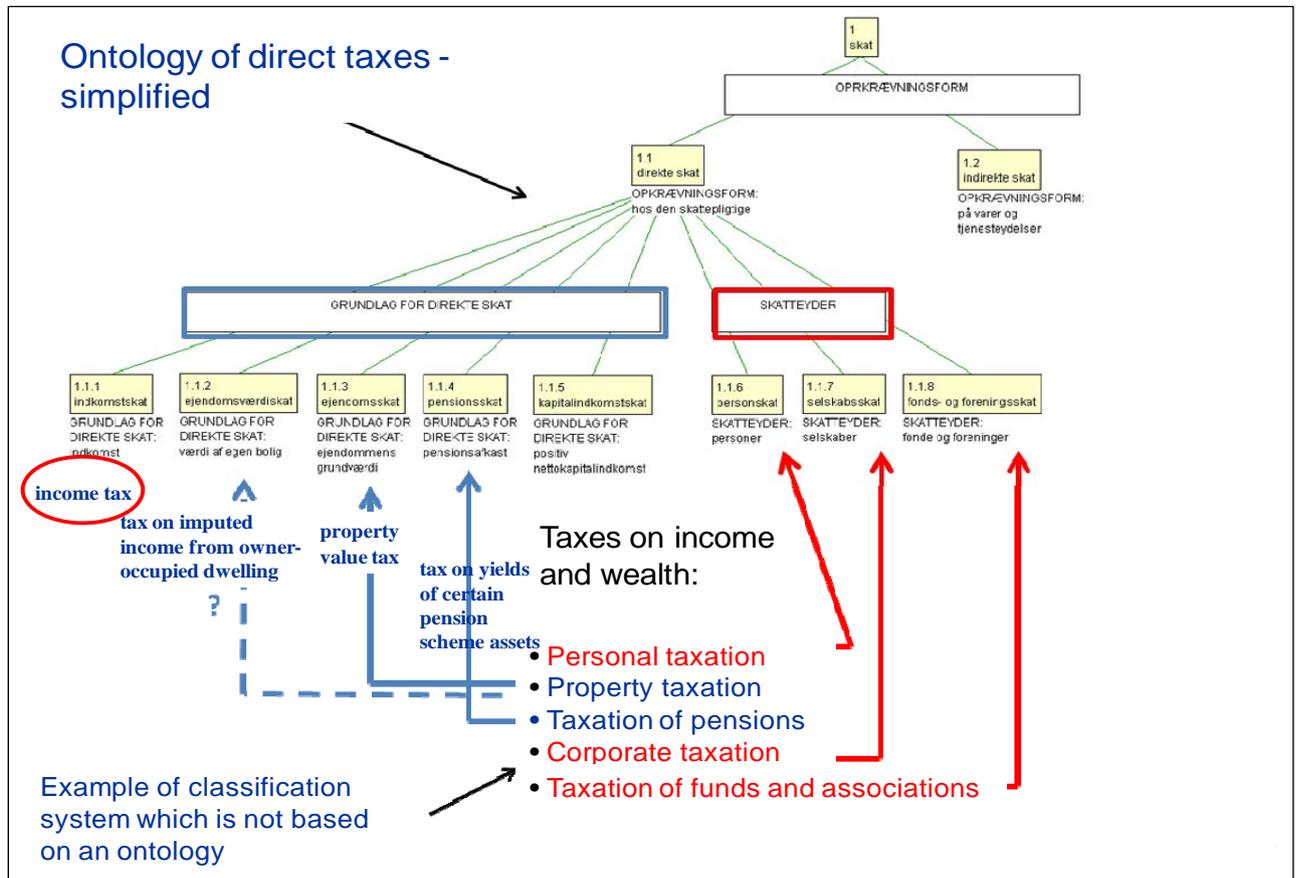Figure 5 shows how the ontology of taxes may grow if all types of tax are added, Nielsen (2009).

Zoom-in on

Figure 4. Comparison of the terminological ontology with an example of a classification system
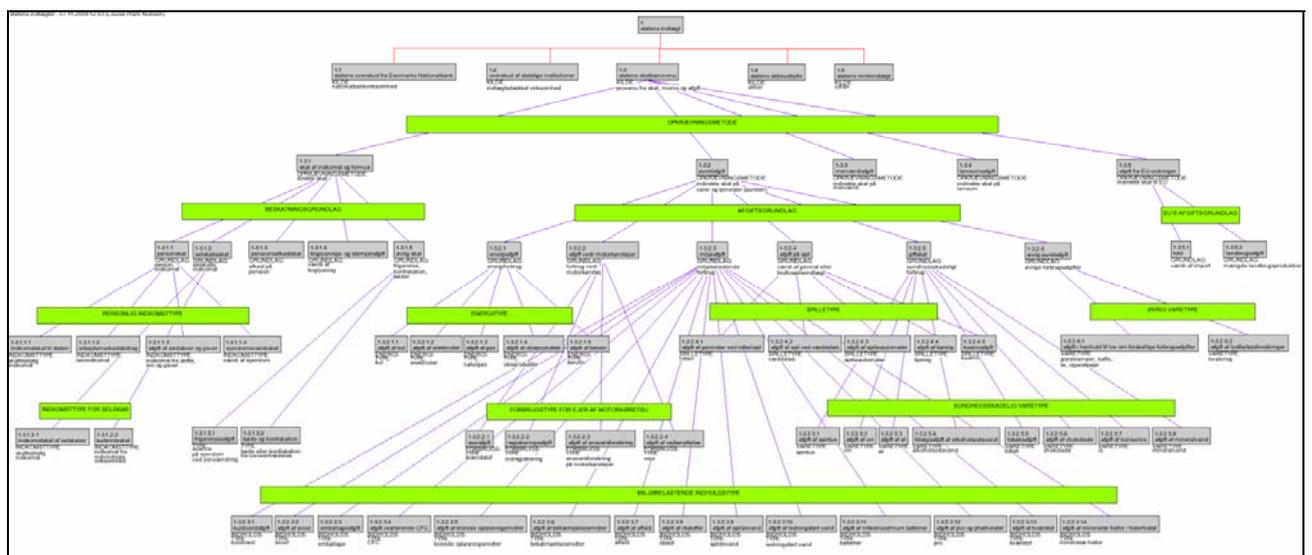
Figure 5. Example 'Statens skatteprovenu'

In the next section, I will go a little more into detail on the differences between ontologies and other knowledge structuring tools.

Ontologies vs. data models and classification systems

Figure 6 presents a small ontology illustrating the differences between ontologies, data models and classification systems (including taxonomies).
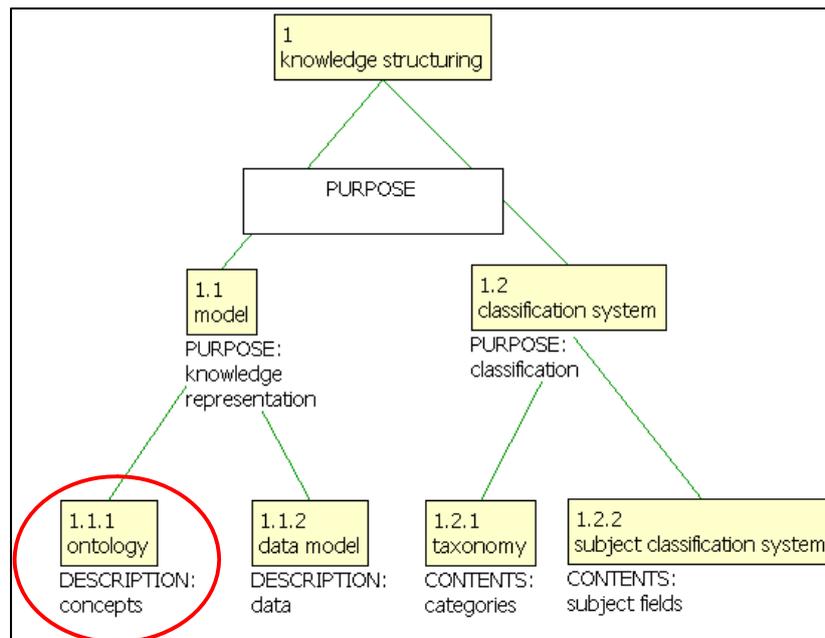
Figure 6. Ontology of central knowledge structuring concepts

On the basis of the characteristics, we suggest the following definitions of these concepts:

model:
simplified representation of knowledge about phenomena

ontology:
model for the description of knowledge about concepts

data model:
formal model for the description of types of data in an IT system

classification system:
system for the division of phenomena into classes

taxonomy:
classification system for the division of categories of a domain

Differences between ontologies and classification systems are also described in Madsen & Thomsen (2009). Madsen & Odgaard (2010) describe differences between ontologies and data models.

Examples of the use of ontologies

Here I will give some examples of the use of ontologies in businesses and organizations, cf. also the examples in Figure 1.

- Development of IT Architecture
- Ontologies as a basis for data models and exchange formats
- Ontologies as a basis for metadata taxonomies and meta models
- Intelligent, ontology-based querying and document management systems
- Software for semantic text control
- Knowledge structuring and knowledge sharing in companies and organisations
- Ontology-based translation systems.

The following figures visualize some of the above mentioned uses of ontologies.

Figure 7 illustrates a well-known situation: Customer and Developer do not speak the same 'language', and probably they do not have the same understanding of concepts that are relevant for development of an IT architecture or an IT system.

Concept clarification by means of knowledge modeling, i.e. formalization of the meaning of concepts, furthers understanding and success of the development task, cf. Figure 8.
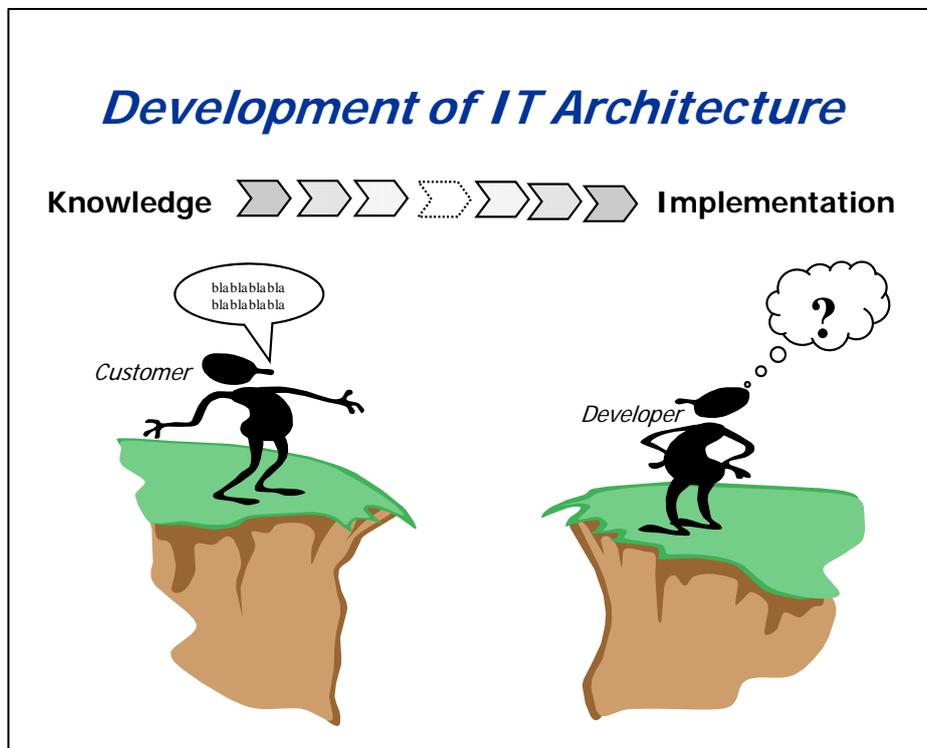
Figure 7. Knowledge gap in the development of IT architecture

Zoom-in on
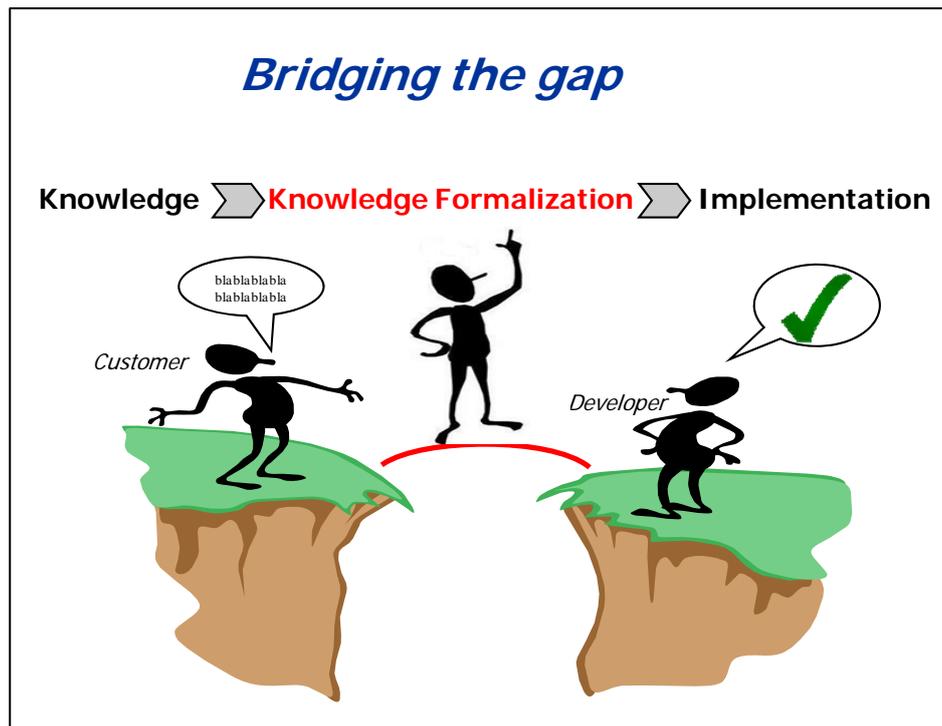
Figure 8. Bridging the gap by means of knowledge clarification

As seen from Figure 6, ontologies and data models have different aims:

- ontologies aim at concept clarification, mutual understanding of concepts and consistent use of terms

- data models aim at specifying the types of information of an IT system and their mutual relationships

Ontologies form a solid basis for the development of data models. Even domain specialists in the same organization may have different understandings of concepts, and in order to avoid errors and inconsistencies, concept clarification by means of ontologies should form the first step in the data modeling process, cf. Figure 9.
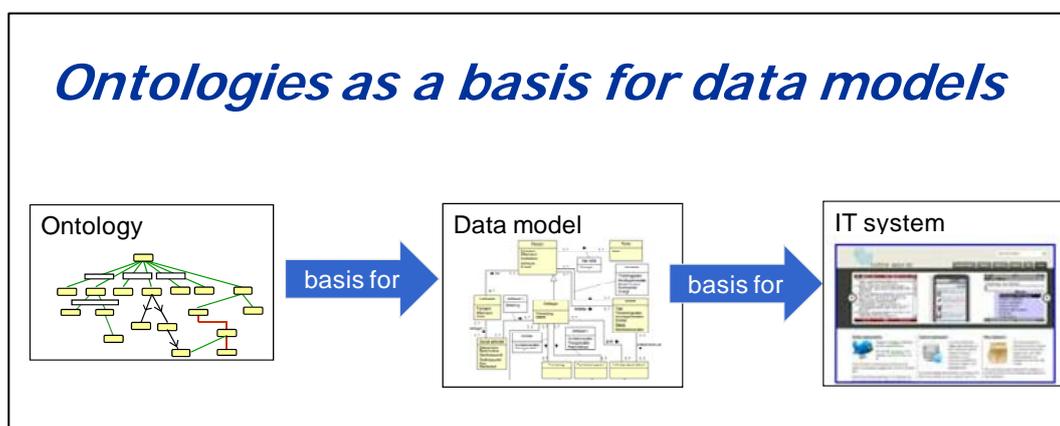
Figure 9. Ontologies as a basis for data models

When The Danish Prison & Probation Service (Kriminalforsorgen) started working on a new IT system, it was clear that the people involved did not speak the same language – they had different understandings of central concepts. Therefore, the Service initiated a cooperation with the DANTERM Centre for concept clarification by means of ontologies.

The problems increase when domain experts from different organisations are involved.

This was obvious in the cooperation between The Danish Prison & Probation Service and other institutions of the Ministry of Justice. The four institutions are going to exchange data, and therefore it is very important that they have the same understanding of the data they are exchanging, cf. Figure 10.

Cooperation between the DANTERM Centre and authorities under the Ministry of Justice:

- The Danish Prison & Probation Service
- The Danish Police
- The Prosecution Service in Denmark
- The Danish Court Administration.
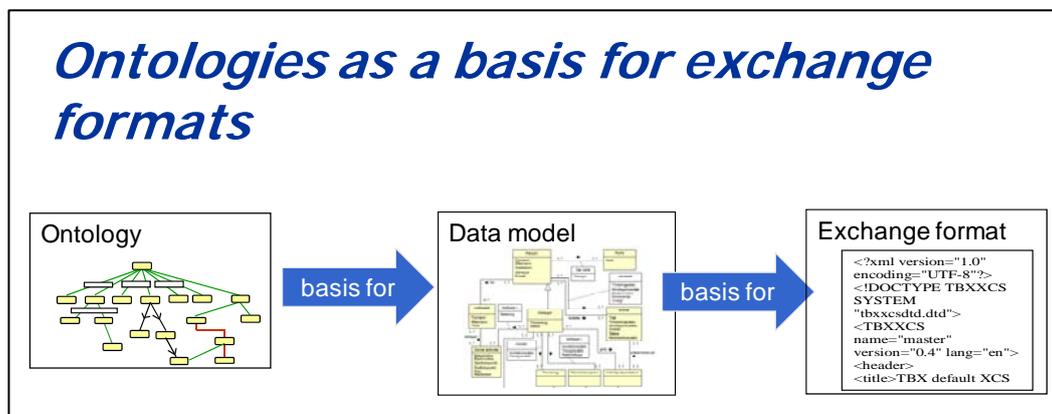
Figure 10. Ontologies as a basis for exchange formats

Also in this case, ontologies contribute to concept clarification and consensus on definitions of concepts. See also 'Modeling of concepts in the Ministry of Justice' below in this volume (Hoffmann, 2011).

The DANTERM Centre has also worked together with the former IT and Telecom Agency (IT- og Telestyrelsen) in order to set up principles for developing metadata taxonomies and to describe components of a meta model for Enterprise Architecture, cf. Figure 11.
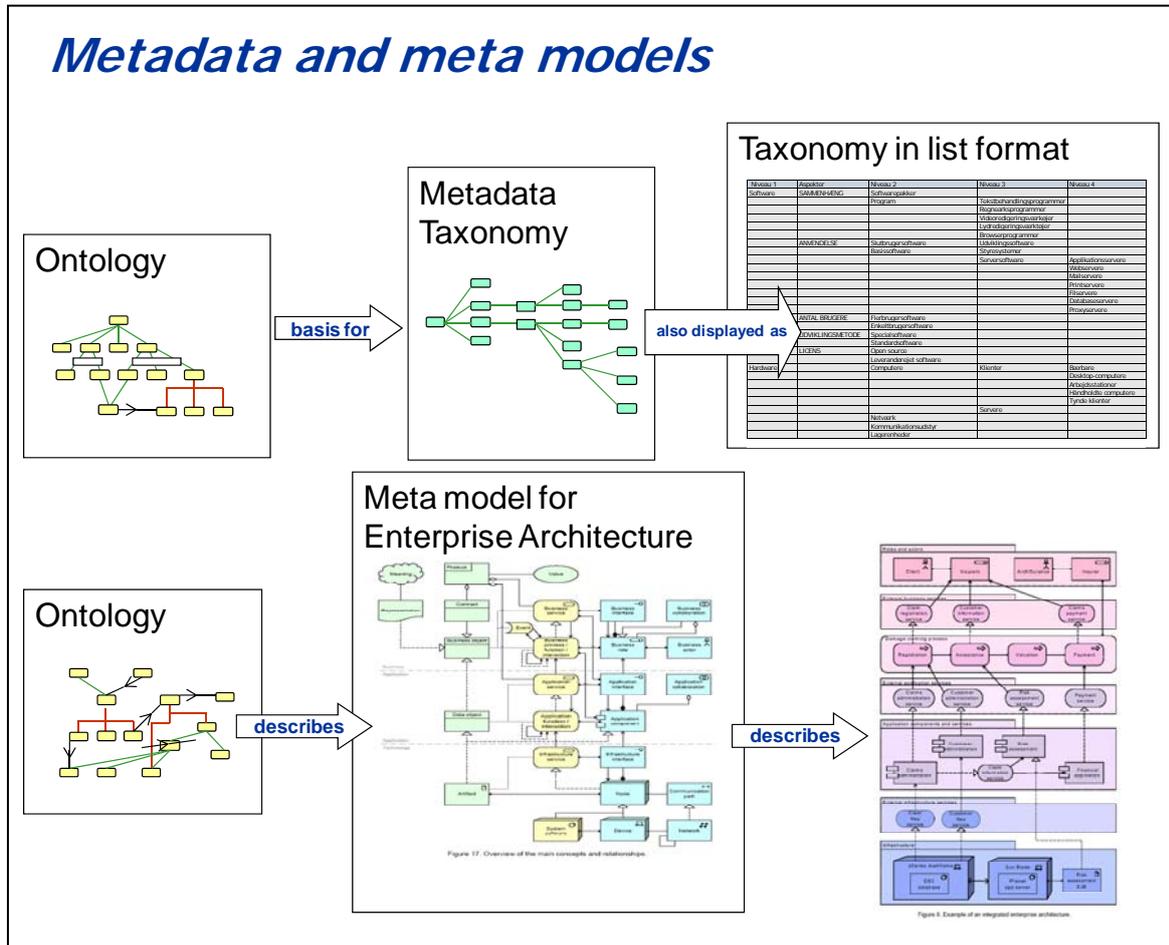
Figure 11. Ontologies as a basis for metadata taxonomies and meta models

Ontology work in Denmark

Many authorities develop ontologies as a basis for their digitization projects. Below some examples of public authorities who have been working or are working on the development of ontologies:

- The National Board of Health (Sundhedsstyrelsen)
    - Concept clarification for example as a basis for *electronic Health Care Records*
- The Capital Region of Denmark (Region Hovedstaden)
    - Concept clarification as a basis for *user interfaces in health care*
- The National Board of Social Services (Servicestyrelsen)
    - *Concept clarification* as a basis for putting social and welfare initiatives into practice

- The Digital Task Force (now under Økonomistyrelsen)
  - Concept clarification as a basis for a *catalogue of public services*
- The (former) IT and Telecom Agency (IT- og Telestyrelsen)
  - Ontologies and classifications (*methods and tools for IT architecture and enterprise architecture*)
- The Danish Prison & Probation Service (Kriminalforsorgen)
  - Concept clarification as a basis for *exchange formats and IT systems*

Huge potential for industrial partners

The potential for industrial partners is also huge. In particular, large businesses and organisations which face challenges due to the complexity of information are likely external partners. Partners in research projects are for example *Novo Nordisk*, *Novozymes*, *IBM* and *Ankiro*.

Most medium-sized and large Danish businesses and organizations have a need for easy access to structured knowledge in several languages. However, knowledge modeling is very time-consuming. Therefore, they need knowledge about and efficient tools for building domain-specific ontologies and systems for managing knowledge as well as a national terminology and knowledge bank, where they can find structured knowledge in Danish and other languages.

Education

The DANTERM Centre gives courses in concept clarification, terminology methods, and knowledge modeling. For example, in 2009: a knowledge modeling course of 5 full days for KL (Local Government Denmark), ITST (Danish IT and Telecom Agency) and KOMBIT (IT cooperation of the Danish municipalities). This course included: principles and tools for ontology work, ontology-based classification systems, thesauri, metadata taxonomies and data models.

It is the vision of the Department of International Language Studies and Computational Linguistics, to develop an executive master programme on Information and IT architecture in cooperation with other CBS departments. One central part of this programme would be knowledge modeling.

Electives will also be offered in other study programmes at CBS.

Need for automatic methods

Ontology construction is very time consuming, and therefore there is a need for automatic ontology construction based on various kinds of input: internet text and text files but also structured data, taxonomies, etc., cf. Figure 12.

On the basis of automatic extraction of information about concepts (concept relations and characteristics) it will be possible to produce draft ontologies, which may be validated against a set of rules for terminological ontologies in an automatic validation module.
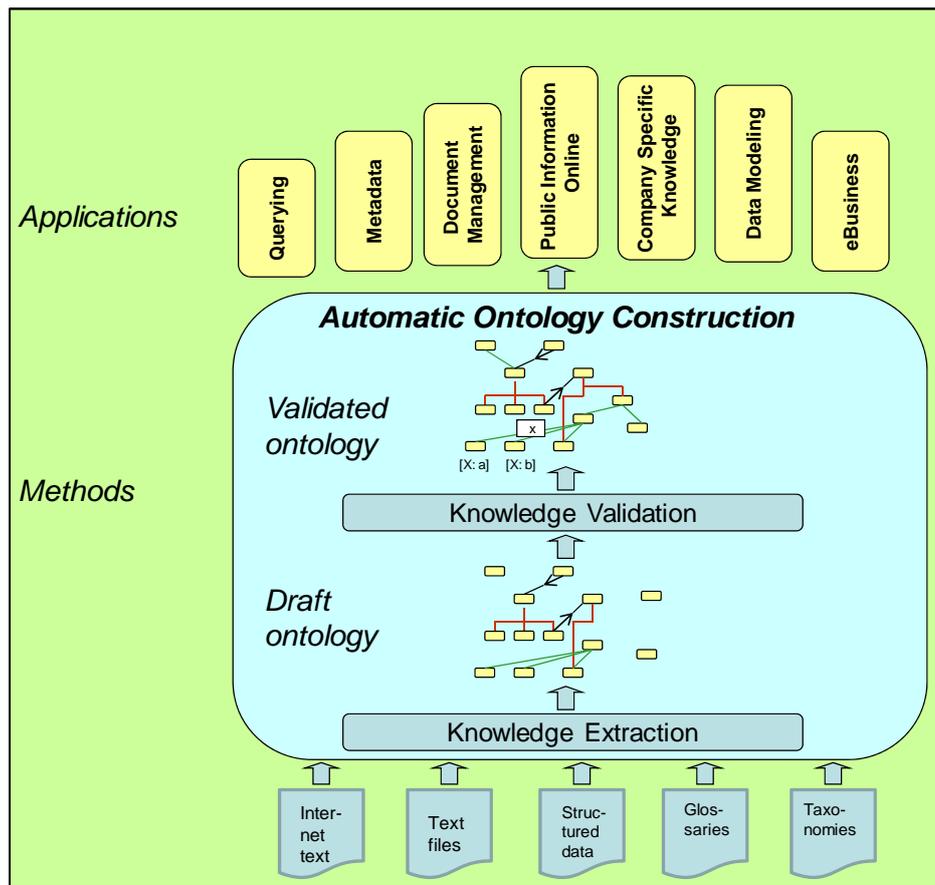
Figure 12. Automatic ontology construction – methods and applications

The DanTermBank Project

This leads us to the DanTermBank project in which we will develop automatic methods for extracting and structuring knowledge for a Danish Terminology and Knowledge Bank.

This Research project aims at establishing the foundation for a National Terminology and Knowledge Bank. As already mentioned, the project is funded by the VELUX Foundation for the years 2011-2014.

The project team comprises:
- Bodil Nistrup Madsen, Hanne Erdman Thomsen (senior researchers)
- Tine Lassen, Louise Pram Nielsen (Post.doc. and PhD-fellow)
- Anna Odgaard, Pia Hoffmann (project managers, the DANTERM Centre)

Background of the team members: terminology, computational linguistics, economy. The project team also includes a system developer.

Background

For many years, CBS has educated translators, and this is the reason why

- Concept clarification and terminology
- Terminology databases and knowledge dissemination
- Computer-aided ontology construction

play an important role in teaching and research.

Already back in the late 1970ies, we had an ambition to establish a term bank, but it was not possible to obtain political and financial support at that time.

In a research project with funding from the Danish Research Council for the Humanities from 1998-2001 we worked on

- Principles and methods for company specific terminology management systems and

- Principles and methods for computer-aided ontology structuring (CAOS).

As a result of this project, the DANTERM Centre started developing the terminology and knowledge management system i-Term, which is used by universities, companies, organizations and public authorities in Denmark and Europe. Ontology modeling in i-Model is based on user input, and has no consistency checking facilities.

Also, research on computer-aided ontology structuring was continued in the CAOS project, and a prototype was developed. This prototype is also based on user input, but it has some facilities for consistency checking and it warns the user, if concepts are placed incorrectly according to their characteristics. In the CAOS project we have formalized principles that have been used for many years in terminology work.

The principles used in CAOS are described in Madsen, Thomsen & Vikner (2004 and 2005). More information on the facilities of i-Model and CAOS may be found in Madsen (2006) and in the article below.

In 2008, the Language Committee of the Danish Government, published the report 'Sprog til tiden' ('Language on demand'), in which the importance of a freely accessible national terminology and knowledge bank was emphasized:

> *A prerequisite for continuous use and development of a national LSP in small countries like for example Denmark is free access to a terminology and knowledge bank comprising domain knowledge in Danish and foreign languages.*

Domain specific knowledge goes beyond traditional dictionary information. In order to clarify and distinguish the meanings of domain specific concepts, these must be described by means of characteristics and relations to other concepts, i.e. in the form of domain specific ontologies. On the basis of these, it is possible to develop consistent definitions that further understanding and the correct use of terms. As already mentioned, terminology work that includes development of ontologies is a very labour-intensive task, and therefore most companies cannot afford this kind of work.

So, when in September 2009 the VELUX Foundation invited researchers from CBS to apply

for funding, we concluded that this was our chance to combine results of previous research and development, and to develop innovative and advanced methods for dynamic and automatic extraction of knowledge about concepts from texts, for automatic construction of ontologies and for target group-oriented knowledge dissemination.

Problems in existing term banks
It is our experience that national term banks suffer from several problems:

1. Insufficient number of terms
2. Little or poor quality of information
3. Too many hits when looking up a term
4. Different needs of user groups not considered

Solutions to these problems may be:

1. Automatic extraction of terms and information about terms from texts
2. Automatic construction of ontologies
3. Use of ontologies for merging terminological entries from different sources
4. Target user-oriented knowledge dissemination

Re 1: If a term bank does not contain a sufficient number of terms, users will not feel encouraged to use it.

A solution to this problem is automatic extraction of terms and information about terms from texts.

Re 2: On the other hand, users will be frustrated if a term bank contains a large amount of terms with only little or poor quality information. Ontologies further understanding and the correct use of terms. Terminology work that includes development of ontologies is, however, a very labor-intensive task, and therefore most term banks do not include ontologies.

A solution to this problem is to use automatic procedures in order to develop ontologies.

Re 3: One way of increasing the amount of terms in a term bank is, as mentioned, to extract terms and information about terms automatically from texts.

Another method is to import terminology from different sources, such as other term banks or existing term lists. However, this approach will often lead to problems, since the term bank will then typically contain many entries comprising the same term, but with varying formulation of the definitions and/or different translations. Therefore, users will often get many hits when looking up a term, and they will find it difficult or impossible to choose the right entry.

A solution to this problem may be to use ontologies in the process of merging data from existing sources.

Re 4: Our project addresses the problems of the different needs of different user groups. We will use eye-tracking as one method for analyzing user needs.

In the next section, Figure 13, I present an overview of the DanTermBank project.

Overview of the DanTermBank project
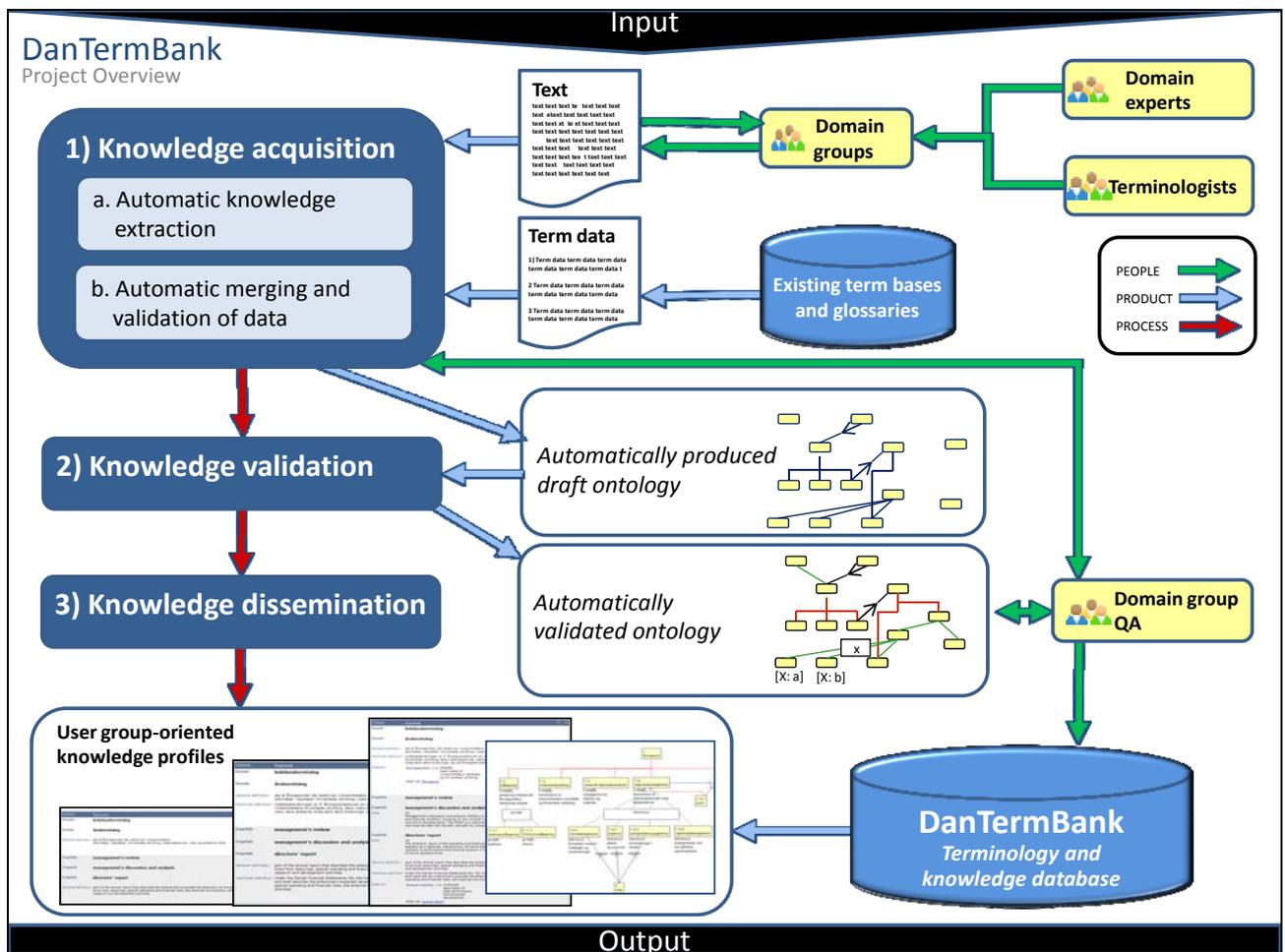
Figure 13: Overview of the DanTermBank project

In this project, we will develop ontologies and definitions within selected pilot domains, and this will result in terminological data within two important economic domains: taxes and auditing. Domain experts and users will be involved in various stages of the project.

The project consists of 3 subprojects, which I will present briefly below. More information may be found in Lassen, Madsen & Thomsen (2011) and Madsen, Thomsen, Halskov & Lassen (2010).

Knowledge acquisition
Subproject 1) a: Automatic knowledge extraction
The aim of this subproject is to develop new models of and methods for automatic extraction of concepts and information about concepts as well as a prototype which can automatically produce a draft version of a terminological ontology on the basis of an existing domain-specific text corpus, or on the basis of domain texts automatically collected from the Internet. Thus, the draft ontologies will contain subdivision criteria and characteristics as formal feature specifications on concepts.

Subproject 1) b: Automatic merging and quality assurance of data
The idea is that the contents of the terminology and knowledge bank should also be extended by means of import of the many term lists from the internet or from authorities, organisations and companies. By means of automatic ontology construction on the basis of the imported data, e.g. definitions, we aim at assuring the quality of the imported data.

Subproject 2): Knowledge validation
The aim is to develop methods and a prototype that may be used for automatic validation and dynamic expansion of the draft ontologies that result from the automatic knowledge extraction.

The Knowledge validation facilities will comprise:
- Automatic positioning of concepts in ontologies
- Automatic consistency checking
- Automatic changes to ontologies
- Dynamic updating of ontologies

Subproject 3): Knowledge dissemination
This subproject will focus on presentation of data in the term bank. Traditionally, terminology and lexicography have been separate research fields with different approaches to compilation and presentation of data. However, modern technology offers unlimited opportunities to meet the needs for several target groups in one database by offering the possibility of choosing between different presentations.

One important user group is businesses and organizations. Another important user group is schools and universities. Since English is often used in business communication and teaching, there is a need for at terminology and knowledge bank comprising structured knowledge about terms in both Danish and English. A third user group is foreigners, who, for example, need to find information in the web portals of the Danish authorities. Within these main user groups, however, a more detailed user group classification is needed.

The project will aim at developing prototypes for dynamic target-user oriented interfaces. Here, we will use eye-tracking methods for investigating the behavior of users while they use the terminology and knowledge bank for various purposes.

Conclusion
For a long period, many resources have been allocated to general language dictionaries, lexical databases and word nets in Denmark. There is, however, a big need for domain specific knowledge within scientific, technical, economical and legal domains which can be made accessible by means of a Danish terminology and knowledge bank.

In this paper, I have described the problems common to many existing term banks: insufficient quantity and quality of terms and information about terms, lack of dynamic, user-oriented information profiles.

In the DanTermBank project we will develop facilities for *automatic* methods for knowledge acquisition, knowledge structuring and validation and dynamic target-user oriented interfaces. To our knowledge, no other national term banks have developed similar facilities.

Acknowledgements

References

Hoffmann, Pia (2011). 'Modeling of concepts in the Ministry of Justice', this volume.

Lassen, Tine; Madsen, Bodil Nistrup; Thomsen, Hanne Erdman (2011). 'Automatic Knowledge Extraction and Knowledge Structuring for a National Term Bank'. In: *NODALIDA 2011 workshop: Creation, Harmonization and Application of Terminology Resources.* NEALT Proceedings Series Vol. 12. http://hdl.handle.net/10062/17274 (2011-05-09)

Madsen, Bodil Nistrup (2006).http://research.cbs.dk/research/terminological ontologies(120631) In: *Semantic Systems: From Visions to Applications*. Proceedings of the Semantics 2006, p. 271-282.

Madsen, Bodil Nistrup & Anna Elisabeth Odgaard (2010). 'From Concept Models to Conceptual Data Models'. In: Úna Bhreathnach & Fionnuala de Barra Cusack(eds.): *TKE 2010: Presenting Terminology and Knowledge Engineering Resources Online: Models and Challenges*, August 14 2010, Dublin.

Madsen, Bodil Nistrup & Hanne Erdman Thomsen (2009). 'Ontologies vs. classification systems'. In: *Proceedings of the NODALIDA 2009 workshop WordNets and other Lexical Semantic Resources — between Lexical Semantics, Lexicography, Terminology and Formal Ontologies*. Editors: Bolette Sandford Pedersen, Anna Braasch, Sanni Nimb and Ruth Vatvedt Fjeld. NEALT Proceedings Series, Vol. 7 (2009), 27-32.

Madsen, Bodil Nistrup, Hanne Erdman Thomsen, Jakob Halskov & Tine Lassen (2010). 'Automatic Ontology Construction for a National Term Bank', In: Úna Bhreathnach & Fionnuala de Barra Cusack(eds.): *TKE 2010: Presenting Terminology and Knowledge Engineering Resources Online: Models and Challenges*, August 14 2010, Dublin.

Madsen, Bodil Nistrup, Hanne Erdman Thomsen & Carl Vikner (2004). 'Principles of a system for terminological concept modelling'. *Proceedings of the 4th International Conference on Language Resources and Evaluation, Vol. I*, pp. 15-18. Lisbon.

Madsen, Bodil Nistrup, Hanne Erdman Thomsen & Carl Vikner (2005).'Multidimensionality in terminological concept modeling'. In: Bodil Nistrup Madsen, Hanne Erdman Thomsen (eds.): *Terminology and Content Development, TKE 2005*, 7th International Conference on Terminology and Knowledge Engineering, Copenhagen, 2005, s. 161-173.

Nielsen, Louise Pram (2009). Terminologisk begrebsafklaring på skatteområdet – en analyse af terminologilæren sammenlignet med fagleksikografi. ED-Afløsningsopgave afleveret på Åben Uddannelse, Copenhagen Business School den 16. november 2009.

***

CAOS
– Construction and validation
of Terminological Ontologies

Hanne Erdman Thomsen and Bodil Nistrup Madsen
Department of International Language Studies and Computational Linguistics
Copenhagen Business School, Denmark
het.isv@cbs.dk / bnm.danterm@cbs.dk

Abstract

This paper presents some principles of terminological ontologies implemented in the prototype that has been developed in the research project CAOS - Computer-Aided Ontology Structuring. Furthermore, some issues that have to be faced to further develop facilities for automatic consistency checking and automatic changes to ontologies, are discussed. The presentation will illustrate central facilities of the current version of the CAOS prototype, which is interactive and presupposes an end-user with a background in terminology rather than in formal ontology.

Introduction

A terminological ontology is a domain specific ontology, cf. for example the categorization of ontologies by Guarino (1998). We use the term *terminological ontology* as synonym to the term *concept system*, which is normally used in terminology work, cf. for example (ISO 704, 2000).

The principles of terminological ontologies presented here, build on the principles of terminology work as presented in (ISO 704, 2000), but have been further developed in the research and development project CAOS - Computer-Aided Ontology Structuring - whose aim is to develop a computer system designed to enable semi-automatic construction of concept systems, or ontologies, cf. (Madsen et al., 2005).

Terminological ontologies model concepts and the relations between them, and a concept is described by means of characteristics that denote properties of individual referents belonging to the extension of that concept. Other ontologies most commonly model classes, described by means of properties, and the relations between classes.

It is possible to use all types of concept relations in CAOS. The system offers a set of concept relations organized in a taxonomy, cf. (Madsen et al., 2002). It is also possible for the user to introduce user defined relations. For other presentations of concept relations, see for example (Nuopponen, 2005).

The CAOS Prototype

The backbone of terminological concept modeling in CAOS is constituted by characteristics modeled by formal feature specifications, i.e. attribute-value pairs, cf. (Carpenter, 1992). The use of feature specifications is subject to a number of principles and constraints.

Figure 1 presents part of an ontology for prevention created in CAOS. As can be seen, the graphical presentation is UML-based.

Consistency checking in CAOS

The technology developed in CAOS enables validation of the inheritance of characteristics when a new concept is introduced into a concept system. In a type hierarchy, subordinate concepts inherit characteristics from their superordinate concepts, and hence it is possible to validate whether the position of a given concept allows for the characteristics associated with it.

   The facilities for semi-automatic construction of ontologies and for consistency checking in CAOS are, among other things, based on the introduction of dimensions and dimension specifications. A dimension of a concept is an attribute occurring in a (non-inherited) feature specification of one of its subordinate concepts, i.e. an attribute whose possible values allow a distinction between some of the subconcepts of the concept in question. A dimension specification consists of a dimension and the values associated with the corresponding attribute in the feature specifications of the subordinate concepts: dimension: [value1| value2| ...]. In this way, the principle of subdivision criteria that has been used for many years in terminology work, has been formalized in CAOS.
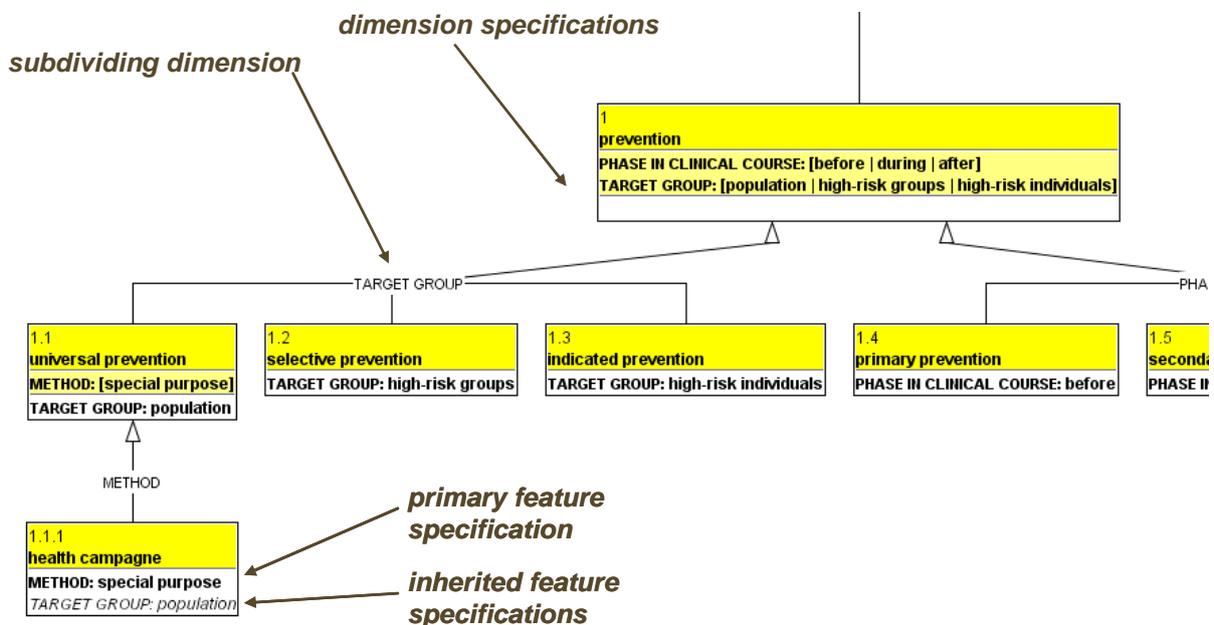

Figure 1. Extract of an ontology for prevention

One or more of the dimensions of a concept must be chosen as the subdividing dimensions. Subdividing dimensions must be chosen in such a way that each daughter concept has one and only one feature specification containing as an attribute a subdividing dimension of the

mother concept. This ensures that there are no overlapping subdividing dimensions, and hence no overlap in partitions.

In the following, a brief description of some important principles of CAOS will be given:
- grouping by subdividing dimensions, including choice of subdividing dimensions and no overlapping of subdividing dimensions,
- uniqueness of primary feature specifications and
- uniqueness of dimensions.

Grouping by subdividing dimensions

From figure 1 it is seen that *prevention* may differ with respect to both target group and phase in clinical course. However, in the case of the three concepts *universal prevention*, *selective prevention* and *indicated prevention* it is obvious that TARGET GROUP must be chosen as the subdividing dimension (subdivision criterion). If the user tries to choose a second dimension as subdividing dimension for the three mentioned subordinate concepts, CAOS will not allow it, and will consequently warn the user. The feature specifications comprising the subdividing dimension (referred to as the delimiting feature specifications) will form the basis for the definition of the three concepts.

Constraints in CAOS related to subdivision criteria are:

- A concept (with only one mother concept) may contain at most one delimiting feature specification
- A concept (of level 2 or below) must contain at least one delimiting feature specification

Another constraint is that an attribute may only be associated with one value in a feature structure on a given concept (a combination of two or more feature specifications on a concept is called a feature structure). If the user attempts to create a concept *universal selective prevention* with two superordinate concepts within the same group (dimension: TARGET GROUP), this would mean that the attribute TARGET GROUP would be associated with two values in the feature structure for *universal selective prevention*: TARGET GROUP: population and TARGET GROUP: high-risk groups. CAOS will not allow this 'illegal polyhierarchy'. This type of error is also known as a partition error (Gómez-Pérez et al. 2003).

In Protégé[1] this can be handled be adding a new superordinate concept to a concept on the basis of the formal definition of the concepts in question. However, this treatment is not feasible for the end users we have in mind, who have no training in formal logic or similar.

Uniqueness of dimensions

The principle of uniqueness of dimensions states that a given dimension may occur on only one concept in an ontology. Uniqueness of dimensions helps to create coherence and simplicity in the ontological structure because concepts that are characterised by means of primary feature specifications with the same dimension must appear as coordinate concepts on the same level having a common superordinate concept.

---

[1] http://protege.stanford.edu/

Uniqueness of feature specifications
The principle of uniqueness of feature specifications stipulates that a feature specification may occur only once in a terminological ontology as primary. A primary feature specification is entered on a concept directly by the terminologist, as opposed to inherited feature specifications, which are inherited from superordinate concepts.

Uniqueness of dimensions (the previous principle) means that a given primary feature specification can only appear on concepts that are daughters of the concept containing the relevant dimension. Uniqueness of primary feature specifications means that a given primary feature specification can only appear on one of these daughters. If the terminologist tries to insert the primary feature specification [TARGET GROUP: population] on the concept *selective prevention*, CAOS will report that [TARGET GROUP: population] is already specified on the concept 1.1 *universal prevention*.

The motivation of the principle of uniqueness of primary feature specifications is that

- characteristics will always serve to distinguish concepts, and
- common characteristics should be located on a common superordinate concept (this principle may contribute to the identification of potential gaps in the ontology).

Characteristics of the CAOS prototype compared to other ontology editors
Several other tools for creating ontologies have been (or are being) developed, e.g. Protégé and WebODE[2].

The main difference between the system for terminological ontologies, described here, and other systems is that in the latter, terminological information cannot be modeled and presented in the same way. This information, i.e. subdivision criteria and dimension specifications, is crucial in the development of terminological ontologies. Furthermore, in order to check conformance to the constraints mentioned in section 2.2 – 2.4, the end user must be able to formulate formal constraints for each subdivision criterion. In CAOS, the constraints are part of the system.

Further Development of the CAOS Prototype

In a new project we aim to develop an additional prototype that will be able to automatically build a first draft ontology on the basis of a domain-specific text corpus. This prototype will be based on a combination of existing and new methods and principles for automatic extraction of concepts and information about concepts, i.e. characteristics and concept relations.

Another aim is to further develop CAOS so that it may be used for automatic validation of draft ontologies that are the result of the automatic knowledge extraction described above. The new prototype will not just be able to detect errors, it will also propose corrections of errors. For example it will automatically handle partition errors. To our knowledge no other systems have such capabilities.

To further develop facilities in CAOS for automatic consistency checking and automatic changes to ontologies, various issues have to be dealt with.

---

[2] http://mayor2.dia.fi.upm.es/oeg-upm/index.php/en/downloads/60-webode

Validation of an ontology vs. validation of one concept

First of all, the technology currently used in CAOS validates one concept at a time, while the new prototype will need to validate an entire ontology provided by the knowledge extraction module.

Characteristics vs. relations

In CAOS, a concept may have both feature specifications and relations to other concepts. However, a given characteristic of a concept can be modeled either as an attribute-value pair or a relation-concept pair, e.g. in Figure 1, the characteristic modeled by the feature specification [TARGET GROUP: population] could have been modeled as a relation (HAS_TARGET_GROUP) to another concept (*population*).

The ontology extraction module will not be able to distinguish between attributes and relations. Therefore, in the new prototype, relations (other than type relations) and attributes of characteristics will have to be treated identically. In the validation they will be treated as attributes of characteristics, and the related concepts will be treated as values. This raises a theoretical research issue: is it necessary to differentiate relations and characteristics? If so, what is the difference?

Multiple values

A problem related to the above is that the CAOS technology allows a given concept to have only one value for a given attribute, while it may be related to several other concepts with the same relation. The extraction tool is bound to deliver more than one concept for a given relation (or value for a given attribute) for any concept. The CAOS technology needs to be modified to handle this.

Some relations may only be applied to a given concept once. For example, no concept can have more than one instance of the relation HAS_LENGTH_IN_CM. This corresponds to the CAOS principle mentioned above, i.e. that for a given attribute a concept can have at most one value. Hence a research issue to be investigated is whether these relations can be distinguished from those allowing for multiple instances, since this is important for validation.

Specialized values

An issue relating to characteristics is that of specialized values. In order to handle this, the CAOS technology needs to be enhanced to include a type hierarchy of values (or related concepts). The use of value hierarchies has been implemented e.g. in the Lexical Knowledge Base system (LKB) first developed by Ann Copestake for lexical semantics and further enhanced for HPSG[3] purposes, c.f. (Copestake, 1993).

Automatic positioning

A prerequisite for making automatic changes in the ontology based on the validation is to be able to position a concept in an existing type hierarchy by employing the characteristics registered for that concept. Techniques for positioning concepts and making automatic changes to the ontology are to be developed.

---

[3] Head Driven Phrase Structure Grammar

Perspectives

Terminological ontologies offer very detailed information about concepts, e.g. feature specifications, subdivision criteria and dimension specifications. The question is whether this information is useful in the various applications of ontologies. Undoubtedly, this information is needed for concept clarification, for example with a view to the definition of central concepts in the use of IT systems for information storage and retrieval.

In the SIABO project, Semantic Information Access through Biomedical Ontologies, cf. http://siabo.org, it is planned to test whether terminological ontologies will also add value to systems for ontology-based information retrieval.

References

Carpenter, Bob. 1992. *The Logic of Typed Feature Structures*. Cambridge University Press, UK.

Copestake, Ann. 1993 *The Compleat LKB, Technical Report No. 316*, University of Cambridge.

CWA 15045. 2005. *CEN Workshop Agreement: Multilingual Catalogue Strategies for eCommerce and eBusiness*.

Gómez Pérez, Asunción, Mariano Fernández-López, Oscar Corcho: (2003) "Ontological Engineering". Advanced Information and Knowledge Processing series. ISBN 1-85233-551-3. Springer Verlag.

Guarino, Nicola. 1998. Formal Ontology and Information Systems. *Formal Ontology in Information Systems, Proceedings of the First International Conference (FOIS'98)*, June 6-8, Trento, Italy, 3-15. Ed. Nicola Guarino. Amsterdam: IOS Press.

ISO 704. 2000. *Terminology work — Principles and methods.* Genève: ISO.

Madsen, Bodil Nistrup, Hanne Erdman Thomsen & Carl Vikner. 2005. Multidimensionality in terminological concept modelling. Bodil Nistrup Madsen, Hanne Erdman Thomsen (eds.): *Terminology and Content Development, TKE 2005, 7th International Conference on Terminology and Knowledge Engineering*, Copenhagen: 161-173.

Madsen, Bodil Nistrup, Bolette Sandford Pedersen & Hanne Erdman Thomsen. 2002. The Role of Semantic Relations in a Content-based Querying System: a Research Presentation from the OntoQuery Project. Simov, Kiril & Atanas Kiryakov (eds.): *Proceedings from OntoLex '2000, Workshop on Ontologies and Lexical Knowledge Bases,* Sept. 8-10 2000, Sozopol, Bulgaria: 72-81.

Nuopponen, Anita. 2005. Concept Relations. Bodil Nistrup Madsen, Hanne Erdman Thomsen (eds.): *Terminology and Content Development, TKE 2005, 7th International Conference on Terminology and Knowledge Engineering*, Copenhagen, 127-138.

***

Modeling of concepts in the Ministry of Justice

Models for co-operation on concept modeling

Pia Lyngby Hoffmann
DANTERMcentre
Copenhagen Business School, Denmark
plh.danterm@cbs.dk

The co-operation between the various authorities under the Danish Ministry of Justice on the clarification and modeling of concepts comprises the following four institutions: The Danish Prison & Probation Service (Kriminalforsorgen), The Danish Police, The Prosecution Service in Denmark and The Danish Court Administration.

The work group is constituted by 12 persons: 4 from the Prison & Probation Service, 4 from the police, 1 from the Prosecution Service and 1 representing the courts. The Prison & Probation Service has undertaken the task of acting as a Secretariat co-ordinating the modeling work.

Originally, the concept modeling work was initiated in The Danish Prison & Probation Service as a prerequisite for redesigning the existing IT system for managing clients – the so-called Client System. But today's modern IT architecture makes it possible to exchange data between authorities, which is exactly the idea of these four authorities, so the scope of the modeling work was extended.

To be able to exchange data, it is important first to clarify which documents are being exchanged between the authorities today, and next to agree on an interpretation of the importance and contents of these documents.

To create an overview of the concepts, we use the DANTERM centre tools i-Term and i-Model – and it is precisely the combination of a traditional term base and a concept modeling tool that is unique and ideal for obtaining results as the one you see below.

Zoom-in on

###.IM-sagsdokumenter der udveksles mellem Justitsministeriets myndigheder/Del 1 UNDER UDARBEJDELSE - 28/09/2011 15:21 (built-in account)

DEL 1 AF DIAGRAM SAGSDOKUMENTER - FORTSÆTTES PÅ DEL 2

FORMÅL MED SAGSDOKUMENT

dokument
en afgrænset samling af information, i en kendt struktur og gemt på et kendt medie

fysisk sag

sagsdokument
dokument som er en del af fysisk sag

manifestation eller affedring af den intellektuelle sag

indkaldelsesdokument
FORMÅL MED SAGSDOKUMENT: udsende indkaldelse
sagsdokument med besked om at give møde på et angivet sted og et angivet tidspunkt

MODTAGER

indkaldelse til afsoning
MODTAGER domfældte
indkaldelsesdokument til domfældte om at møde til afsoning af dom

stævning
MODTAGER tiltalte
indkaldelsesdokument til tiltalte om at møde til hovedforhandling

indridelsesindkaldelse
MODTAGER vidne
indkaldelsesdokument til vidne med pligt til at møde i retten

anmeldelsesdokument
FORMÅL MED SAGSDOKUMENT: indgive anmeldelse
sagsdokument der indeholder formel indberetning af forhold til en myndighed

RESULTAT AF ANMELDELSE

højesteddet til anstedet
RESULTAT AF ANMELDELSE indsættkel i arrest
anmeldelsesdokument med lovmæssig hjemmel til at indsætte en person i arresthus

anmeldelse til afsoning
RESULTAT AF ANMELDELSE fuldbyrdelse
anmeldelsesdokument fra Anklagemyndigheden til Kriminalforsorgen om at fuldbyrde dommen

anmeldelse af straffart forhold
RESULTAT AF ANMELDELSE politianmeldelse
anmeldelsesdokument til Politiet vedrørende straffart forhold

indberetning om vilkårsovertrædelse
RESULTAT AF ANMELDELSE
vurdering angående skærpelse af dom
anmeldelsesdokument fra Kriminalforsorgen til Anklagemyndigheden om overtrædelse af vilkår der kan give anledning til skærpelse af dom

afmelding af efterlysning RESULTAT AF ANMELDELSE
annullering af efterlysning
anmeldelsesdokument til Politiet om at eftersøgning af person skal indstilles

efterlysning
RESULTAT AF ANMELDELSE
efterlysning af person
anmeldelsesdokument fra Politiet om at en bestemt person eftersøges

høringsdokument
FORMÅL MED SAGSDOKUMENT: iværksætte høring
sagsdokument med anmodning om en udtalelse fra en myndighed til en anden myndighed

UDTALENDE MYNDIGHED

politihøring
UDTALENDE MYNDIGHED politi
høringsdokument hvor en myndighed indhenter udtalelse fra Politiet vedrørende en klient i en konkret sag

HØRINGSFORMÅL

politihøring vedr. udgang
HØRINGSFORMÅL tilfældeighed
politihøringsdokument med Kriminalforsorgens indstilling af en klient til udgang og in- imodtning om Politiets udtalelse herom

politihøring vedr. verserende sag
HØRINGSFORMÅL formodet dom
politihøringsdokument med Kriminal- forsorgens anmodning om Politiets udtalelse om klientens even- tuelle verserende sager

politihøring vedr. prøveløsladelse
HØRINGSFORMÅL indstilling til prøveløsladelse
politihøringsdokument med Kriminal- forsorgens indstilling af en klient til prøveløsladelse og anmodning om Politiets udtalelse herom

HØRINGSFORMÅL ???????
skal ændres

udtalelse vedr tilsynsforløb
HØRINGSFORMÅL tilsynsforløb
høring vedrørende fuldbyrdelsesforløb hvor Politiet indhenter udtalelse fra Kriminalfor- sorgen vedrørende en klients tilsynsforløb

erklæringsdokument
FORMÅL MED SAGSDOKUMENT: fremlægge erklæring
sagsdokument der indeholder saglig udtalelse om hvordan noget forholder sig

UDTALENDE MYNDIGHED

RESULTAT

høring vedrørende fuldbyrdelsesforløb
UDTALENDE MYNDIGHED Kriminalforsorgen i Frihed
høringsdokument hvor en myndighed indhenter udtalelse hos Kriminalforsorgen vedrørende en klients fuldbyrdelsesforløb

straffeattest
RESULTAT
oplysning af faktiske algerelser
erklæringsdokument fra kriminalregisteret, der viser udvalgte strafferetlige algerelser vedrørende en person

erklæring fra sagkyndig
RESULTAT vurdering
erklæringsdokument med ekspertudtalelse

FAGEKSPERT

mentaliserklæring
FAGEKSPERT psykiater
erklæring fra psykiater om en sigtets psykiske tilstand

lægeerklæring
FAGEKSPERT praktiserende læge
erklæring fra en persons helbredsmæssige tilstand

teknisk erklæring
FAGEKSPERT specialist
erklæring fra specialist eller organisation med specialviden vedrørende teknisk forhold

obduktionsrapport
FAGEKSPERT patolog
erklæring fra patolog om undersøgelses- resultater for en afdød person

begæringsdokument
FORMÅL MED SAGSDOKUMENT: fremsætte begæring
sagsdokument med anmodning om at få oplyst et fortsættende

BEGÆRINGSFORMÅL

GENSTAND FOR APPEL

anmodning om saglyndig udtalelse
BEGÆRINGSFORMÅL saglyndig udtalelse
begæringsdokument til at indhente saglig special viden om en vurdering af spørgs- mål af betydning for en sag

anmodning om kendelse
BEGÆRINGSFORMÅL kendelse
begæringsdokument til Politi eller Anklagemyndighed om at få rettens afgørelse af en kendelse vedrørende indgreb eller efter- forskningsskridt

personundersøgelse
FAGEKSPERT kriminalforsorgen
erklæring fra Kriminalforsorgen i Frihed (KiF) efter anmodning fra Anklagemyndighed eller Politi til afklaring af en klients personlige forhold med henblik på fastsættelse af straf eller anvendelse af anden retslige endf

kæreskrift
GENSTAND FOR APPEL kendelse
begæringsdokument til en højere instans med appel af den af retten afsagte kendelse eller beslutning

ankemeddelelse
GENSTAND FOR APPEL dom
begæringsdokument til enhejæringsdokument til en højere appelfhøjere instans med appel af den af retten afsagte dom

HVEM ER EKSPERT

anmodning om personundersøgelse
HVEM ER EKSPERT kriminalforsorgen
anmodning om saglyndig udtalelse fra Politi eller Anklage- myndighed til Kriminalforsorgen om udfærdigelse af en personundersøgelse til brug for retssagen

41

It is not readable. However, it does give you an idea of just how extensible merely the concept "case document" is. There are 68 concepts. (We started out with over 100 concepts.)

Concept modeling is the basis for:
- Data being interpreted in the same way in all systems
- Data being able to flow freely and correctly
- Data being presented in a way that makes it possible to get an overview of accessible information

That data is interpreted in the same way in all systems means that it will be possible to compare, quantify and create statistics from certain data. That data is able to flow correctly means that you avoid manual transfers and that errors occur. To get an overview of the accessible information eases the exchange of data between authorities.

The Danish Prison & Probation Service's concept modeling was conducted in smaller groups consisting of domain experts who had been introduced to the methods and principles of terminology work. Every group worked with certain subjects and they used i-Term as the modeling tool. No homework was necessary – the groups came to the meetings and discussed the subjects until they agreed.

When the modeling work under the Ministry of Justice was initiated, we tried to copy this way of working in what we call Plenum Workshops where all parties participated on equal terms. This type of workshop was planned to take place once a month and took place over two days' of concentrated work. These workshops were repeated until everybody was satisfied with the diagram and the corresponding definitions.

However, in the process, it was necessary to change the way of working, as it became clear that these Plenum Workshops were not the ideal way of working in this forum. Perhaps because there were too many people in the group – bear in mind that there were 10 domain experts from four different but more or less overlapping domains. Perhaps because there were cultural differences between the four institutions. Perhaps because under the surface there was a non-explicit power clash between the authorities.

Therefore, we changed working method and chose to work in Authority Specific Workshops. This working method required an increased co-ordination effort in the Secretariat. The Secretariat conducts a shorter workshop with each of the authorities. It is necessary to document carefully what has been discussed during each workshop so we know who said what and why.

Suggestions to concepts, definitions, changes and comments were registered in an Excel-sheet. However, the definitions might as well have been documented in i-Term where we could have entered one definition suggestion from each authority.

On the basis of suggestions to concepts made by the authorities, the Secretariat creates a concept system which is commented on and revised later on in a workshop.

Once all the authorities have been heard, a concept system is created, where after the authorities may comment on it and add to it in a second workshop.

The Secretariat decides when the work has been finalised.

Advantages and disadvantages to Authority Specific Workshops:

Advantages:
- Faster results
- Fewer hours spent with the authorities
- A more structured work form
- Cheaper in the long run

Disadvantages:
- Time consuming for the Secretariat
- More consultancy hours
- More co-ordination
- Difficult to preserve the overview
- Domain experts have to do homework
- No on site agreement

Originally, we estimated that the time invested in the latter type of workshops would be more or less the same seen over a period of half a year. This was not the case. As mentioned earlier, the Secretariat spent more time, and the DANTERM centre delivered more consultancy hours, whereas the domain experts spent considerably less time.

In this way we achieve our results faster and therefore, the new way of working is actually cheaper and more productive. Although we apparently list more disadvantages than advantages, we do see that in this specific case under these specific circumstances and with this specific constellation of participants this is the method that is most effective and efficient. This is the method that our participants have indicated to be the best and the method that has rendered the most progress.

We may conclude that from our experience when you need a larger group of people to work together you should always bear the following in mind:
- Consider the number of participants/interested parties
- Beware of cultural differences in the various organisations
- Remember that there might be hidden hierarchies
- Decide how to facilitate the process
- Remember to promote/communicate the necessity of working with concepts
- Create ownership

***

Report

From print to web 2.0: The changing face of discourse for special purposes

Elizabeth Rowley-Jolivet[a] and Sandra Campagna[b]

[a] Laboratoire Ligérien de Linguistique
   Université d'Orléans, France
   elizabeth.jolivet@univ-orleans.fr

[b] Facoltà di Economia
   Università degli Studi di Torino, Italia
   campagna@econ.unito.it

Abstract

This report addresses the recent changes induced in professional discourse genres as they migrate from traditional print media to the Web 2.0 platform on the Internet. After a brief characterization of the specificities and affordances of the Web 2.0 environment, a selected number of themes that are currently being researched are outlined, namely: the reshaping of existing discourse genres and a rethinking of genre theory; the implications of collaborative authoring; the impact of narrative elements on professional discourse; the personalization and informality of Web-mediated communication. Some of the challenges that these developments pose to discourse for special purposes are highlighted.

1 Introduction

As Marshall McLuhan presciently remarked many years ago, "A new medium is never an addition to an old one, nor does it leave the old one in peace. It never ceases to oppress the older media until it finds new shapes and positions for them" (McLuhan 1964: 158). Over the last decade, the new medium to which specialized discourses have had to adapt is the second-generation web or Web 2.0: while the term itself has been criticized by some (e.g. Tim Berners-Lee 2006), it remains a useful shorthand for the wide-ranging and qualitative changes that have taken place in how the web is now being used. These changes can be briefly characterized - in contrast to earlier forms and uses of the Internet - by user participation and user-generated content, where consumers have become *prosumers* covering both roles of producer and consumer engaged in uploading rather than just downloading material. Recent web developments have also involved: the creation of virtual communities such as social networking sites, blogs, twitters, podcasts or mashups (comprising new ways of indexing

material such as tags and folksonomies); new forms of archiving, data sharing and collaboration which are now possible thanks to the interoperability of the software used; and changes in 'footing' and participation frameworks (Goffman 1981) between the various participants interacting in this new communicative setting. To these more recent changes, one can add the typical technological and semiotic affordances of the Internet such as hyperlinking, the non-linearity of navigation compared to the traditionally linear reading mode, and heightened multimodality, as well as an unprecedented acceleration in the time-frame of interaction.

The implications of these developments have generated a substantial amount of research which has addressed the issue from a variety of perspectives: globalized business practices (dubbed 'the flat world' by Friedman 2007); political accountability (Dutton 2007; Cardon 2010); e-publishing; discourse analysis; communication studies; societal evolution; or the socio-technical infrastructure underpinning these developments, as evidenced by the existence of numerous journals devoted to these topics (e.g. *New Media and Society*; *Studies in Communication Science*; *Media, Culture and Society*; *Language@Internet*; *Journal of Computer-Mediated Communication*; *Information, Technology & People*, to name but a few). The range and diversity of these approaches preclude any comprehensive review of the field. The much more modest aim of this short report is to attempt to outline, from a discourse analysis perspective, some of the research themes that are currently being investigated in relation to the impact of Web 2.0 on the discourse of specialized communities.

Professional communities – business, academia, journalism… - have over time generated the discourse genres that correspond to their communicative and epistemological needs: company annual reports and press releases, research articles and dissertations, news reports and editorials, etc. These established, though constantly evolving, genres have until recently been expressed mainly via print media. The features of Web 2.0 summarized above, while opening up exciting possibilities for novel ways of accessing, creating, and communicating knowledge and information, pose a number of challenges to existing print media as they attempt to come to terms with the new medium and find their place in a communicative environment that has been, and is still being, profoundly reshaped. Though it is difficult as yet to foresee where some of these developments will lead us, an interim assessment of some of the trends may hopefully provide food for discussion and stimulate further research. The coverage of the topic presented here is, perhaps unavoidably, incomplete, given our own particular research interests, and interested readers are invited to contribute to enriching this ongoing debate (see the final paragraph of the present report).

2 Rethinking genre
A dominant trend in recent discourse analysis research of this topic has been investigations into how the Internet is shaping and (re)defining genre when old genres are transferred to the Web or newly-shaped Internet genres appear (Campagna & Garzone, forthcoming; Giltrow & Stein, 2009; Garzone, Poncini & Catenaccio, 2007; Askehave & Nielsen, 2005a, 2005b). The questions raised about this process of 'genre migration' are the following: What happens when a genre migrates to the Web? Will it retain the same properties it possessed in the older medium (the print format) or will it adapt, and if so, how? Does the Web environment generate new, web-native genres? What are the linguistic/rhetorical features of these emerging or adapted genres? Because of the inherent fluidity of the Internet and the pace at which changes are taking place, there are no clear-cut answers to these questions. As Giltrow & Stein point out (2009: 1-2): "The question of genre, then, is really an old one, but the advent

of new media has highlighted the issue with new full force. Seemingly, there are new genres on the Internet, but in some cases it is a matter of contention whether the genre is new, or an old one in new medial garb" - an uncertainty anticipated a decade earlier by Crowston & Williams (2000).

Ongoing research seems to point, however, to the existence of a cline, going from the replication of existing genres at one pole to the emergence of novel genres at the other, along which Internet genres can be (provisionally) placed: at one extreme are genres that have moved intact to the web, as pdf downloads of existing documents, for example. The process of recuperating genres is somehow to be expected since, as Orlikowski & Yates (1994) suggested, when faced with a new communicative situation, individuals will typically draw on their existing genre repertoire, reproducing genres they are already familiar with as members of other communities. Further along the cline one finds other document types which, although they have not undergone major changes in this migration process, have nevertheless accrued considerable 'added-value' (Luzón 2007), or new Internet-enabled features and functions, in the process. The academic research article is a good illustration of this trend: online journal articles now include many interactive tools (e.g. hyperlinks to citations, supplementary material, social networking sites), a large amount of visual and multimodal material (e.g. color photos, video clips, audio summaries), a blog, a space for readers' comments, even online polls to vote on the most favorite article. Yet further along are many genres formed from new combinations, that can be considered either as hybrids or as new web-native genres: blogs (Herring et al., 2005; McNeill 2005), e-zines, homepages, FAQs, hotlists etc. (Crowston 2010). Beyond this are the unclassifiable combinations, leading Santini to argue that "Web pages need a zero-to-multi-genre classification scheme, i.e. a scheme that allows zero genre or multi-genre classification, in addition to the traditional single-genre classification" (Santini 2007:71).

This fast-paced proliferation generated by the communicative needs of new Internet communities entails as yet unresolved implications for genre theory itself. The question that arises is whether traditional genre theory can handle these evolving and novel genres, or whether a 're-purposing' (Askehave & Swales 2001) of genre theory is necessary, that takes the characteristics of the medium into account (Miller & Shepherd 2009). Askehave & Nielsen (2005a), for example, propose an 'up-grade' of the traditional genre model to incorporate media features that cannot be accounted for in existing theory, arguing that the medium itself should form part of a revised genre model, since "the medium adds unique properties to the web genre in terms of production, function and reception which cannot be ignored in the genre characterisation" (2005a: 125).

In addition to this rich debate on genre, other trends that impact on professional discourses as they migrate from print to the web have come to the fore, and are briefly discussed below.

3 Collaborative authoring

Ever since the launch of wiki software by Ward Cunningham in 1994 which enabled the collaborative writing of documents by using a simple markup language, the traditional concept of 'authorship' has been challenged by web-enabled texts. Web users can now upload rather than only download, which means that they can express their opinions, exchange ideas, and produce materials that bypass hierarchical institutions or organizations. A significant repercussion linked to this new medium affordance is that the role of single expert-in-the-field which was in the past clearly assigned to an authority figure/body, is now blurred by several

'experts in the field' taking the floor. The most prominent examples of this trend are participatory journalism (Domingo et al., 2008; Howe 2006), open-source business platforms, online encyclopedias such as Wikipedia (Emigh & Herring 2005), and open science. Open science data, like Wikipedia, are mutable: the documents evolve through a large number of edits made by a worldwide community of scientific contributors. The underlying philosophy of many of these initiatives is to challenge the stranglehold of institutions and corporations on the discourse of the profession: Open Notebook Science (http://usefulchem.wikispaces.com/), for example, provides other researchers with access to failed experiments ('dark data') which are almost never published in academic journals; the Synaptic Leap is a collaborative research project into tropical diseases that are neglected by major pharmaceutical companies because not economically profitable (http://www.thesynapticleap.org). From a discourse analysis angle, however, collaborative authoring raises some thorny questions: how is textual coherence constructed in these multi-authored texts? How do the opposing centrifugal and centripetal forces of discourse (Bakhtin, 1986; Gatto, forthcoming) balance out? Can one talk of an authorial 'voice' anymore? As Yates & Sumner (1997) point out, "the distinction between producers and consumers of digital documents is being blurred and we are seeing the democratization of genre production".

4 Narrativity

Another prominent theme in the analysis of Internet genres is narrativity: not the grand narratives of classical literature but the 'small stories' of private individuals. This is particularly evident in blogs, of course, which abound in stories of daily life (Myers 2010), but is also a pervasive feature of citizen journalism, readers' and users' comments on websites of various kinds, or the Talk pages of wiki sites. This development can be related to the recent shift, in narrative studies, from viewing "narrative as text (i.e., defined on the basis of textual criteria and primarily studied for its textual make-up) to narrative as practice within social interaction" (de Fina & Georgakopoulou 2008: 275). Narrative fulfills this social function on the Internet by helping to create a sense of community between individuals who are geographically far-flung and culturally diverse but who, by recounting and sharing stories, can nevertheless make contact and recognize themselves in one another's stories.

This is a particularly interesting development as far as LSP is concerned: many professional genres are expository, argumentative or informational, rather than narrative; anecdotal evidence and individual stories have little epistemic or institutional weight in these discourse genres. In science, for instance, though narrative may surface in oral genres such as conference presentations (Thompson 2002) or laboratory talk (Lynch 1984), written genres such as the research article construct conceptual arguments rather than give a chronological account of the research process. As demonstrated by Myers (1990), the 'story' told by the research article is a far cry from the 'narrative of nature' recounted in popular science articles. The Web 2.0 platform, however, has given rise to new forms of scholarly exchange such as academic blogs (Kjellberg 2009; Kouper 2010; Bukvova, Kalb & Schoop, 2010), social-cum-scientific networking sites (some hosted by mainstream journals such as the Nature Network by *Nature*, at http://network.nature.com/), forums on open access journals, or wiki sites of user-generated scientific resources (e.g. laboratory protocols on http://openwetware.org/): narrative elements feature prominently in all of these initiatives. It will be interesting to see whether narrative will remain confined to these newer forms of exchange or whether a gradual hybridization of the longer-established genres will take place, and what this will imply for the construction of knowledge itself.

Likewise, blogs on corporate websites and bloggers' contributions to the websites of newspapers or TV channels often have a narrative component. A potential conflict may arise, however, between these unsolicited, individual stories, and the editorial line of the journal or the marketing image a company wishes to project: the centrifugal forces of participatory media may threaten the centripetal forces of institutions and corporations, possibly leading, as in the case documented by Puschmann (2009) to a *fake blog* (or flog) of fabricated narratives on the corporate website as the company seeks to maintain control.

5 Personalization of discourse

Closely related to the theme of narrative is that of the individualization or personalization of discourse on the Internet. Many professional discourses have traditionally been filtered or vetted before publication by gatekeepers of various kinds: the editorial boards of scholarly journals, the company management, or the chief editor of a newspaper, for example. This institutional control fostered the now discredited illusion that such discourses were objective, impersonal representations of reality. While this discourse control by one (or by a selected few) of the many no doubt contributed to the stability of many genres, and potentially also to the disciplinary reliability or validity of the texts, from a critical discourse analysis perspective it appears hegemonic and a negation of empowerment. Computer-mediated communication has changed the rules of the game by complementing one-to-many channels of communication with 'new combinations of N-to-N' (Giltrow & Stein 2009: 9) - many-to-many, some-to-some – opening up spaces for the expression of individual voices. As Borgman (2007) makes clear in the domain of scientific research, however, this development means that material whose quality and legitimacy is difficult to ascertain is now circulating. Another problem generated by these less constrained modes of communication concerns the potential challenge to scholarly and professional norms through more personalized forms of discourse. Self-portrayal and self-expression are recognized features of blogs, which position themselves at different places on the cline between the private and the public (Mortensen & Walker 2002). Many professionals such as journalists or academics (Dennen, 2009; Kirkup, 2010), in addition to their 'professional' voice expressed in news reports or research publications, are increasingly taking advantage of the Web 2.0 platform to express a more private voice in their blogs. What impact will this increased personalization eventually have on the professional discourse of these same authors?

6 Informality of language

As is probably clear from the above, much of the research has approached the field of Internet discourse from a socio-constructivist perspective, foregrounding the contextual features which have engendered these radical changes and privileging discursive practices and function rather than formal, or formalist, aspects (Devitt 2009). In the wake of earlier work on the language of emails, however (Gains, 1999; Lan, 2000), which highlighted the increasing 'conversationalization' of discourse (Fairclough 1995), and the blurring of the writing/speech distinction, some researchers have focused on more specifically linguistic features. A study of the language of blogging, netspeak, chat groups and virtual worlds can be found in Crystal (2006), for example, and that of blogs and wikis in Myers (2010), while Santini (2005), *inter alia*, engages with more technical issues of web language analysis. This remains for the moment, however, a relatively under-researched area.

7 Concluding remark

This short report has left many interesting facets uncovered - issues of identity and reflexivity for example, or multimodality. A forum for further discussion and the presentation of ongoing

research will be provided in Seminar n°1 "From Print to Web 2.0: What future for professional discourses?" to be held at the 11th ESSE (European Society for the Study of English) Conference which will take place at Boğaziçi University, Istanbul, Turkey, from 4th to 8th September 2012. Interested scholars are cordially invited to submit proposals to the convenors, Elizabeth Rowley-Jolivet and Sandra Campagna, before January 31st 2012. Details about the seminar itself, the submission process, and the conference can be found on the ESSE website at http://www.esse2012.org.

8 References

Askehave, I. & Swales, J. M. (2001): Genre identification and communicative purpose: A problem and a possible solution. *Applied Linguistics,* 23(2): 195-212.

Askehave, I & Nielsen A.E. (2005a): Digital genres: A challenge to traditional genre theory. *Information Technology & People,* 18(2): 120-141.

Askehave, I & Nielsen A.E. (2005b): What are the characteristics of digital genres? Genre theory from a multi-modal perspective. *Proceedings of the 38th Hawaii International Conference on System Sciences*.

Bakhtin, M. (1986): *The dialogic imagination*. Austin: University of Texas Press.

Berners-Lee, T. (2006): Interview with S. Laningham, online at: http://www.ibm.com/developerworks/podcast/dwi/cm-int082206txt.html.

Borgman, C.L. (2007): *Scholarship in the digital age: Information, infrastructure, and the internet*. Cambridge, Mass.: MIT Press.

Bukvova, H., Kalb, H. & Schoop, E. (2010): What we blog? A qualitative analysis of researchers' weblogs. *Publishing in the Networked World: Transforming the Nature of Communication*, 14th International Conference on Electronic Publishing 16-18 June 2010, Helsinki, Finland: 89-97. http://dhanken.shh.fi/dspace/bitstream/10227/599/9/8bukvova_kalb_schoop.pdf .

Campagna, S. & Garzone, G. (Eds) (Forthcoming): *Evolving genres in web-mediated communication*. Bern: Peter Lang.

Cardon, D. (2010): *La démocratie internet*. Paris: Le Seuil.

Crowston, K. (2010): Internet genres. *Encyclopedia of Library and Information Science.* http://flosseb.floss.syr.edu/system/files/elischapter.pdf.

Crowston, K & Willians, M. (2000): Reproduced and emergent genres of communication on the world wide web. *The Information Society,* 16(3): 201-215.

Crystal, D. (2006): *Language and the internet*. 2nd ed. Cambridge: Cambridge University Press.

De Fina, A. & Georgakopoulou, A. (Eds) (2008): Narrative analysis in the shift from texts to practices. Special issue of *Text & Talk* 28(3).

Dennen, V. P. (2009): Constructing academic alter-egos: Identity issues in a blog-based community. *Identity in the Information Society*. 2(1): 23-38.

Devitt, A. J. (2009): Re-fusing form in genre study. In J. Giltrow & D. Stein (eds.): *Genres in the Internet* (pp. 27-47). Amsterdam: Benjamins.

Domingo, D., Quandt, T., Heinonen, A., Paulussen, S., Singer, J.B. & Vujnovic, M. (2008): Participatory journalism practices in the media and beyond. *Journalism Practice*, 2(3): 326-342. Special Issue: The Future of Newspapers. http://dx.doi.org/10.1080/17512780802281065.

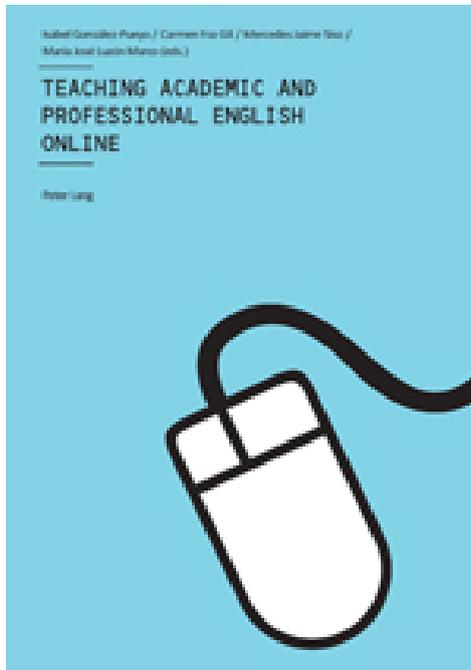Dutton, W. (2007): Through the network (of networks) - The fifth estate. Oxford Internet Institute.

Emigh, W. & Herring, S. (2005): Collaborative authoring on the web: A genre analysis of online encyclopedias. *Proceedings of the 38th Hawaii International Conference on System Sciences.*

Fairclough, N. (1995): *Critical discourse analysis*. London: Longman.

Friedman, T.L. (2007): *The world is flat, 3.0. A brief history of the twenty-first century*. Picador.

Gains, J. (1999): Electronic mail – A new style of communication or just a new medium? An investigation into the text features of e-mail. *English for Specific Purposes,* 18(1): 81-101.

Garzone, G., Poncini, G & Catenaccio, P. (Eds) (2007): *Multimodality in corporate communication: Web genres and discursive identity*. Milano: FrancoAngeli.

Gatto, M. (Forthcoming): Centrifugal and centripetal forces in web 2.0 genres: The case of wikipedia. In S. Campagna & G. Garzone (eds.): *Evolving genres in web-mediated communication*. Bern: Peter Lang.

Giltrow, J. & Stein, D. (Eds) (2009): *Genres in the internet. Issues in the theory of genre*. Amsterdam: John Benjamins.

Goffman, E. (1981): *Forms of talk*. Oxford: Blackwell.

Herring, S., Scheidt, L. A., Wright E. & Bonus, S. (2005): Weblogs as a bridging genre. *Information Technology & People*, 18(2): 142-171.

Howe, J. (2006): The rise of crowdsourcing. *Wired*. http://www.wired.com/wired/archive/14.06/crowds.html .

Kirkup, G. (2010): Academic blogging: Academic practice and academic identity. *London Review of Education*. 8(1): 75-84.

Kjellberg, S. (2009): Scholarly blogging practice as situated genre: An analytical framework based on genre theory. *Information Research*. 14(3): Paper 410. http://InformationR.net/ir/14-3/paper410.html .

Kouper, I. (2010): Science blogs and public engagement with science: Practices, challenges, and opportunities. *Journal of Science Communication*, 9(1).

Lan, L. (2000): Email: A challenge to standard English? *English Today* 16(4): 23-29.

Luzón, M. J. (2007): The added value features of online scholarly journals. *Journal of Technical Writing and Communication,* 37(1): 59-73.

Lynch, M. (1984): *Art and artefact in laboratory science: A study of shop work and shop talk in a research laboratory*. London: Routledge.

McLuhan, M. (1964): *Understanding media*. Cambridge, Mass.: MIT Press.

McNeill, L. (2005): Genre under construction: The diary on the internet. *Language@Internet* Vol. 2. http://www.languageatinternet.org/articles/2005/120 .

Miller, C.R. & Shepherd, D. (2009): Questions for genre theory from the blogosphere. In J. Giltrow & D. Stein (eds.): *Genres in the internet. Issues in the theory of genre* (pp. 263-290). Amsterdam: John Benjamins.

Mortensen, T. & Walker, J. (2002): Blogging thoughts: personal publication as an online research tool. In A. Morrison (ed.) *Researching ICTs in context* (pp. 249-279). InterMedia: University of Oslo.

Myers, G. (1990): The social construction of popular science: The narrative of science and the narrative of nature. In *Writing biology: Texts in the social construction of scientific knowledge* (pp. 141-192). Madison: University of Wisconsin Press.

Myers, G. (2010): *The discourse of blogs and wikis*. London: Continuum.

Orlikowski, W. J. & Yates, J. (1994): Genre repertoire: The structuring of communicative practices in organizations. *Administrative Sciences Quarterly,* 33: 541-574.

Puschmann, C. (2009): Lies at Wal-Mart. Style and the subversion of genre in the *Life at Wal-Mart* blog. In J. Giltrow & D. Stein (eds.): *Genres in the internet* (pp. 49-84). Amsterdam: Benjamins.

Santini, M. (2005): Web pages, text types, and linguistic features: Some issues. *ICAME Journal No. 30.*

Santini, M. (2007): Characterizing genres of web pages: Genre hybridism and individualization. *Proceedings of the 40th Hawaii International Conference on System Sciences.*

Thompson, S. (2002): 'As the story unfolds': The uses of narrative in research presentations. In E. Ventola, C. Shalom & S. Thompson (eds.): *The language of conferencing* (pp. 147-168). Frankfurt: Peter Lang.

Yates, S. J. & Sumner, T. R. (1997): Digital genres and the new burden of fixity. *Proceedings of the 30th Hawaiian International Conference on System Sciences.* VI, 3-12.

***

Book Review

Reviewed by:
Associate Professor Lisbet Pals Svendsen
Educational Research Group,
Dept. of International Language Studies and Computational Linguistics,
Copenhagen Business School, Denmark.
lps.isv@cbs.dk

In today's university setting where professional ESP/EAP competencies are considered one of the keys to success in the business community upon graduation, it is only natural that didactic practices are expanded to comprise the growing range of ICT tools that are being offered in support of language, culture and communication teaching and learning.

This book, edited by Isabel Gonzáles-Pueyo, Carmen Foz Gil, Mercedes Jaime Siso and María José Luzón Mardo and published by Peter Lang, addresses a range of interesting and relevant issues in connection with the use of ICT for the professionalizing and academization of the English skills of adult language learners.

In the words of the editors, the book aims to offer an empirical rather than theoretical perspective on how ICT tools may be put to good use in connection with efforts to professionalize learners' competences, and the book offers an interesting variety of contributions, ranging from those that address the more general considerations involved when the question of applying ICT in the ESP/EAP classroom comes up to those that discuss specific initiatives taken to promote language learning within a specific topic area.

The nine articles in the book are ordered in a logical way that guides the reader through the topic from the more general to the very specific exemplifications of the use of ICT in the ESP/EAP classroom, including lessons learned in the process and examples of how the application of ICT tools played out in practice.

From the point of view of this reviewer, the volume offers a solid insight into the considerations that will have to be made by any teacher contemplating the introduction and use of ICT in his/her teaching activities, just as it contributes the experiences from courses where different ICT tools have been used with specific aims and reports the outcomes of these didactic experiments. Interestingly enough, none of the contributors seem to suggest that ICT will be replacing face-to-face interaction in a classroom setting anytime soon or in a more distant future, but they all – reassuringly enough – tell the story of how the proper mix between face-to-face and ICT contribute to enhancing learning outcomes for the participating students. Which is, after all, the whole point!

***

Call for papers

The Editorial Board of the international peer reviewed LSP Journal is currently considering papers for inclusion in forthcoming 2012-2013 issues.

The LSP Journal is dedicated to publishing high-quality contemporary research into language for special purposes (LSP), professional communication, knowledge management and cognition in theory and practice.

Focus is on interlinking research across these fields which are of vital importance to the development, exchange and acceptance of new ideas and products in scientific domains as well as in trade and public services at national and international levels.

Operational and innovative methods and practices of relevance to creation and dissemination of knowledge in intralingual, interlingual and intercultural environments are key focal points.

Researchers are given a peer-reviewed (2 reviewers) international forum for exchanges of new insights into theoretical and practical approaches.

The LSP Journal is a respected electronic journal with a worldwide audience and a history going back to 1977, as it is the successor of *Unesco Alsed-LSP Newsletter* (1977-2000) and *LSP & Professional Communication* (2001-2008).

Potential topics
The journal welcomes contributions from around the world and calls in particular for trans-disciplinary innovative approaches which cover but are not limited to topics such as:

- LSP , ESP, Terminology, Professional communication, knowledge management, and cognition
- Trans-disciplinary approaches linking e.g. knowledge dissemination, crowd sourcing and quality assessment
- LSP in monolingual and multilingual contexts
- Cutting edge research in emerging fields of relevance to LSP, e.g. the semantic web
- Case studies involving traditional LSP research disciplines including translation
- Conceptual modeling and ontology building in theory and practice.

The contributions may be in English, French, German or Spanish.

Submission:
Articles should be submitted in Microsoft Word format (.doc or .docx). Submissions should be made using the online submission and peer review system (Open Journal System) on the LSP journals website **http://lsp.cbs.dk** where authors will be guided through the submission process.

Colophon

Editor

The Journal is hosted by:

Web access to the electronic journal: http://lsp.cbs.dk

Indexed in:

Contact: