User-Sensitive Lexical Databases: A case of lexical knowledge management

Virpi KALLIOKUUSI and Krista VARANTOLA, Helsinki and Tampere, Finland

Abstract

This article focuses on the user-friendliness of lexical information sources. Whereas our previous study on user-friendliness (Euralex 1998) emphasized the context-sensitive needs of dictionary users, our present study goes one step further and suggests that it would be possible to compile interactive lexical databases that would be both context- and user-sensitive. We approach the function of lexical databases from two perspectives: from their role as primary information sources and from their role as lexical interfaces to other knowledge bases. Our approach is generally based on frame-semantics. We apply semantic frames to capture the different ways of conceptualization used when searching a knowledge base for social and health care services.

1 Introduction

This study focuses on the user-friendliness of lexical information sources. In a previous article on user expectations, we argued that users of lexical reference sources often feel frustrated because they have unrealistic expectations about the sources and their adaptabilility to individual needs. Whereas users normally have context-sensitive, specific questions, dictionary makers try to formulate the information in a lexical entry so that the answers are valid irrespective of the context that the individual user has in mind. We suggested that, in the electronic age, dictionary makers could pay more attention to user needs and even re-evaluate some editorial principles, be less orthodox and more pragmatic in their approaches to dictionary and glossary compilation (Kalliokuusi and Varantola 1998).

More concretely, we encouraged dictionary makers to simulate user needs in order to provide more specific answers to user questions. This type of approach would necessitate the inclusion of context-sensitive, extralinguistic and culture-specific information in the entry that could be made available to the user on request, maybe with a little prompting from the dictionary maker. The entries would then have a hypertext structure and links to other types of textual information sources, such as tailored corpus collections or encyclopedic knowledge bases.

In this article, we take our position one step further and envisage lexical databases that would be both context- and user-sensitive. In fact, we would like our dictionaries or glossaries to become so user-friendly and interactive that they would guide the users to the information and the advice they need, even when the users do not really know how to ask for it.

2 Lexical databases and their users

Users need lexical information for a number of purposes. They may try to find a suitable equivalent for a lexical expression in another language or they may be looking for information that

is related to phraseological units, grammar, or longer passages in a text. Users frequently also resort to dictionaries to check their hunches and to find confirmation for their tentative choices.

In addition to their traditional functions, lexical databases also serve as primary user interfaces to other types of knowledge bases. The indexes in reference works, such as manuals, text books, on-line help options, function as the first user interface to the specific information the user wants to find in the reference work. With electronic reference databases, the role of accurate search words becomes even more pronounced. When retrieving information, the user does not want to be flooded by a list of hits or references in which the relevant ones are hidden among a large number of totally irrelevant pointers. The user becomes frustrated by the information overload and the time spent on sifting the desired from the undesired information.

On the other hand, when using electronic indexes to find the desired information, users may not find a relevant passage because they do not know the correct search words. Most of us have undoubtedly encountered this problem when trying to make our word processor do what we would like it to do. We end up doing fuzzy searches with a number of right-sounding keywords, which the manual writers, however, have not provided as search or index words or have used in a different sense than we have. We can thus claim that the electronic format has added information overload as a new search problem to cope with, without taking away the old problems of getting too little information, or of not accessing the relevant information to solve an extralinguistic, context-sensitive problem [1].

3 Lexical databases in a larger communicative context

We can think of lexical databases as communication products which aim to provide relevant information to their users. The information need is defined by the user, but the user is often unable to express his or her exact need in a precise manner or in accurate terms. It would therefore be beneficial if the product could help and guide the user during the search process. The users would thus profit from lexical databases which would become active mediators of information and adjust to the users' needs instead of being primarily passive reference sources.

The purpose of lexical databases could then be compared to that of technical communication. Technical communication can be defined "as transferring knowledge from those who know to those who need to know" (cf. Carliner 1999). Dictionaries and glossaries serve the same purpose. Users consult dictionaries to get information on the form, meaning and use of words and expressions, and this information becomes knowledge when users apply the acquired information in an appropriate context.

The electronic medium has changed the way in which this information can be offered to the user and this leads us to knowledge management, a key issue in the information age. Knowledge management, in turn, can be defined as access to the information you need to have (Carliner 1999:85). If we elaborate this idea further, we can claim that, in the present phase of our information society, it is crucial to discuss how to ensure that the transformation of information into knowledge is smooth. Information technology has greatly facilitated the collection of data, its analysis and availability as structured information but it has not yet solved the problem of how to offer this information in a manageable and relevant way to the human users. In addition,

individuals have a wide range of information and application needs but also very different levels of background knowledge to anchor the new information to.

4 Lexical knowledge management

4.1 Present-day problems in acquiring user-friendly lexical information

Dictionary makers and terminologists face major tasks when they try to customize lexical entry information to suit individual competencies. It is not easy to incorporate tailor-made, user-friendly information in a dictionary or terminological glossary. Kalliokuusi and Varantola point out that "the editor of a specialized glossary, working according to terminological principles and methods, aims at presenting knowledge structures of a certain subject field in such a way that they are not bound to any particular context". We also maintain that "the information given in the definition is not intended to solve problems related to a particular linguistic or textual context. Consequently, no information is given, for example, on collocational probabilities, derivational characteristics or inflection" (Kalliokuusi & Varantola 1998). And, yet, users "try to find non-dictionary type information in dictionaries because it is not readily and systematically available in other sources" (Varantola 1998).

Translators expect terminological glossaries to contain terms and concepts that are less central to the field, whereas terminologists usually prefer to include only the central concepts and their definitions. Glossary users are often dissatisfied with the terminologically correct definitions of concepts because they are too general or too difficult to understand. Users also complain that the definitions do not have enough contextual, real-world information and that different user groups use the terms for different purposes. An additional problem is that, in some fields, many terms look like general words but are actually employed in a narrower and more restricted sense used in a special field.

Does this mean that compiling a user-friendly lexical database is an impossible undertaking? In the following, we try to suggest some ways in which user-specified needs could be taken into account in a lexical database.

4.2 Interactive lexical databases as solutions

Interactive databases are already everyday applications in restricted domains, such as in banking or on-line shopping. Cash dispensers or automatic ticket machines at railway stations are concrete examples of such applications. They guide and navigate the user through the transaction or operation and list the different options available at any decision point during the operation. It is also possible to envisage this type of interactivity for lexical databases. A glossary or an index could navigate the user through the information that is available in the glossary database. It would thus alert the users to the alternative routes that the search can take and to the decisions they must make in order to refine the search. The glossary could also monitor user performance and give advice in the manner of the Microsoft "paper clip" but in a less intrusive fashion (cf. Carliner 1999:90 who says that improvements are being planned for future versions of the word processor).

To be helpful, a performance monitor in a lexical database would need to have respect for the user and the various unpredictable reasons why the user is following a particular search strategy. Furthermore, the actual decision making should be placed firmly in the hands of the user and the monitoring system should be discreet in its advice.Yet, it could highlight the alternative search routes for the user and give vital information or advice that the user has not thought of asking.

5 A possible model for an interactive lexical database

Our ideas are based on a terminology project in the field of social and health care services. During the project, comments were elicited from representatives of two interest groups, professionals in social services and in health care services. The social and health care sectors in Finland have a pressing need for systematic terminology work. Centrally-managed, seamless service chains are envisaged in order to cut down public sector costs and to produce more client-oriented services. However, the effective management of service processes and service chains cannot be achieved without developing information systems which are harmonized and integrated.

The development of these integrated information systems and service databases has led to an increased interest in terminology work and especially in the systematic organization of concept-related information. Furthermore, new terminological solutions are needed to support the joint information systems.

When designing special field information systems we have to study the nature of the terminology used in the fields. The terminology of social and health services is both special and general in character. Many terms are easily recognizable as special terms (e.g. *social work, nursing care, surgical procedure, inpatient service, outpatient service, referral*), whereas others are common also in general language (e.g. *service, service process, service plan, service provider, client, patient*). Subject specialists, however, use the terminology in a "technical" sense, some of it in many technical senses, depending on the domain and their specialization.

The clients of social and health services are also a potential user group of the future information systems. For them it would be natural to give the "general-language" terms a general-language interpretation. This could be very confusing and result in serious misunderstandings. Misunderstandings are already common when specialists communicate with their clients/patients because of divergent interpretations of the meanings of concepts. Systematic efforts in terminology work are therefore necessary, in general, to bridge the gap between specialist and non-specialist understanding, i.e. to , i.e. to "transfer knowledge from those who know to those who need to know" (see above).

The comments received during the terminology project, referred to above, clearly indicate the need for novel lexical and terminological solutions.These glossaries were compiled in a traditional fashion to facilitate the use of the integrated information systems. Before their release, social service and health care professionals were asked to comment on the usability and intelligibility of the definitions and explanations used in the entries.The examples given below illustrate some concerns the reviewers had about the methodology:

- The concepts are often defined only from the service provider’s point of view. In the

glossary, this results in an approach that is far too organization-oriented, whereas the actual services provided are based on client-oriented work methods.

- Approaches typical to specialized health care tend to dominate in the glossary descriptions. The definitions are thus too narrow in scope and basic health care services and social services receive too little attention. Problems arise when concepts common to various subdomains are defined on the basis of one specific domain only,(e.g. when *referral* or *care report* are defined only in relation to medical services or just in relation to clinical or surgical procedures).

- Concept interpretations tend to be based on the current legislation. This often results in delineations of concepts that are too narrow in scope and do not take into account the wider applicability of the concepts in social and health services. The reason is that, in the legislation, concepts related, e.g., to patient rights, fields of specialized health care or social and health service organizations are defined by prototypical examples instead of exhaustive lists of real life solutions or phenomena.

- The glossaries are not part of the information systems. This means that professionals are not able to employ glossary information in an efficient way in their decision making. The demands for multi-professional teamwork and cross-sectional communication impose special requirements on the information systems and necessitate the inclusion of terminological information in the databases.

- Information about social and health services should be accessible also to clients, and the information systems should contain databases that are also open to the users of the services. Clients should be able to use the databases for searching information on available services (e.g. hospitals, day care centres and other institutions).

5.1 The philosophy behind our model

We can imagine that the potential users of social service and health care databases would approach the information they need from at least three different directions: either from the specialist directions of a social service specialist, or health care specialist or from a non-specialist lay direction. To access the information, the users would first enter their search word(s) to the system at the lexical or terminological "index" interface. When choosing their search word(s), the users would either explicitly or implicitly be thinking of a particular question that they have conceptualized in their minds. These conceptualizations would depend on the background knowledge of the users, and the non-specialist users would probably have a much looser way of conceptualizing their questions than would the expert users.

This lexical and terminological interface would, in its turn, try to determine the exact purpose of the user’s question by trying to narrow down the number of possible answers by means of specifying questions to the the user about his or her information need. The purpose of these questions would be to guide the users and help them decide which of the available search routes they should to take. The idea is that, instead of giving the user a large number of possibly relevant answers, the interface would, at this stage, try to eliminate the irrelevant answers and routes. The original search words would probably often point to several concept entries and

concept systems in the database. The intermediate questions, on the other hand, would act as a buffer that would prevent the users from being confused or frustrated by a multitude of opaque hits that they would need to investigate before they could determine their relevancy.

How could these intermediate questions be generated? The search words can vary and be very inaccurate in a strictly terminological sense. The interface would have to recognize a continuum of search words as potential keywords for one or several entries in the system. We would therefore need to predict what search words or combinations of search words different types of users might use.

A frame semantics-oriented approach could help in envisaging the structure of this buffer interface. The authors of the FrameNet project description state[2] that,

> "...in order to understand the meanings of the words in language we must first have knowledge of the conceptul structures, or semantic frames, which provide the background and motivation for their existence in the language and for their use in discourse". [...]
>
> "In developing a frame-semantic description, we must first identify the phenomena, experiences, or scenarios represented by the meanings of the target words and the sentences in which they occur."

Later on in the description, the authors use the "Commercial Transaction Frame" as an example of their line of reasoning and point out that it consists of a "richly structured collection of concepts" (possession, change of possesion, exchange, money), which are described by means of a wide range of words, such as *buy, sell, pay, spend, price, change, debt, credit, merchant, broker, shop*, etc. (see also, Fillmore and Atkins 1998).

5.2 Description of the model

In our terminological context, we need a slightly modified version of a semantic frame that will better suit our purposes. We therefore suggest the notion of fuzzy general-language frames that can be made more accurate by gradually shrinking them to the size of terminological frames. This gradual shrinking would take place at the buffer interface and be accomplished by the interactive questions. The buffer questions would both narrow down inaccurate general language search words to terminologically proper expressions and also help to determine the various special senses in which the terminologically correct search words are employed by the users. In the latter case, the buffer questions would refine and redefine the query by presenting the user with a set of possible terminological frames to choose from. For the user, this process would resemble a guided tour through the database, all the way to the intended destination. This search model could also be described as a do-it-yourself index to a reference source, or as a sophisticated, interactive lexical map to the addresses in the database.

The information system in our proposed model would consist of several databases with clearly defined user levels. It would be a decision-making tool for the social and health care professionals, but it could also serve as a database for clients needing information about social and health services.

The information system would display its data at least on three levels of knowledge:

- Level one: a general instruction base for the non-specialist, with general-language semantic frames.
- Level two: the shared knowledge bases of the two special fields, with shared terminological concept frames.
- Level three: separate knowledge bases for each field; one for social services (including social security) and one for health services (including basic and specialized health care), with separate terminological concept frames.

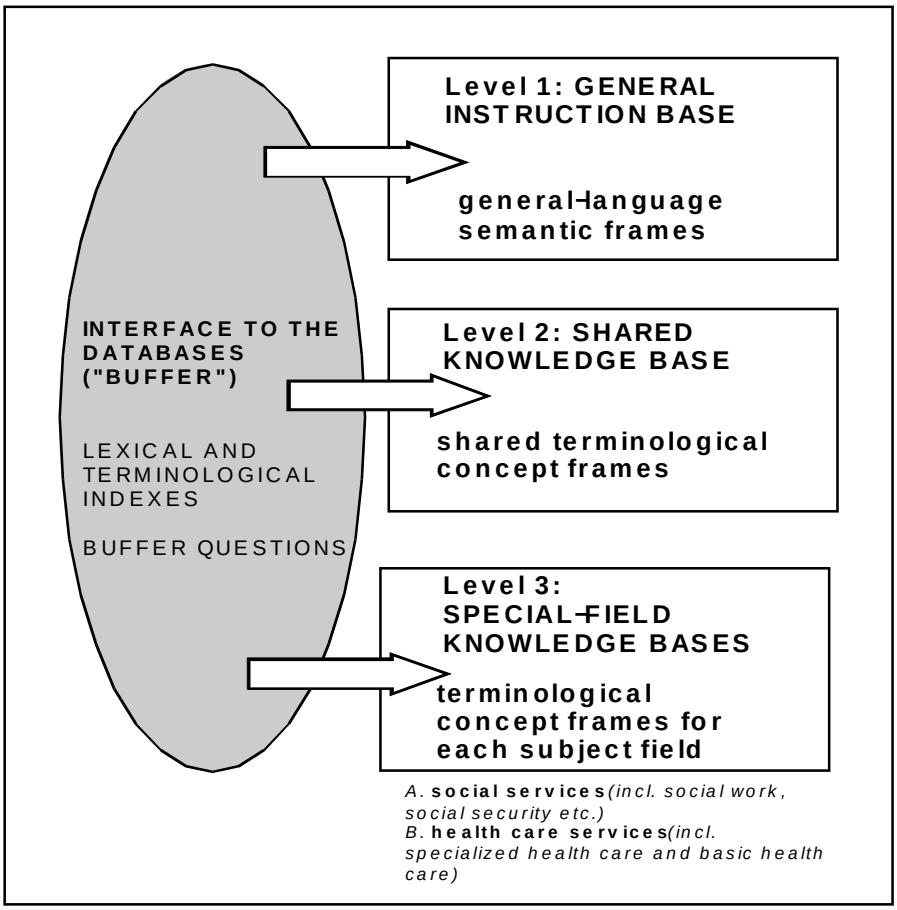

Figure 1: Display of data at three levels of knowledge

Non-specialist users. The aim of the first-level instruction base would be to provide additional information to the non-specialist users. These users would then be able to enter the information system by using context-sensitive search words that would correspond to their experience and conceptualizations. The first-level instruction base, however, would help the users to answer the buffer questions and conceptualize their knowledge in a more specialist-like fashion so that it would be easier for the non-specialist to find the "right" information in the special-field knowledge bases. In other words, the instruction base would help the users to anchor their fuzzy

"general description" search words and semantic frames to the terminologically more stringent concept frames and conceptulizations in the specialist knowledge bases.

A client who needs information on benefits related to the incapacity for work or unemployment could enter the system by using such general-language search words as *assistance, unable to work, out of work, unemployed, unemployment* without knowing the exact terms for the various allowances and subsidies available, such as *unemployment allowance, unemployment pension, unemployment benefit, unemployment supplement, employment pension, invalidity pension*. Likewise, a person needing medical treatment for a toothache could enter the system by giving general-language search words, such as *dental appointment, pain in tooth ,doctor, dentist, inflammation of teeth, toothache* without having to define the exact problem (e.g. *gingivitis,inflammation of the gum, caries, tooth decay*), or the exact health-service provider needed.

The general-language search words would be placed in the appropriate semantic frames in the system, and the frames would, in turn, trigger further specifying questions when necessary. These questions would gradually guide the user towards the "right" information. The benefit of this interactive and instructive interface would be that specialized information could be retrieved through fuzzy searches.

Specialist users. The professional users of the information system would naturally have a more direct access to the information they need because they would be able to use more accurate search terms leading them directly to the possible concept frames. They could then choose the intended frame which would take them to the right location in the specialist core knowledge bases. The knowledge structures in the specialist bases would correspond to the terminologically analyzed and structured concept systems and systematically drafted definitions. The database structure would, however, enable the collection and presentation of additional information about the concepts. The traditionally "context-free" definitions of concepts, such as *specialized health care, health service provider, care episode* or*care plan* could be accompanied by extra information about how these concepts are treated in legislation, in statistics or in administration, for example.

An important feature of the information system would be information about concepts that are common to several fields, in this case to both social and health services. Concepts, such as *client, service, consultation, service order, referral* etc., are basically identical in meaning for both fields, but have in practice acquired certain "sub-technical" interpretations and associations which are different in the two fields. It would be possible to include and present these "associated" meanings in the specialist knowledge bases so that social and health service professionals would be able to benefit also from this type of information in their decision making. Working as multi-professional teams would be more efficient if specialists had access to interpretations of other experts in other fields and with different background knowledge.

Notes

[1]Cf. e.g. Kohonen on self-organizing maps and data mining, `http://www.cis.hut.fi` and `http:// websom.hut.fi`, Nielsen 1993 or Schriver 1997 for discussions on information overload and on the problems of accessing relevant information.

[2]The FrameNet Project. What is Frame Semantics?
`http://www.icsi.berkeley.edu/~framenet`.

References

Carliner, Saul. (1999). Knowledge Management, Intellectual Capital, and Technical Communication. In *Communication Jazz: Improvising The New International Communication Culture*. Proceedings 1999 IEEE International Professional Communication Conference. New Orleans, September 1999. 85-91.

Fillmore, Charles J. and B.T.S. Atkins.(1998). FrameNet and Lexicographic Relevance. In print.

Kalliokuusi, V. and Varantola, K. (1998). From general dictionaries to terminological glossaries. User expectations vs editorial aims. In Thierry Fontenelle, Philippe Hiligsmann, Archipald Michiels, André Moulin, Siegfried Theissen (Eds*.) Euralex '98 Proceedings.* Papers submitted to the Eighth Euralex International Congress on Lexicography in Liège, Belgium. Liège: Université de Liège. 600-610.

Kalliokuusi, Virpi (1998). Från ordlisteredigerarens principer till användarens förväntningar. In Henning Bergenholtz, Sven-GöranMalmgren (Eds.) *LexicoNordica 5*. Oslo: Nordisk förening för lexikografi, 25-37.

Laurén, Christer, Myking, Johan, Picht, Heribert (1997). *Terminologi som vetenskapsgren.* Studentlitteratur. Lund.

Myking, Johan, Saebøe, Randi, Toft, Bertha (ed.) (1996). *Terminologi ? system og kontekst.* Nordisk minisymposium 1996. KULTs skriftserie nr. 71. Oslo.

Nielsen, Jacob. (1993). *Usability Engineering.* AP Professional, Academic Press. Boston, San Diego, etc.

Pearson, Jennifer (1998). *Terms in Context. Studies in Corpus Linguistics.* John Benjamins Publishing Company. Amsterdam, Philadelphia.

Schriver, Karen A. (1997). Dynamics in Document Design. John Wiley & Sons, Inc. New York, Chichester, etc.

Varantola, K. (1998). "Translators and their Use of Dictionaries". User needs and user habits. In Atkins, B.T.S. Ed. *Using Dictionaries.* Studies of Dictionary Use by Language learners and Translators. Niemeyer. 179-192.

Varantola, K. (1998). Translators, dictionaries and text corpora. In Proceedings of the Workshop "*Corpus use and learning to translate*". Forli/ Bertinoro, 14-15.11. 1997. Available under: `http://www.sslmit.unibo.it/cultpaps`

WEBSOM. The WEBSOM home page on the Internet. `http://www.cis.hut.fi`