Design, Development and Compilation of a bilingual multifunctional Intranet-based Differential Telecom Lexinome at a major Danish Telecoms Group

Henrik KØHLER SIMONSEN, Århus, Denmark

Abstract

This paper discusses the conceptual considerations for the design and development of an Intranet-based lexicographic knowledge management system at Tele Danmark (TD). The *Convergence Telecom Lexinome* (CTL) is thus the result of the interplay between theory and practice. The design and development of CTL are based on an analysis of the communicative environment and of end-users' skills and competencies. On the basis of this analysis, the CTL functions are identified, and the conclusion is drawn that multiple lexicographic and non-lexicographic data sources must be integrated and networked to solve the lexicographic problem identified. Finally, this paper outlines the theoretical framework of the system, which is based on an integrated database and Internet platform converging a multitude of lexicographic and non-lexicographic data sources.

1. Introduction

This R&D project is based on an official three-party agreement between Tele Danmark A/S[1] (TD), the Aarhus School of Business[2] (ASB), and the Academy of Technical Sciences[3] (ATV). The project is part of the Danish Industrial Ph.D. Fellowship Programme, which is designed to facilitate knowledge transfer between Danish universities and the corporate sector and stimulate and promote research and development in the corporate sector. The Danish Industrial Ph.D. Fellowship Programme is thus a bipartite three-year programme consisting of a development task at TD and a theoretical research education leading to the Ph.D. degree.

2. Problem Identification

TD is the result of a relatively recent merger of five former regional telephone companies. Despite the fact that Denmark is a fairly small country, there are considerable regional differences. Each regional telephone company thus had its own set of working procedures and its own "corporate culture" when Tele Danmark A/S was formed in 1996. In addition to regional differences, the five regional telephone companies had a monopoly in the Danish telecommunications market, and these five companies had to go through a hard adjustment process from being public companies to forming a single private company. The privatization process was completed when US Ameritech bought the remaining 42% of the shares from the Danish State. Consequently, in the wake of its formation, the new private company was facing a wide variety of difficulties while at the same time trying to operate in a highly competitive international market. Almost simultaneously TD implemented a nation-wide harmonization process of the working procedures, an effective privatization campaign and entered into a strategic partnership with Ameritech.

All these factors naturally resulted in a number of problems for TD. One of these problems was the terminological and linguistic black hole, which characterized TD. The problem was further aggravated by the fact that on the one hand TD has a high-profile international image and is an important player in the international telecommunications markets, but on the other hand TD has neither a lexicographic knowledge management tool, nor any formal language policy. The multi-faceted lexicographic problem is:

1. TD is in urgent need of a system that facilitates harmonization and standardization of terminology used both internally and externally.

2. TD is in urgent need of a system that covers virtually all areas (due to the complexity and multitude of the products and services provided) and all types of information (both encyclopedic and linguistic) in as many languages as possible (at least Danish and English).

3. Analysis of Possible Project Solutions

To find a solution to the lexicographic problem outlined above, a total of ten potential project solutions were analyzed and studied. The ten project solutions under scrutiny were:

1. Outsourcing of all text production to external translation firms
2. Establishment of implants of external translation firms at TD
3. Training and appointment of a large number of linguistic experts
4. Training and appointment of a central communications department
5. Acquisition of a large number of paper-based and/or electronic dictionaries
6. Acquisition of computer-aided translation memory tools
7. Utilization of existing Internet dictionaries, databases, glossaries, and term lists
8. Utilization of free machine translation services on the Internet
9. Comprehensive linguistic training of all employees
10. Design and development of a custom-built lexicographic knowledge management system accessible from TD's Intranet.

The analysis, which included empirical studies, practical tests, and introspection, first of all revealed that there was much to learn from already existing solutions. However, the analysis also showed that solution number ten would be by far the optimum solution for TD. The decision to focus on some kind of Intranet-based solution was based on three arguments. First of all, a custom-built lexicographic knowledge management system would be based on the actual needs of the intended users, and the system would thus have a *"genuiner Zweck"* as stated by [Wiegand 1988] and further discussed by [Bergenholtz/Tarp 1995]. Second, a custom-built system would be able to fulfil the widely varying demands of the intended user group, or in other words fulfil a number of *"dictionary functions"* as discussed by for example [Bergenholtz/Tarp 1995]. Third, a custom-built system would fit perfectly into TD's Intranet allowing the telecommunications group's approx. 17,000 employees from numerous geographical locations in Denmark and abroad to access information from the system, and at the same time the system would be future-oriented.

3.1. Empirical Analysis of Internet Reference Works

On the basis of the above arguments it was decided that an empirical analysis of various Internet reference works[4] would be beneficial. The purpose of the analysis was thus partly to learn from already existing lexicographic systems on the Internet and partly to find out whether Internet reference works could be the solution, or perhaps part of the solution, to the lexicographic problem at TD.

The Internet is a dynamic and anarchist medium, and is therefore not easy to measure or quantify. However, the analysis nevertheless gives a quite clear picture of the quality of the Internet reference works at the time of the survey, and even though the empirical analysis only yields exemplary results, these results seem to be scientifically valid. This analysis backs the conclusion arrived at in a survey conducted by [Storrer/Freese 1996]. The analysis was conducted in the fall of 1999, and the empirical basis of the analysis included 151 Internet reference works, distributed on 117 LSP Internet reference works and 44 LGP Internet reference works. A number of measuring points (lexicographic elements) were identified in advance, and an overall typology following the typology in HSK 5.1-5.3 was used. Finally, specific selection criteria were defined in order to ensure the relative exemplarity of the data. The results of the analysis are shown below in figures 1 and 2.
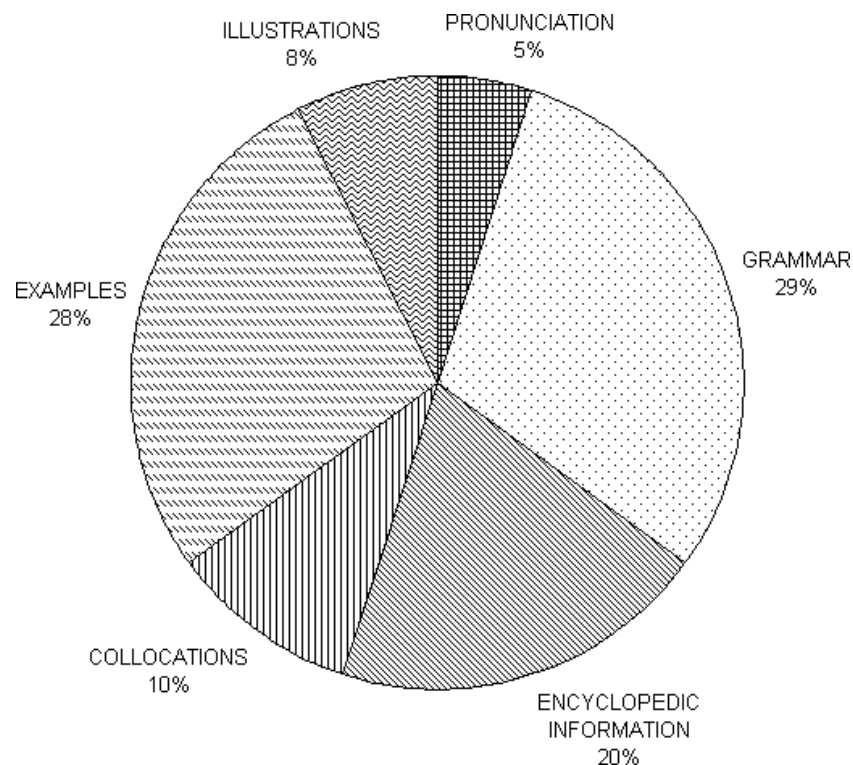

Figure 1: Lexicographic elements - LGP

PRONUNCIATION
2%

ILLUSTRATIONS
13%

GRAMMAR
2%

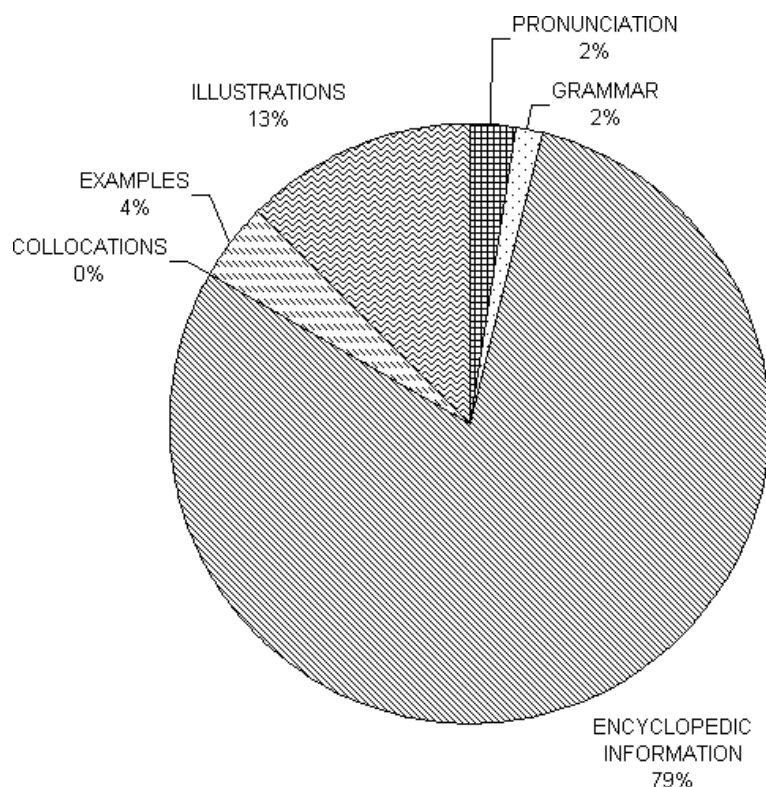EXAMPLES
4%

COLLOCATIONS
0%

ENCYCLOPEDIC
INFORMATION
79%

Figure 2: Lexicographic elements - LSP

The two figures show the distribution of selected lexicographic elements in the two classes of Internet reference works. When taking a closer look at the individual lexicographic elements, it is particularly interesting to see that the distribution of lexicographic elements in the two overall classes of Internet reference works is dramatically different. On the basis of this empirical analysis it can be concluded that the majority of LSP reference works do not contain a satisfactory amount of lexicographic elements. The LSP reference works in this survey seem to focus almost solely on offering encyclopedic information to the users, even though the use of illustrations seems to be more typical in LSP reference works than in LGP reference works. It is very surprising to see that linguistic information is almost non-existent in LSP reference works, for example the almost negligible amount of examples and collocations, which are all-important and decisive for successful text production in L2 in an LSP context.

Consequently, there is much work to be done in this area. The quality of LSP Internet reference works clearly needs to be improved before they can be used in a professional context at TD. The conclusion of this empirical analysis thus resembles the conclusion made by [Storrer/Freese 1996] who state that *"Wer jedoch glaubt, ein Internet-Anschluss mache die Anschaffung eines deutsch-englishen Wörterbuches überflüssig, wird bei näherer Bekanntschaft mit den verfügbaren Online-Nachschlagewerken für Deutsch-English bald gerne zum gedruckten oder Off-line Wörterbuch zurückkehren"*. The two analyses thus reveal that lexicographers have not yet realized the full potential of the Internet. The typical approach still seems to be to generate mere naked word solutions and then upload these wordlists to the Internet. This is not in any way satisfactory from a lexicographic point of view because such a solution does not at all

match the user requirements to a professional translation and text production tool. In conclusion, the R&D project will attempt to remedy the situation at TD by designing a lexicographic system that contains a satisfactory amount of encyclopedic and linguistic information. However, before outlining the conceptual considerations for the CTL, it is first necessary to discuss some results from the analysis of the intended CTL users and not least the *"dictionary functions"*, as discussed by [Tarp 1995].

4. CTL Users and Functions

It is universally accepted that the degree of success of any product is heavily dependent on the underlying end-user research conducted. A clear picture of the intended end-user thus plays a vital role for the success of any product or service, and the qualifications and situation of the end-user govern any production process. This is also the case for this R&D project where the paramount factor is the requirement profile of the CTL end-users. The purpose of the analysis conducted was thus to obtain a clear picture of the intended end-users (CTL users) and to identify the different situations in which the end-users intend to use the CTL (CTL functions). The objective of the analysis was to obtain sufficient information on the CTL users and functions so that the ultimate version of the CTL would match the user requirements to the lexicographic knowledge management system and thus to design and build a user-oriented tool that would solve the lexicographic problem at TD. The research design model used in the analysis included introspection, existing theories and practical experience, and an Intranet-based questionnaire form, which was particularly useful in the somewhat isolated lexicographic environment at TD.

The analysis, which focused on specially selected measuring points as discussed by, for example [Tarp 1999], not surprisingly revealed that the user group is extremely heterogeneous (due to widely differing linguistic and encyclopedic competencies and widely different requirements to the CTL). The analysis also revealed that the CTL would have to be designed in such a way that it would satisfy the demands of the user in a number of specific situations - the CTL functions. In conclusion, it was now possible to draw a profile of the user and to identify a number of *"knowledge-oriented"* and *"communications-oriented functions"* as discussed by [Tarp 1999]. Now the basis for the design and development phase was formed, and it was time to prepare the theoretical framework and the actual architecture of the CTL, which is the subject of the next section.

5. Theoretical Framework of Convergence Telecom Lexinome

With the above conclusions in mind, it was soon realized that traditional paper-based lexicographic solutions could not be used because the inherent limitations as to the contents and structure of paper-based lexicographic solutions rendered them unfit for use in the special lexicographic environment at TD. In the quest for a satisfactory lexicographic solution, selected elements of the idea discussed in [Almind/Bergenholtz 1998] about an integrated CD-ROM and Internet concept were of particular relevance for the purpose of this R&D project, but the theoretical framework of the CTL really started to materialize when I was inspired by [Lemberg et al. 1998] who state that *"Hypertextualisierung, die Einbindung von Bild-, Film- und*

Tonmaterial sowie Mwie Möglichkeiten der Vernetzung mit anderen Nachschlagewerken oder Forschungsinstituten fördern die Interdisziplinarität der Wörterbücher und bringen insgesamt eine beachtliche Qualitätssteigerung". The idea of networking is also discussed by [Storrer 1998], who outlines some considerations on how to utilize the Internet for dictionary purposes. These ideas were particularly relevant for this project, as integration and networking of different lexicographic and non-lexicographic data sources would be part of the solution. However, it was soon realized that the special lexicographic environment at TD more or less stipulated that the users should be involved in the lexicographic process. [Carr 1997] refers to this as *"bottom-up lexicography"*, but I have chosen to go a step further as the end-users are seen as *"active fellow players"* in all project phases. This concept is referred to as *"user involvement"* or *"lexicographic democracy"* which is only possible in an isolated lexicographic environment such as TD. In line with the concept of *"lexicographic democracy"*, and as the primary functions of the CTL is text production in English and translation from Danish into English, it was decided that what was called for was a lexicographic knowledge management system that enabled users to get access to both encyclopedic and linguistic data in the two parallel and specially selected corpora, which form the basis of the CTL. However, a similar idea was envisaged as early as in 1996 by for example [Atkins 1996].

In conclusion, the theoretical framework of the CTL may best be illustrated by figure 3.
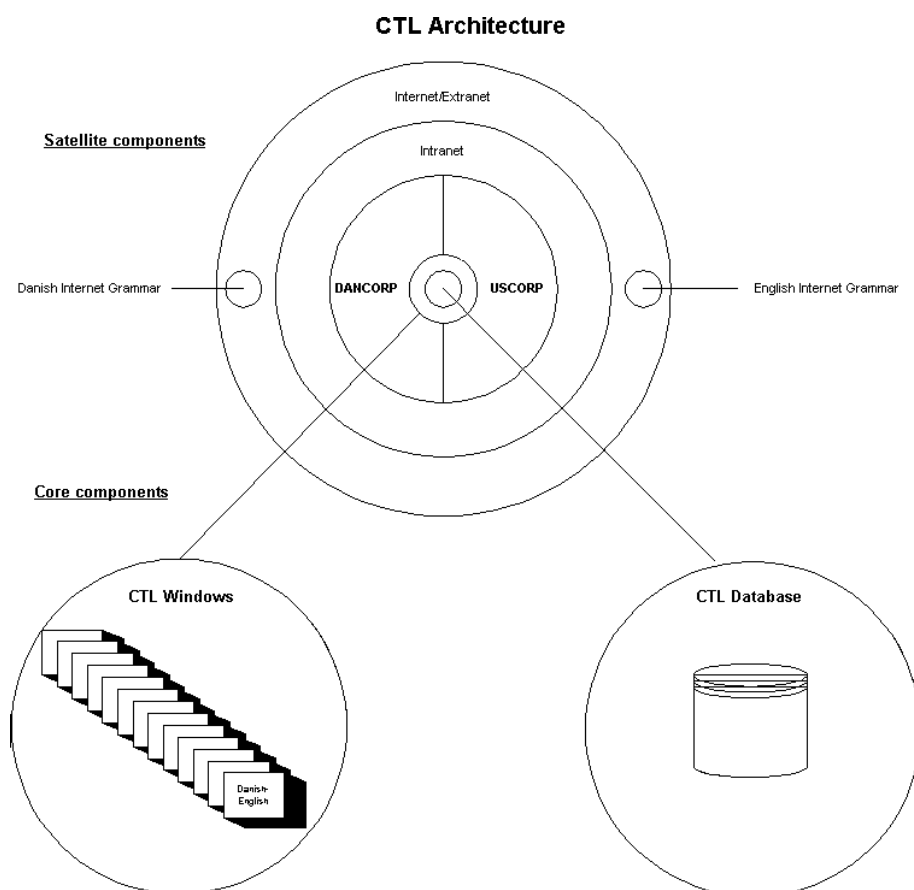

Figure 3: Framework and architecture of CTL

As figure 3 illustrates, a number of lexicographic and non-lexicographic data sources are integrated, networked, and converged into a combined database and Intranet-based platform. Users have access to conventional "static" lexicographic articles, but in addition to that they also have access to encyclopedic and linguistic data from the two corpora, just as they can initiate structured searches for "dynamic" data on TD's Intranet and on the Internet by clicking an icon from inside the CTL windows. The CTL core components consist of the CTL database and the individual CTL interface windows, which contain a wide variety of functionalities, for example an interactive terminology window, linguistic and telecom-related chatrooms giving CTL a dynamic and holistic dimension where lexicographic quality management and knowledge management are important elements. The satellite components are the "auxiliary components" such as the two corpora, TD's Intranet and the Internet. The surface-lexicographic result of the R& D project is shown in the CTL windows. They contain a wide variety of lexicographic data, including spoken pronunciation and illustrations, and the Danish-English CTL window may be divided into three overall sections: an Encyclopedic Section containing encyclopedic information, a Linguistic Section containing linguistic information and finally a Search/Reference Section containing icons that initiate structured searches in the satellite elements. This article and a selection of CTL windows formed the basis of a paper presented at EURALEX 2000.

6. Conclusion

On the basis of the analyses it must be concluded that Convergence Telecom Lexinome is a satisfactory solution to the lexicographic problem at TD. By means of the new lexicographic knowledge management system, TD is now able to register, harmonize, and quality assure the terminology used both internally and externally. It is also the ambition of this R&D project to make a modest contribution to lexicographic research by outlining some theoretical considerations on how to utilize and network static lexicographic data with dynamic lexicographic and non-lexicographic data from a wide variety of sources.In conclusion, the new and innovative CTL architecture implemented at TD's Intranet is thus estimated to solve the multi-faceted lexicographic problem identified above for the benefit of TD's employees and customers.

Notes

[1] `http://www.teledanmark.dk/`

[2] `http://www.asb.dk/`

[3] `http://www.atv.dk/`

[4] The term "internet reference works" covers dictionaries, databases, glossaries, and term lists on the Internet.

Bibliography

Almind, Richard/Henning Bergenholtz (1998). "Integreret cd-rom og internetkoncept til elektronisk referenceværk." In: Anna Garde/Pia Jarvad/Knud Troels Thomsen. *Elefant – se ogse også myg. Festskrift til Jens Axelsen.* København: Gyldendal, 205-215.

Atkins, B.T.S (1996). "Bilingual Dictionaries – Past, Present and Future". In: *Euralex Š96 Proceedings*. Göteborg University, Department of Swedish.

Bergenholtz, Henning/Sven Tarp (1995). *Manual of Specialised Lexicography*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Carr, Michael (1997). "Internet Dictionaries and Lexicography". In: *International Journal of Lexicography 10,* 209-230.

Lemberg, Ingrid/Sybille Petzold/Heino Speer (1998). "Der Weg des Deutschen Recthswörterbuchs in das Internet". In: Herbert Ernst Wiegand (ed.). *Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*, Tübingen: Niemeyer, 262-284.

Storrer, Angelika/Freese, Katrin (1996). "Wörterbücher im Internet". In: *Deutsche Sprache* 1996/2, 97-153.

Storrer, Angelika (1998). "Hypermedia-Wörterbücher: Perspektiven für eine neue Generation elektronischer Wörterbücher". In: Herbert Ernst Wiegand (ed.). *Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*, Tübingen: Niemeyer, 106-131.

Tarp, Sven (1995). "Wörterbuchfunktionen: Utopische und realistische Vorschläge für die bilinguale Lexikographie". In: Herbert Ernst Wiegand (Ed.): *Studien zur zweisprachen Lexikographie mit Deutsch II. Germanistische Linguistik 127-128, 1995.* Hildesheim/New York: Olms, 17-51.

Tarp, Sven (1999). "Theoretical foundations of the so-called cross-reference structures". In: *Lexicographica 15, 1999.* Tübingen: Max Niemeyer Verlag, 114-137.

Wiegand, Herbert Ernst (1998). "Was eigentlich ist Fachlexikographie? Mit Hinweisen zum Verhältnis von sprachlichen und enzyklopädischem Wissen". In: Horst haider Munske/Peter von Polenz/Oskar Reichmann/Reiner Hildebrandt (Eds.): *Deutscher Wortschatz. Lexikologische Studien. Ludwig Erich Schmitt zum 80. Geburtstag von seinem Marburger Schülern.* Berlin/New York: de Gruyter 1988, 729-790.