

Abstracts

Gregor Büchel, Bernhard Schröder

The choice of coding systems for complex text structures depends on numerous relevant criteria. In most cases the choice of an SGML or XML based coding system will meet these criteria. The data can be managed with the help of database management systems. For this purpose the text objects and relations have to be mapped onto the expressive means of the database type chosen. It depends strongly on the way the documents are structured and on the way the data will be used, which kind of database modelling is most adequate for retrieval and maintenance purposes.

Ingrid Schmidt, Carolin Müller

Dictionaries that are currently available in electronic form only exploit the potential offered by the medium in a very restricted way. It is clear the high quality of dictionaries in printed form is often lost in the electronic version. Since the requirements of the media landscape are constantly shifting, it is necessary to develop a new starting point for dealing with lexicographic content. To this end, we first explore various aspects of SGML, multiple media publishing, TEI, and meta lexicography. We then use this as a basis for discussing a fresh attempt to develop a durable lexicographic model that will outlive rapidly outdated media and with which we are thus in a position to react flexibly to market requirements.

Angelika Storrer

The paper examines the new prospects that hypertext technology opens up for creating innovative electronic dictionaries. It first discusses the ideas which are central to the concept of hypertext: The non-sequential storage and presentation of the data, the integration of text, sound, images and video, and the tools for searching and filtering lexical information that are provided by hypertext management systems. It will be argued that most existing electronic dictionaries are more or less digital copies of paper dictionaries which only utilize a small portion of the technical possibilities available.

For this reason, the main part of the paper will establish seven theses and discuss ways of using the value-adding features of hypertext technology in future electronic dictionaries more efficiently.

Ingrid Lemberg

The Internet delivers a multifunctional medium for publishing, production and communication. The essential components of Internet-dictionaries are their global and around-the-clock availability (Chapter 4.1), doing away with a limited amount of printed pages, the possibility for continual corrections and updates of dictionary articles that have already been published (Chapter 2.2), hypertextualization and the use of multimedia integration.

Within this new medium, dictionaries can be adapted and expanded using other forms of presentation, for example by creating direct links between the dictionary and the text-corpus that it is based upon or between other informational systems (Chapter 2.1.1). Using multimedia integration is considered to be particularly suitable for the presentation of lexical or lexicographic knowledge with different types of expressive mediums thus contributing to an efficient conveyance of knowledge (Chapter 2.1.2).

Internet availability also means that lexicography has a totally new production technology at its disposal: making lexicographic databases available on the Internet makes it possible to create homogenous and completely networked dictionaries from different locations spread out all over the world (Chapter 3). Being a medium for communication, the Internet makes it possible to establish direct contact between the user and the lexicographer: users of an Internet-dictionary can request comprehension support via e-mail, depending on the type of situation that users find themselves in. In addition, users can also make valuable and developmental contributions to the informational status of the dictionary (Chapter 4.2).

Annette Klosa

This article starts from a fundamental definition of "quality" and discusses which quality characteristics for CD-ROM dictionaries are mentioned in reviews. On this basis a list of quality criteria for electronic dictionaries is developed, which, finally, is compared with current practice in publishing houses.

Ulrike Haß-Zumkehr

This article contains reflections on the design of an abstract microstructure for a hypertext information system for the German language, which results from a lexicographical project at the Institut für Deutsche Sprache, Mannheim. The new medium confronts lexicographers with different conditions and consequences. Among these are the separation of data structure and data presentation, the creation of several ways of accessing the entry via various dimensions of information (semantics, grammar etc.), the context-free formulation of information elements, the removal of most of the text compression, the explicit naming of information elements to help the user select what he wants to know, the creation of a typology of links and the decision between a structure pivoting on the word form or on the single word meaning.

Thomas Gloning, Rüdiger Welter

Documenting a vocabulary as an electronic data base offers some considerable benefits as compared with the one-dimensionality of alphabetically arranged works of reference: it is a more adequate representation of the complex interior structure of a vocabulary concerning questions of onomasiology, text types, style(s), the history of ideas and concepts etc.; it provides ways of data retrieval that are differentiated and flexible in selection and combination; it is continuously open to amendments and revisions.

The article starts by explaining how the complex structure of a vocabulary (parts 1 and 2), in particular the interior structure of Goethe's vocabulary (part 3), may be characterized. Part 4 discusses some of the new possibilities the Goethe Dictionary as a SGML-coded data base offers its users, e.g. multiple use of the data, improved methods of analysis, and links to other related documents. It is argued that on the basis of these data the Goethe-Wörterbuch should be made available free of charge via the internet.

Thomas Burch, Johannes Fournier

The most important dictionaries of Middle High German presently are the ‚Mittelhochdeutsches Wörterbuch‘ by Benecke/Müller/Zarncke, the ‚Mittelhochdeutsches Handwörterbuch‘ by Lexer, its ‚Nachträge‘ and the ‚Findebuch zum Mittelhochdeutschen Wortschatz‘. These dictionaries are closely interconnected and have to be used simultaneously. They therefore are ideal candidates for the composition of an electronically interlinked dictionary. Additionally the lexicographical information within the entire contents of all four works of reference will be accessible via a database supplied with complex retrieval options. The application of SGML, defined as ISO standard, provides the ground for an encoding which does not depend on specific hardware or software, neither for the CD-ROM nor for the Internet. Moreover this markup will guarantee the longevity of the data concerned. The application of TEI Guidelines which have already been successfully employed in a variety of projects was essential to the steady progression of the retrospective digitisation.

The essay discusses problems that occurred when applying TEI Guidelines to the electronic dictionaries. It is however apparent that problems did not stem from the application of TEI Guidelines as such, but instead were primarily due to the fact that the articles of BMZ and Lexer's dictionary do not always follow a well defined overall structure, which has made automatic SGML encoding an often difficult task. In many cases only manual markup led to TEI compliant documents. Nevertheless the results achieved so far fully justify the decision in favour of TEI Guidelines.

Ralf Plate, Ute Recker

This essay discusses current possibilities and future prospects of the computer-aided composition of a historical citation dictionary (*Belegwörterbuch*), as they present themselves to the authors after five years of preliminary work towards a new Middle High German dictionary. The phase of elaboration of the dictionary, that is to say, the writing of entries, will in all probability begin in 2001. A period of twenty years will see the completion of the dictionary, the printed version consisting of four volumes, containing approximately 1000 pages each. The new dictionary is meant to assume the role hitherto fulfilled by its predecessors dating from the 19th Century, for the language of the literary sources from the period 1050 to 1350. Being the most recent of current projects concerning historical citation lexicography, it was possible to base all of the work on the new Middle High German dictionary entirely on electronic data processing from the very beginning. The compilation of data for the new dictionary is based on an comprehensive electronic archive containing all texts of the basic corpus in full, as well as a list of potential headwords (*Lemmakandidatenliste*) consisting of app.ly 80,000 lemmata, obtained from the

preceding dictionaries. After preparing the digitized text sources with specific markings, the lemmatized archive of citational evidence is then produced by employing a system of programmes devised for semi-automated lemmatization, drawing informational data from the texts. Additionally, the programming also inputs to the archive all word forms as derived independently from the texts by way of an especially devised lemmatization component. Thus the electronic archive of citational evidence will serve as the basis for the future dictionary, complete with an additional system of programmes which will support the composition of entries, from the coordination and presentation of citational data through to the stage of typesetting the finished entries.

By entirely basing the new dictionary on data processing, it will be possible to publish an electronic version alongside the printed edition of the dictionary, which will appear periodically in separate issues/installments. With little extra preparation the electronic publication will furthermore provide access to the digitized text material itself as well as offer a wide range of further data retrieval opportunities.

The future prospects of computer-aided lexicography as discussed in the contribution with regard to the new dictionary of Middle High German refer to the chance of improving traditional lexicographical research strategies by employing computers for the tasks of reorganizing and economizing the process of information retrieval. The greater part of former lexicographical endeavor used to consist of the copying, editing, and lemmatizing of individual pieces of citational evidence. When employing the aid of computers and semi-automated lemmatization software, the resources previously invested in this time-consuming process of compiling data can be used for the production of once-and-for-all carefully edited digitized texts which will be electronically processed for all further applications. Proceeding in this way means that the author of dictionary entries him/herself is retains full control over the choice of citational evidence, rather than being limited to the result of an earlier excerpt, and can thus make his/her selection on lexicographical grounds.

The composition of the entries themselves is supported by a system of programmes designed to perform most of the mechanical tasks such as the sorting, writing and controlling of the data concerned, thereby enabling the lexicographer to focus attention entirely on the analysis of the citational evidence and the formulation of lexicographical information.

Additionally, due to the obligatory formatting routines devised in Trier, the entry files produced will be supplied with a certain (minimal) amount of encoding as it relates to the overall structure of the electronic dictionary, thus allowing the later realization of sophisticated lexicographic information retrieval strategies.

Other prospects for computer-based lexicography can be discerned in the fact that much of the data which is compiled in order to produce the dictionary will not only be employed and made accessible in combination with the published work but can also be made accessible for independent use in support of other linguistic and philological research interests. In the case of the new Middle High German dictionary, additional advantages take the form of an electronic textarchive, the electronic archive of citational evidence, and, finally, an efficient instrument in the semi-automated production of lemmatized indices and concordances of any given Middle High German text.

The prospects discussed in the contribution refer to the particular advantages of publishing electronic dictionaries as well as to questions pertaining to lexicographical concepts. These questions are basically concerned with the nature of the relationship between the electronic publication of the dictionary and the data material it is based on (particularly the text- and citation archive), should they be published in combination and thus present the opportunity of proceeding directly from the article into the archive. We have also addressed

the question of whether and to what extent the dictionary should – in addition to its essential functions – be designed to meet the demands of new classes of users for whose purposes various kinds of structured listings or registers might be required.

Gerd Richter¹

Electronic forms of presentation such as hypertext provide new opportunities for the design of toponymic repertoires in the fields of presentation, retrieval and analysis. Based on the printed version of the *Südhessisches Flurnamenbuch* (SHFLNB), this article shows the advantages brought about by the transformation of the traditional book version into a hypertext. A model analysis of the SHFLNB which includes an analytical description of its medial structure will show the hypertextual potential of a toponymic repertory. An example realisation as a hypertext will then serve as a basis for a thorough description of the potential new modes of presentation for toponymic repertoires, especially in comparison with the traditional print version.

Krzysztof Petelenz

The availability of reference works on CD-ROM has led to the practical and commercial significance of hypertext for lexicography. The development of the WWW has lead to improved qualities for linked information; open systems, topical contents, easier communication between the authors and their audience and, consequently, an improved consideration of user needs.

It therefore appears appropriate to consider ways of improving on-line dictionaries. This paper attempts to lay the foundations for the presentation of a bilingual dictionary in high-quality hypertext. The structuring and storage strategies for dictionary data will be dealt with in particular and described using the example of a section of a Polish-German dictionary. First, however, we will discuss the position of bilingual computer-dictionaries within hypermedial systems and special features of bilingual lexicography which are significant for the structuring of bilingual on-line dictionaries. Particular attention will be paid to use and categorisation of pictures which can finally be given adequate space in this new medium. In conclusion, some "added value" features of an on-line dictionary will be discussed, such as the networking with other web-projects and the continuous expansion of the macro- and microstructure.

Claudia Kunze, Andreas Wagner

This paper presents GermaNet, a lexical-semantic network and on-line thesaurus which covers the German basic vocabulary. GermaNet encodes the basic semantic relations like synonymy, hyponymy, antonymy, and the causal relation that hold among the lexical items being implemented. This semantic network constitutes an important resource for word

¹ Für die Übersetzungen danke ich Frau Sabine Prechter. – G.R.

sense disambiguation which is a prerequisite for various applications within natural language processing like information retrieval and the semantic annotation of corpora.

Lothar Lemnitzer

Sound knowledge of user needs is crucial for the planning and realization of new and revised dictionaries. However, such data are costly to acquire. Nowadays the World Wide Web enables a new kind of contract with users. They are allowed to use dictionary data free of charge but in turn agree that their requests are logged and analysed. The present study presents such an approach with a set of bilingual dictionaries which are available on the net. User requests have been logged for more than a year. A quantitative and qualitative analysis of these data is presented.