# Natural Language Processing Techniques for Improved User-Friendliness of Electronic Dictionaries

Ulrich Heid
Universität Hildesheim
heidul@uni-hildesheim.de

## Abstract

There is no doubt that electronic dictionaries should be user-friendly. Lexicographic approaches inspired, for example, by the Function Theory of Lexicography (Tarp 2008, etc.) place the user and his/her needs in the center of their discussions about dictionary design.

In this talk, we aim at reviewing a number of proven to work techniques from Natural Language Processing, with respect to their impact on the user friendliness of electronic dictionaries. Our impression is that the potential of NLP tools used as components of online lexical information systems has not yet fully been exploited. We thus intended to discuss possibilities for combining dictionary resources with processing tools, to enhance the user friendliness of the integrated product.

We start from a brief recall of basic assumptions of the Lexicographic Function Theory, using a simplified model of reception and production oriented tasks as a background for the subsequent discussion of techniques from NLP. We then focus on aspects of data representation, and on comparatively simple, mainly symbolic tools and procedures that support users in text reception or text production tasks, as well as in bilingual tasks.

As dictionaries are seen by most users as authoritative reference works, we do not address approximative techniques, but concentrate on those approaches which can produce an assessment of the quality of their output. In this sense, reliability is a major issue, and the tools should either produce correct results or (at least) warn the user about the unverified status of their output (as it is done, for example, within the word formation analyzer available for German on the website of canoo.com).

An important prerequisite for language processing tools is an appropriate representation of the lexical data they use: this includes in particular a fine-grained categorization of the different typed of lexic(ographic)al data. Similarly, also lexicographers interested in providing appropriate data for a given dictionary function can profit massively from a fine-grained classification and description of different types of lexicographical data. This view has gained more wide-spread acceptance over the last few years. We will discuss it in the framework of a model of the function-theoretical approach which assumes a central lexicographical data collection and a set of filters to select, order and present lexicographical data for different target user groups and/or dictionary functions. In this context, we will report about the experience of publishers who use similar concepts for dictionary production.

We then address techniques for text reception dictionaries. Such NLP techniques mainly serve as access tools for the user; a prime example is inflectional morphology, which allows the user to insert (or, in an interactive dictionary related with text reading tools, to highlight) an inflected item and to get access to the appropriate lemma entry. We also discuss tools for word formation analysis and possibilities to include them into dictionary access devices. Another type of access tool that allows users to efficiently query the dictionary are devices that can identify idioms: if a user point this mouse to an item that is a part of an idiom, the tool verifies the idiom status of the full expression and displays only the dictionary entry of the idiom, and not that of the compositional meaning of the targeted item (cf. Seretan/Wehrli 2013). We assess to what extent such functions can be extended beyond idioms.

In terms of support for text production, inflection paradigms and/or inflection generation are relatively standard; we also think that links between dictionary and corpus, as they have been much discussed over the last few years, are very useful for text production, if combined with NLP techniques, for example to select appropriate examples for syntactic constructions of a given type. In a similar way, example selection from bilingual corpora may be a way forward towards tailor-made support for the verification of users' translation hypotheses.

This overview may help to identifying domains where the cooperation between lexicographers and NLP researchers and/or language technologists may lead to considerable improvements in the user friendliness of electronic dictionaries.

Sven Tarp (2008): Lexicography in the Borderland between Knowledge and Non-knowledge: General Lexicographical Theory with Particular Focus on Learner's Lexicography. Max Niemeyer.

Violeta Seretan, Eric Wehrli (2013): Context-sensitive look-up in electronic dictionaries. In: Rufus H. Gouws, Ulrich Heid, Wolfgang Schweickard, Herbert Ernst Wiegand (editors) *Dictionaries. An international encyclopedia of lexicography. Supplementary volume: Recent developments with special focus on computational lexicography, Handbooks of Linguistics and Communications Science (HSK-5/4)*. Walter de Gruyter, Berlin/New York.

**Language of Presentation**: English