

What Makes a Good Online Dictionary? – Empirical Insights from an Interdisciplinary Research Project

Carolyn Müller-Spitzer, Alexander Koplenig, Antje Töpel

Institute for German Language (IDS), Mannheim

E-mail: mueller-spitzer@ids-mannheim.de, koplenig@ids-mannheim.de, toepel@ids-mannheim.de

Abstract

This paper presents empirical findings from two online surveys on the use of online dictionaries, in which more than 1,000 participants took part. The aim of these studies was to clarify general questions of online dictionary use (e.g. which electronic devices are used for online dictionaries or different types of usage situations) and to identify different demands regarding the use of online dictionaries. We will present some important results of this ongoing research project by focusing on the latter. Our analyses show that neither knowledge of the participants' (scientific or academic) background, nor the language version of the online survey (German vs. English) allow any significant conclusions to be drawn about the participant's individual user demands. Subgroup analyses only reveal noteworthy differences when the groups are clustered statistically. Taken together, our findings shed light on the general lexicographical request both for the development of a user-adaptive interface and the incorporation of multimedia elements to make online dictionaries more user-friendly and innovative.

Keywords: dictionary use; empirical research; online dictionary

1. Introduction

Research into the use of online dictionaries is still quite a new field. Although some 250 to 300 studies have been carried out to date, the current state of knowledge still needs to be improved (Wiegand, 1998; Loucky, 2005; Welker, 2008; Engelberg & Lemnitzer, 2008; Tarp, 2009). Most studies are methodologically limited to the analysis of log files (e.g., de Schryver & Joffe, 2004; Bergenholtz & Johnson, 2005). While log file studies are able to provide reliable data about requested lemmas and related types of information, this method is not well suited to gaining insights into actual user demands. Take for instance the following hypothetical but plausible situation: Alex does not know the spelling of a particular word. To solve this problem, he visits an online dictionary. However, when trying to find the search window, he stumbles across various types of innovative buttons, hyperlinks and other distracting features. Instead of further using this online dictionary, he decides to switch to a well known search engine, because he prefers websites that enable him to easily find the information he needs. In this example, there would not be any data to log (except for an unspecified and discontinued visit to the website). In contrast, the market for online dictionaries is expanding both for academic lexicography and for commercial lexicography, with sales figures for printed reference works in continual decline. This has led to a demand for reliable empirical information on how online dictionaries are actually being used and how they could be made more user-friendly. As the example above indicates, relying completely on log file data can lead to biased conclusions in this context.

The remainder of this paper is structured as follows. In section 2, we will give a short overview of our project. Section 3 presents some of the hypotheses to be tested regarding online dictionary users' demands, while section 4 explains the methodological procedure. Section

5 describes some basic results. Finally, this study concludes with a short discussion of the implications of our findings (section 6) and briefly outlines our future work (section 7).

2. Project background

The project "User-adaptive access and cross-references in *elexiko* (BZVelexiko)" (www.using-dictionaries.info) aims to make a substantial contribution to closing this research gap. BZVelexiko is an externally funded joint research project at the Institute for German Language in Mannheim. For a period of three years, a group of researchers from different academic backgrounds (lexicographers, linguists, social scientists) is undertaking several extensive studies on the use of online dictionaries, using established methods of empirical social research. The first two studies focused on online dictionaries in general; subsequent studies in our project are restricted to monolingual German online dictionaries such as *elexiko* or the dictionary portal OWID (www.owid.de).

3. Demands on online dictionaries

Providing reliable empirical data that can be used to answer the question of how users rate different aspects of online dictionaries is an important issue for practical lexicography, because it can be used as the basis of various decisions that have to be made in this context. Is it more important to use financial and human resources to extend the corpus and improve its accessibility for the user community, or to focus on keeping the dictionary entries up to date? Which is more user-friendly, a fast user interface or a customizable user interface? Do different user groups have different preferences? For example, one of our hypotheses was that, compared to non-linguists, linguists would have a stronger preference for the entries to be linked to the relevant corpus, because this documents the scientific basis of the given information.

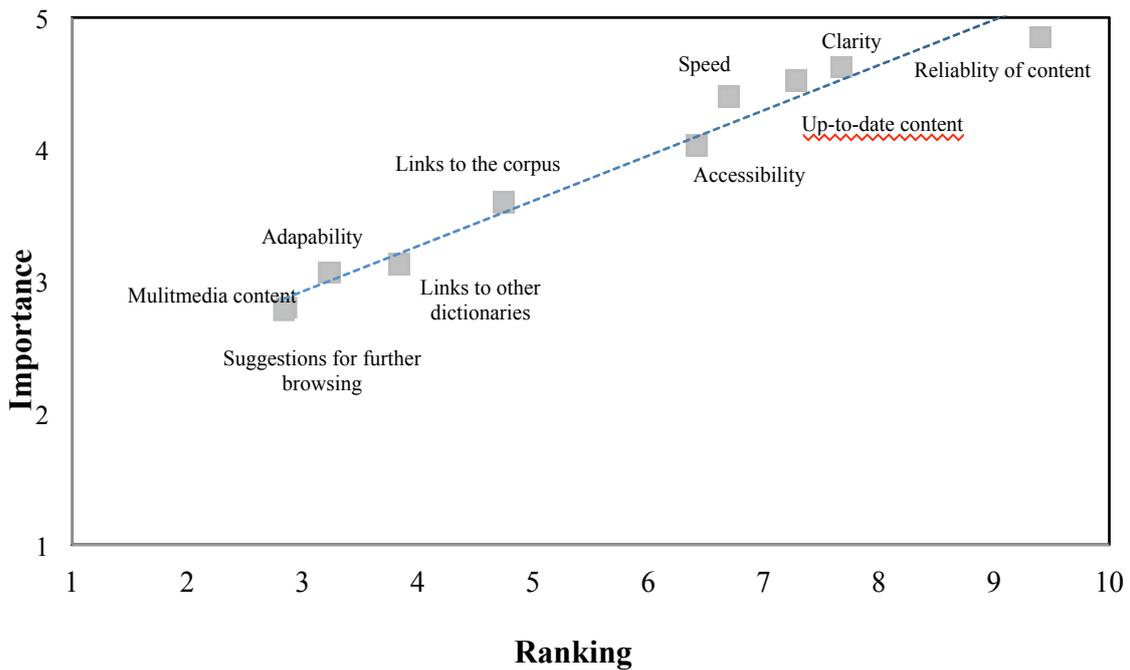


Figure 1. Correlations between means of ranks and means of importance regarding the use of an online dictionary. *Note.* Means of ranks are on 10-point scales and means of importance are on 5-point scales; both with higher values indicating higher levels of benefit.

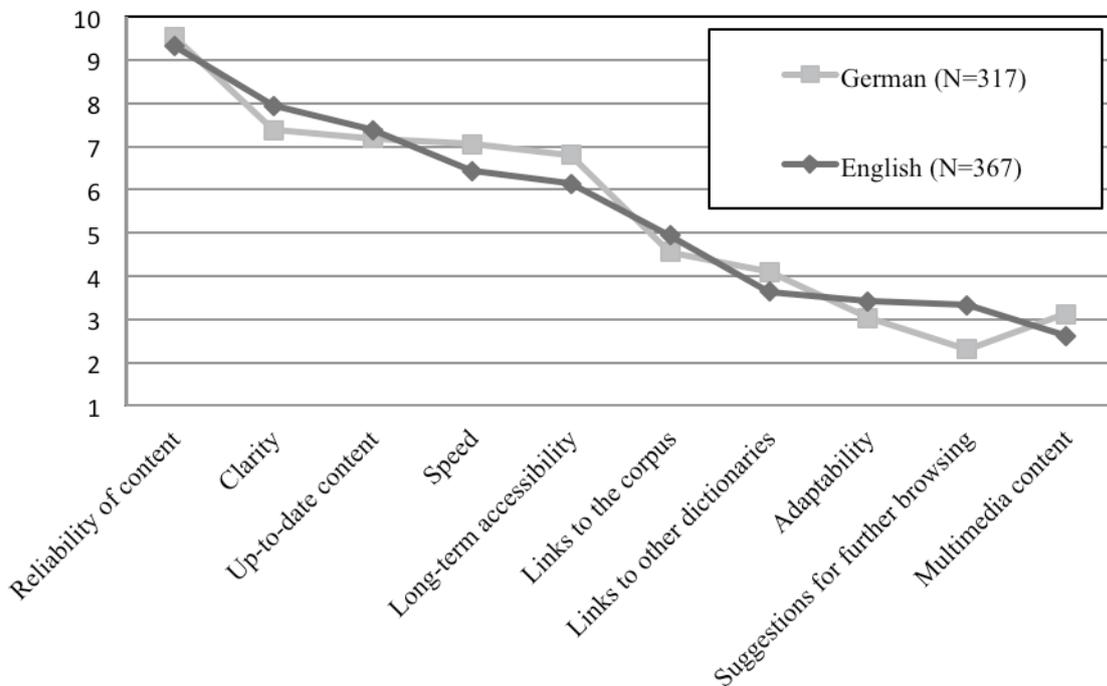


Figure 2. Means of Rankings as a Function of Language Version. *Note.* Means (Fig. 2, Fig. 3 and Fig. 4) are on 10-point scales with higher values indicating higher levels of importance regarding the use of an online dictionary.

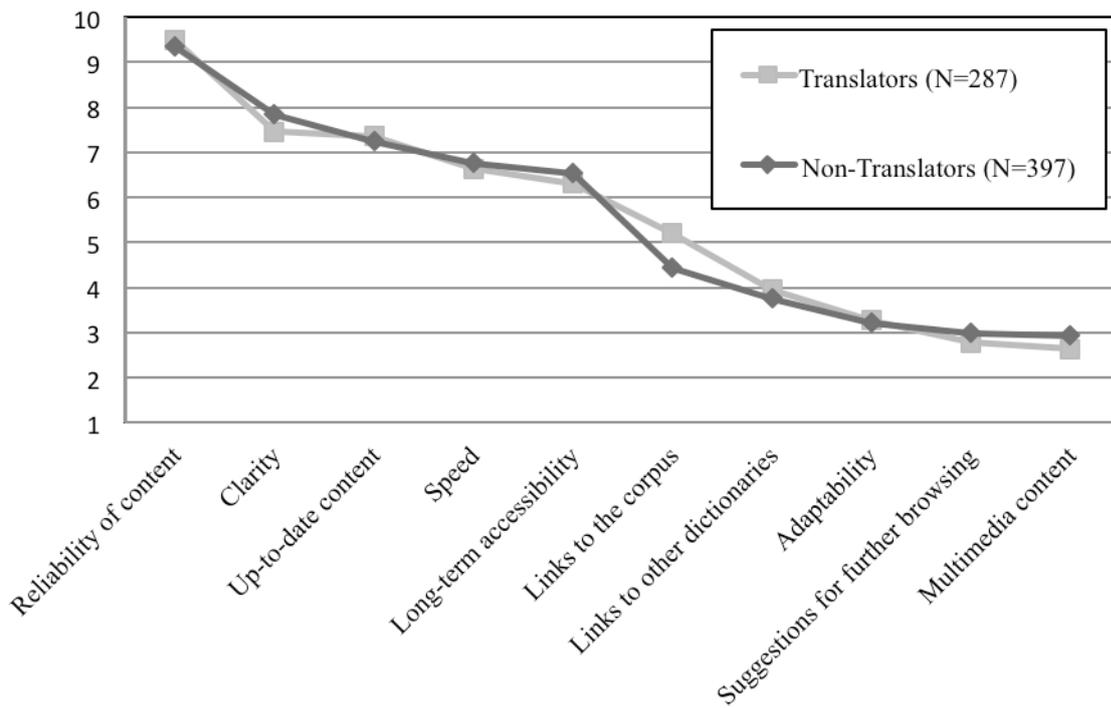


Figure 3. Means of Rankings as a Function of Professional Background (*Translators vs. Non-Translators*).

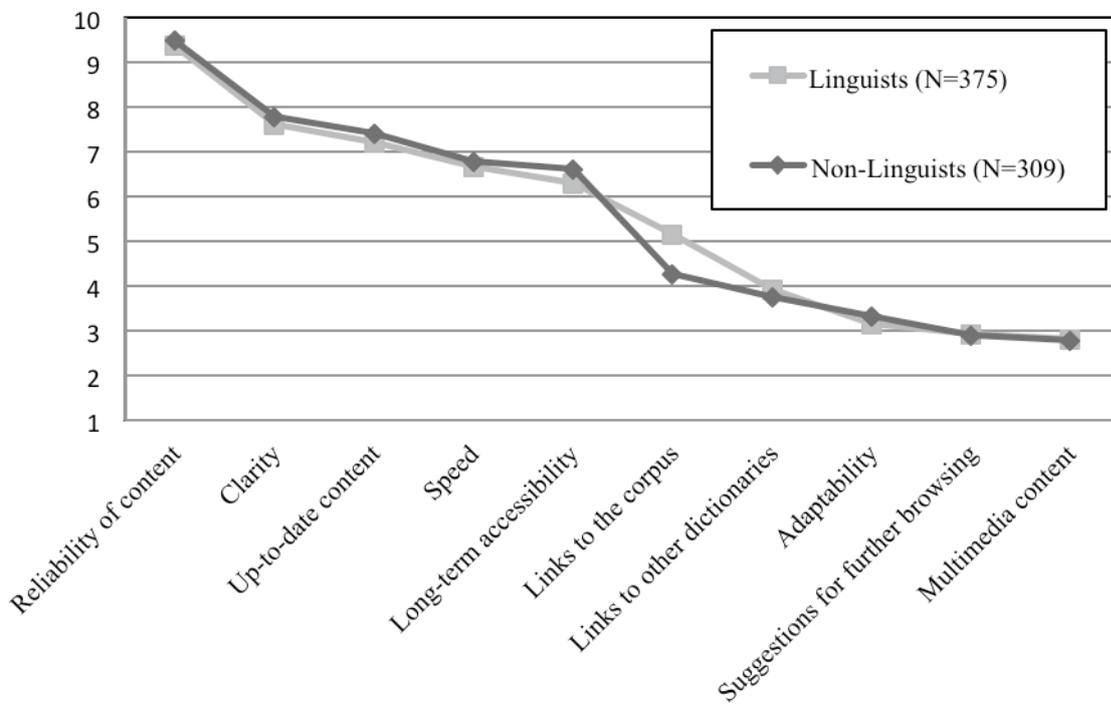


Figure 4. Means of Rankings as a Function of Academic Background (*Linguists vs. Non-Linguists*).

Another hypothesis was that we expected translators to rate, on average, a user interface that is adaptable to be more important for an online dictionary than non-translators, since professional translators rely heavily on dictionaries in their daily work. An adaptable user interface could enhance their individual productivity.

4. Method

To identify different user demands, we conducted two online surveys in English and German in 2010. A total of 1,074 respondents participated. Among other questions, respondents in the first survey (N = 684) were asked to rate ten aspects of usability on 5-point Likert scales (1 = not important at all, 5 = very important) regarding the use of an online dictionary (in the questionnaire, all criteria were explained fully): *Adaptability, Clarity, Links to other dictionaries, Links to the corpus, Long-term accessibility, Multimedia content, Reliability of content, Speed, Suggestions for further browsing, Up-to-date content.*

After this, participants were asked to create a personal ranking according to importance. The most important criterion was placed in tenth position, whereas the least important criterion was placed in first position. Furthermore, participants could choose in which language they wanted to complete the questionnaire (English/German) and were asked whether they work as a linguist and/or as a translator (yes/no) in order to analyze whether different users groups have different demands.

5. Results

5.1 Correlation Analysis

Analysis of (Spearman's rank) correlation revealed a significant association between importance and ranking; $r = 0.39$ [0.20; 0.56]; $p < .01$. These results indicate that the individual ranking can be used as a reliable indicator of users' demands as intended (cf. fig. 1).

5.2 Subgroup analyses

As mentioned in Section 3, another objective of the study was to assess whether the size of this difference depends on further variables, especially the participants' background (linguistic vs. non-linguistic; translator vs. non-translators) and the language version of the online survey chosen by the participants (German vs. English). Surprisingly, there are no noteworthy rating differences – on average – between different groups, as a visual inspection clearly demonstrates (cf. fig. 2, fig. 3, and fig. 4).

Statistical analyses of variance (not reported here) reveal that some of the differences in average ratings across subgroups are significant. However, this is mainly due to the high number of participants. In fact the F-Value (1, 682) as a test for statistical significance ranges from 0.20 to 59.11 with 8.08 on average, yielding highly significant differences ($p < .001$) in only 8 out of 30 cases.

Another way of framing these findings is to state that the relative ranking orders represented by the shapes of the curves correspond in each figure except for fig. 2, where a small difference between the two criteria rated on average as least important and second least important occurs. In other words, these results indicate that knowledge of the participant's background allows hardly any conclusions to be drawn about the participant's individual ranking.

5.3 Cluster Analysis

In order to better interpret these results, we conducted a cluster analysis to see how users might group together regarding their individual ranking. A two-cluster solution was identified. Means, standard deviations, and N of each cluster are presented in Table 1.

| | Cluster 1 (N = 206) | | Cluster 2 (N = 478) | |
|----------------------------------|------------------------|------|------------------------|------|
| | M | SD | M | SD |
| Criterion | | | | |
| Reliability of content | 9.09 | 1.79 | 9.54 | 0.91 |
| Clarity | 6.96 | 1.98 | 7.97 | 1.35 |
| Up-to-date content | 6.89 | 2.28 | 7.45 | 1.50 |
| Speed | 5.52 | 2.56 | 7.21 | 1.47 |
| Long-term accessibility | 5.43 | 2.47 | 6.86 | 1.86 |
| Links to the corpus | 7.01 | 1.93 | 3.77 | 1.60 |
| Links to other dictionaries | 4.72 | 2.11 | 3.46 | 1.47 |
| Adaptability | 3.59 | 2.04 | 3.08 | 1.73 |
| Suggestions for further browsing | 3.35 | 2.19 | 2.64 | 1.55 |
| Multimedia content | 2.43 | 1.75 | 3.02 | 1.89 |

Table 1. Means and Standard Deviations of Rankings as a Function of the Cluster Analysis

Analyses of variance with the cluster as independent variable and the respective criterion as a response variable yielded highly significant differences ($p < .001$) for every criterion (10 out of 10 cases) with F (1, 682) ranging from 11.22 to 520.30 (93.08 on average). Most strikingly, only preceded by "Reliability of content", respondents in Cluster 1 rate the criterion "Links to the corpus" on average as the second most important aspect of a good online dictionary (M = 7.01, SD = 1.93), whereas this criterion only plays a minor role for respondents in Cluster 2 (M = 3.77, SD = 1.60), $F(1, 682) = 520.30$, $p < .000$ (cf. fig. 5). Taken together, the findings reported here suggest that our initial hypothesis that different groups have different demands was too simple.

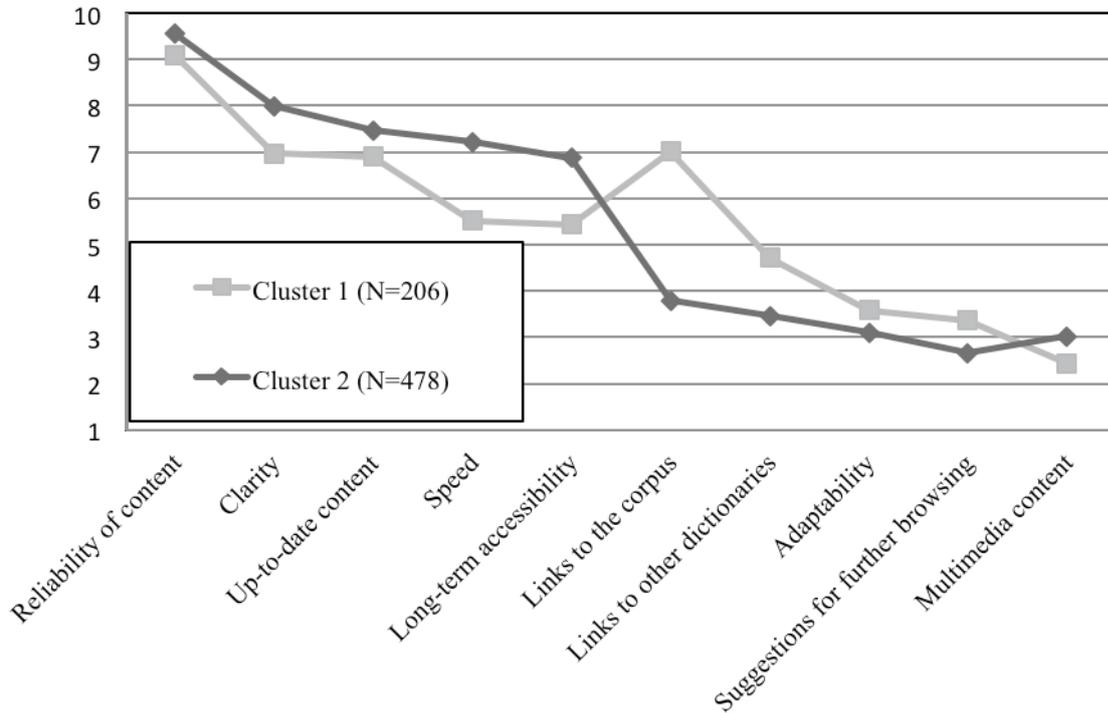


Figure 5. Means of Rankings as a Function of the Cluster Analysis

In Müller-Spitzer/Koplenig (*manuscript in preparation*), we argue that different background variables seem to interact with each other. By using a binary logistic regression model, we show that the probability of belonging to one of the two clusters (as an indicator for sharing similar individual demands regarding the use of an online dictionary) depends on academic background and on professional background and on the language version chosen. Our model indicates, for example, that the probability of belonging to the first cluster (N=206) for subjects in the English language version who work as translators and who have a linguistic academic background is 0.42 (0.95 confidence interval: 0.33 - 0.55), compared to a likelihood of only 0.13 for subjects in the German language version who do not work as translators and who do not have a linguistic background (0.95 confidence interval: 0.08 - 0.21).

6. Discussion

In our study, the classical criteria of reference books (e.g. reliability, clarity) were both ranked and rated highest, whereas the unique characteristics of online dictionaries (e.g. multimedia, adaptability) were rated and ranked as (partly) unimportant.

This result conflicts with the general lexicographical request both for the development of a user-adaptive interface and the incorporation of multimedia elements to make online dictionaries more user-friendly and innovative (e.g., de Schryver, 2003; Müller-Spitzer, 2008; Verlinde & Binon, 2010 present evidence challenging that view).

As is the case for printed dictionaries, our results indicate that online dictionaries are initially being used as a reference work providing reliable and accurate information. The unique characteristics of online dictionaries (e.g. multimedia, adaptability) only seem to play a minor role.

Nevertheless, this does not mean that the development of innovative features of online dictionaries is pointless. As we show elsewhere in detail (Koplenig, 2011; Müller-Spitzer & Koplenig, in preparation), users tend to appreciate good ideas, such as a user-adaptive interface, but they are just not used to online dictionaries incorporating those features. As a result, they have no basis on which to judge the usefulness of those features. Thus, in order to make an online dictionary more user-friendly by implementing innovative features, it is essential that users are also shown the potential benefits of those features.

7. Future Research

The results presented in this paper are still at a preliminary stage. Nevertheless, we believe that they show that both practical lexicography and theoretical lexicology can benefit from this research agenda by shedding some light on an important aspect of dictionary usage in a way that would not be possible through the use of log file analyses alone.

As a next step, to further enhance our understanding of online dictionary usage, we plan to incorporate the

insights gained from an eye-tracking study that we have conducted.

8. Acknowledgements

This project is financed by the Gottfried Wilhelm Leibniz Scientific Community ("Joint Initiative for Research and Innovation").

9. References

- Bergenholtz, H., Johnson, M. (2005). Log Files as a Tool for Improving Internet Dictionaries. *Hermes*, (34), pp. 117-141.
- de Schryver, G.-M. (2003). Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography*, 16(2), pp. 143-199.
- de Schryver, G.-M., Joffé, D. (2004). On How Electronic Dictionaries are Really Used. In G. Williams, S. Vessier (eds.) *Proceedings of the Eleventh EURALEX International Congress, Lorient, France, July 6th–10th*. Lorient: Université de Bretagne Sud, pp. 187-196.
- Engelberg, S., Lemnitzer, L. (2008). *Lexikographie und Wörterbuchbenutzung*. Tübingen: Stauffenburg Verlag.
- Koplenig, A. (2011). Understanding How Users Evaluate Innovative Features of Online Dictionaries – An Experimental Approach (Poster). Presented at the *eLexicography in the 21st century: new applications for new users (eLEX2011)*, organized by Trojina, Institute for Applied Slovene Studies, Bled, November 10-12, 2011.
- Loucky, J.P. (2005). Combining the Benefits of Electronic and Online Dictionaries with CALL Web Sites to Produce Effective and Enjoyable Vocabulary and Language Learning Lessons. *Computer Assisted Language Learning*, 18(5), pp. 389-416.
- Müller-Spitzer, C. (2008). Research on Dictionary Use and the Development of User-Adapted Views. In A. Storrer, A. Geyken, A. Siebert & K.-M. Würzner (eds.) *Text Resources and Lexical Knowledge Selected Papers from the 9th Conference on Natural Language Processing KONVENS 2008*. Berlin: de Gruyter, pp. 223-238.
- Müller-Spitzer, C., Koplenig, A. (in preparation). *Demands on Online Dictionaries: An Exploration of Group Differences and its Lexicographical Consequences (working title)*.
- Tarp, S. (2009). Beyond Lexicography: New Visions and Challenges in the Information Age. In H. Bergenholtz, S. Nielsen & S. Tarp (eds.) *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Frankfurt a.M./Berlin/Bern/Bruxelles/NewYork/Oxford/Wien: Peter Lang, pp. 17-32.
- Verlinde, S., Binon, J. (2010). Monitoring Dictionary Use in the Electronic Age. In A. Dykstra, T. Schoonheim (eds.) *Proceedings of the XIV Euralex International Congress*. Ljouwert: Afûk, pp. 1144-1151.
- Welker, H.A. (2008). Sobre o Uso de Dicionários. *Anais do 8º Encontro do CELSUL*, pp. 1-17.

Wiegand, H.E. (1998). *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. Berlin, New York: de Gruyter.