

Understanding How Users Evaluate Innovative Features of Online Dictionaries – An Experimental Approach

Alexander Koplenig

Institute for German Language (IDS), Mannheim

E-mail: koplenig@ids-mannheim.de

Abstract

Compared with printed dictionaries, online dictionaries provide a number of unique possibilities for the presentation and processing of lexicographical information. However, in Müller-Spitzer/Koplenig/Töpel (2011) we show that – on average – users tend to rate the special characteristics of online dictionaries (e.g. multimedia, adaptability) as (partly) unimportant. This result conflicts somewhat with the lexicographical request both for the development of a user-adaptive interface and the incorporation of multimedia elements. This contribution seeks to explain this discrepancy, by arguing that when potential users are fully informed about the benefits of possible innovative features of online dictionaries, they will come to judge these characteristics to be more useful than users that do not have this kind of information. This argument is supported by empirical evidence presented in this paper.

Keywords: dictionary use; empirical research; online dictionary

1. Introduction

In contrast to the long tradition of printed dictionaries, online dictionaries are a very recent phenomenon in lexicography. In addition to the classical criteria of reference books (such as reliability of content), online dictionaries have unique and innovative features resulting from the possibilities of the electronic medium, e.g. multimedia applications or a user interface that is customizable (de Schryver, 2003). Thus, it is not clear *a priori* how users judge the potential usefulness of those features. In this paper, we present and discuss the findings of an interdisciplinary research project that tries to investigate this question by applying various methods of empirical social research (Müller-Spitzer et al., 2011).

The first studies carried out as part of our project combined online surveys with experimental elements. Throughout this project, we plan to conduct some lab usability tests including an eye-tracking study. Based on this research agenda, we aim to illustrate these points by showing how empirical research can foster new insights and increase our understanding of dictionary use.

This paper is structured as follows. In section 2, we give a short overview of the project background and the hypothesis to be tested. Sections 3 and 4 describe the experimental approach and the methodological procedure, while section 5 presents the result. Finally, this study concludes with a brief discussion of the implications of the findings (section 6).

2. Background and Hypothesis

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 of the first study (N = 684) were asked to rate ten aspects of usability with respect to their im-

portance regarding the use of an online dictionary and then to rank these aspects in order of importance. The classical criteria of reference books (e.g. reliability, clarity) were both rated and ranked highest, whereas the unique characteristics of online dictionaries (e.g. multimedia, adaptability) were rated and ranked as (partly) unimportant¹ (Müller-Spitzer et al., 2011). 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 (de Schryver, 2003; Müller-Spitzer, 2008).

We assume that one possible explanation for this result is the fact that respondents are not used to online dictionaries incorporating those features. Thus, respondents currently have no basis on which to judge their potential usefulness. This line of reasoning predicts a learning effect. That is, when users are fully informed about possible multimedia and adaptable features, they will come to judge these characteristics to be more useful than users who do not have this kind of information.

To test this assumption, we incorporated an experimental element into our second survey.

3. Experiment

The participants of our second online study were presented, both visually and linguistically, with several possible multimedia applications and various features of an adaptable online dictionary in a set of statements (S1). Each feature was explained in detail (cf. Table 1) and/or supplemented by a picture illustrating its potential function. The participants were then asked to rate each feature with respect to three different characteristics regarding the use of an online dictionary (importance/benefit/helpfulness).

¹ Analysis of correlation revealed a significant association between importance and ranking; $r = 0.39$ [0.20; 0.56]; $p < .01$.

| Domain | Feature | Explanation |
|-------------------|--------------------------------|---|
| Multimedia | Audio pronunciations | In contrast to a printed dictionary, an online dictionary can include audio files illustrating the pronunciation of a word, a phrase or a whole sentence. |
| | Collocation graphs | An online dictionary can represent collocations, i. e. frequently occurring word combinations, in a visual form. |
| | Illustrations | An online dictionary can contain illustrations. |
| User adaptability | Customized user interface | To facilitate access to relevant personal information, the user interface of the online dictionary automatically adapts to the user's preferences depending on the item classes used in previous search requests. In this process, the online dictionary "remembers" a particular item class. For example, if the user frequently consults an online dictionary to search for synonyms, then a special search window for this kind of request appears on the homepage. A widely known commercial example is the homepage of the mail-order company Amazon, which changes according to the user and his/her previous shopping preferences. |
| | Alternative profiles | This means that the user of the online dictionary can choose between different profiles that optimally adjust the content according to the user's needs. For this purpose, the user first chooses between different user types and/or different usage situations. Certain defaults are then used to structure the mode of content presentation. |
| | Dynamic visual representations | This refers to the possibility of creating a personalised user view of the online dictionary. This can be done by choosing between different item classes, e. g. paraphrase, sense relations, information on grammar or citations. |

Table 1: Features of user adaptability and multimedia presented in the survey

In a second set (S2), participants were asked to indicate how much they agree with the following two statements: *The application of multimedia and adaptable features ...*

- (A) ... makes working with an online dictionary much easier.
(B) ... in online dictionaries is just a gadget.

To induce a learning effect, we randomized the order of the two sets: participants in the learning-effect condition (*L*) were first presented with the examples in S1. After that, they were asked to indicate their opinion in S2. Participants in the non-learning-effect condition (*N*) had to answer S2 followed by S1. Thus, to judge the potential usefulness of adaptability and multimedia, the participants in the learning-effect condition could use the information presented in S1, whereas the participants in the non-learning-effect condition could not rely on this kind of information. If our assumption is correct, participants in the learning-effect condition *L* will judge adaptability and multimedia to be more useful compared with participants in the non-learning-effect condition *N*.

Another objective of the experiment was to assess whether the size of this difference depends on further variables, especially the participants' background (lin-

guistic vs. non-linguistic²) and the language version of the online survey as chosen by the participants (German vs. English).

4. Method

Three hundred and eighty-one people participated in the bilingual online survey on the use of online dictionaries. Participants were randomly assigned to one of the conditions of the 2 (learning vs. non-learning-effect) factorial design.

The dependent variables were measured as described above (S2). Both ratings were made on 7-point Likert scales (1 = *strongly disagree*, 7 = *strongly agree*). The answers to these two items were averaged and oriented in the same direction to form a reliable scale of adaptability and multimedia benefit judgments ($\alpha = .75$), with higher values indicating more benefit.

5. Results

An ANOVA yielded a significant effect of outcome, $F(1, 379) = 12.27$, $p < .001$. As hypothesized, the results showed that participants in *L* judged adaptability and multimedia to be more useful ($M = 5.02$, $SD = 1.30$, $N = 175$) than participants in *N* ($M = 4.50$, $SD = 1.54$, $N = 206$).

² We asked the participants whether they work as a linguist and whether they study linguistics (yes/no).

In order to better interpret these results, we conducted a three-way ANOVA with condition, background and language version as independent factors. The statistical analysis revealed significant main effects for condition ($F(1, 373) = 18.29, p < .01$), for background ($F(1, 373) = 8.75, p < .05$), and for language version ($F(1, 373) = 13.56, p < .01$). In addition, a significant three-way interaction between experimental condition, background, and language version was found ($F(3, 371) = 7.59, p < .05$); cf. table 2.

Post hoc comparisons using the Tukey HSD test indicated that the mean difference in the German language version between the conditions was significant for the non-linguists ($p < .05$) and insignificant for the linguists ($p = .99$), whereas the difference between the two conditions was highly significant ($p < .00$) for the linguists and insignificant for the non-linguists ($p = .41$) in the English language version.

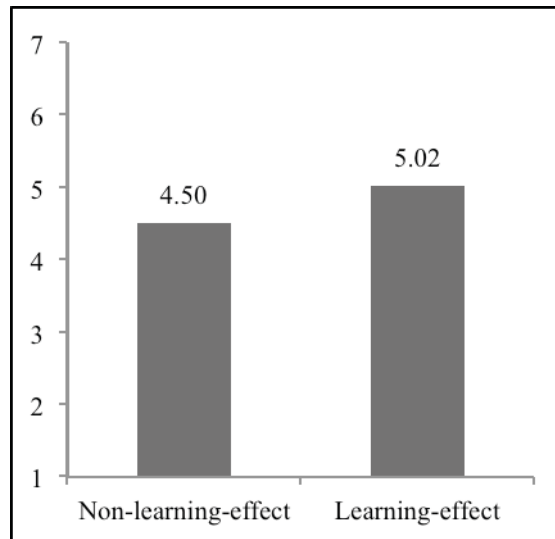


Figure 1: Means of adaptability and multimedia benefit judgments as a function of the learning-effect condition. *Note.* Means are on 7-point scales with higher values indicating higher levels of judgments of benefit

| German Language Version | | |
|--------------------------|--------------------|--------------------|
| | Background | |
| | Linguistic | Non-Linguistic |
| Condition | | |
| Non-learning-effect | 5.02 (1.47) | 4.45 (1.66) |
| Learning-effect | 5.02 (1.18) | 5.09 (1.35) |
| English Language Version | | |
| | Background | |
| | Linguistic | Non-Linguistic |
| Condition | | |
| No-learning-effect | 4.23 (1.47) | 4.12 (1.63) |
| Learning-effect | 5.15 (1.26) | 4.45 (1.50) |

Table 2: Means of adaptability and multimedia benefit judgments as a function of condition, background and language version. *Note.* Means are on a 7-point scale, with higher values indicating higher levels of benefit. Significant differences in bold. Standard deviations in parentheses

6. Discussion

As predicted, the results revealed a learning effect. This effect turned out to be modest in size (about half a point on a 7-point scale), but highly significant: from a Bayes-

ian point of view, the probability of observing this difference, given that it is just due to a random variation, is 0.05%. Furthermore, it should be noted here that we implemented only a weak manipulation of the learning

effect. Due to the nature of our survey design, we simply presented several features of multimedia and adaptability. It seems plausible to assume that, if the participants had the opportunity to actually use the presented features, the observed learning effect would be even bigger.

However, a closer inspection (cf. Table 2) showed that this difference is mediated by linguistic background and language version: while there is a significant learning effect in the German version but only for non-linguists, there is a highly significant learning effect in the English version but only for linguists. This leaves room for further studies focusing on the reasons for this interaction effect.

7. Acknowledgements

This project is financed by the Gottfried Wilhelm Leibniz Scientific Community ("Joint Initiative for Research and Innovation"). I thank one anonymous reviewer for a helpful comment on the title of this paper.

8. References

- de Schryver, G.-M. (2003). Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography*, 16(2), 143-199.
- 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. & Töpel, A. (2011). What Makes a Good Online Dictionary? – Empirical Insights from an Interdisciplinary Research Project. Presented at the eLexicography in the 21st century: new applications for new users, organized by Trojina, Institute for Applied Slovene Studies, Bled, Slovenia, November 10-12, 2011.