

Carolin Müller-Spitzer, Frank Michaelis, Alexander Kopleinig  
**Evaluation of a new web design for the dic-  
tionary portal OWID**

An attempt at using eye-tracking technology

**Abstract:** The main aim of the study presented in this chapter was to try out eye-tracking as form to collect data about dictionary use as it is – for research into dictionary use – a new and not widely used technology. As the topic of research, we decided to evaluate the new web design of the IDS dictionary portal OWID. In the mid of 2011 where the study was conducted, the relaunch of the web design was internally finished but externally not released yet. In this regard, it was a good time to see whether users get along well with the new design decisions. 38 persons participated in our study, all of them students aged 20-30 years. Besides the results the chapter also includes critical comments on methodological aspects of our study.

**Keywords:** eye-tracking, web design, screen layout, dictionary portal, sense navigation

---

**Carolin Müller-Spitzer:** Institut für Deutsche Sprache, R 5, 6-13, D-68161 Mannheim, +49-(0)621-1581429, mueller-spitzer@ids-mannheim.de

**Frank Michaelis:** Institut für Deutsche Sprache, R 5, 6-13, D-68161 Mannheim, +49-(0)621-1581423, michaelis@ids-mannheim.de

**Alexander Kopleinig:** Institut für Deutsche Sprache, R 5, 6-13, D-68161 Mannheim, +49-(0)621-1581435, kopleinig@ids-mannheim.de

## 1 Introduction

Just like all the studies described in this volume (with the exception of the log file study, Kopleinig et al., this volume), the eye-tracking study described here was conducted as part of the project on research into dictionary use at the Institute for German Language (IDS), which was externally financed and ran from 2009 until 2011 (cf. Müller-Spitzer: Introduction, this volume). During the project, we mainly conducted online studies in the form of surveys. In the interests of methodological diversity, however, we wanted to try out another form of data collection, namely eye-tracking technology. The main aim was therefore to try out this way of collecting data about dictionary use as it is – in research into dictionary use – a new and not widely used technology. This is also mentioned as an aim of the eye-tracking study by Lew et al.; similarly, Tono concludes his report of an eye-tracking study with the

request that more eye-tracking studies should be conducted in the area of research into dictionary use:

“As eye tracking has been used very little in dictionary user studies so far, another goal of the study is to examine the applicability of this technique to the study of dictionary entry navigation.” (Lew, Grzelak, & Leszkowicz, 2013, 230)

“I hope that this study will trigger more interest in taking rigorous research methods using such apparatus as an eye mark recorder.” (Tono, 2011, p. 152)

In this case, our approach was rather exploratory (cf. Koplenig, this volume). Rather than putting the research question first and then starting the process of finding an appropriate study design, the method of data collection was the starting point. One of the reasons was that our project was a good opportunity for the IDS to gain experience in this kind of user study, because we had the funding which made it possible to rent the lab at the University of Mannheim, to pay for the software, and to pay a research assistant to supervise the study and the participants. In other respects, as far as project organization is concerned, however, it was not the best time to conduct the study: at the end of the project, many tasks had to be finished, and in addition to that, the project team was reduced in number (e. g., due to maternity leave). These are some of the factors which led to the methodological shortcomings mentioned later.

As the topic of research, we decided to evaluate the new web design of the IDS dictionary portal OWID.<sup>1</sup> In the middle of 2011, the new web design was finished but had not yet been launched. In this respect, it was a good time to see whether users got along well with the new design decisions.

This chapter is structured as follows: Section 2 provides a brief insight into eye-tracking technology; Section 3 includes a brief description of OWID which will make the questions asked in the study easier to understand; Section 4 provides a summary of the study aim (4.1), procedure and apparatus (4.2) and the participants (4.3); an evaluation of several aspects of the new web design of OWID is presented in Section 5, where all the results of our eye-tracking study are presented and discussed; and instead of general concluding remarks, this chapter ends with critical comments on methodological aspects of our study.

---

<sup>1</sup> [www.owid.de](http://www.owid.de) (last accessed 13 July 2013).

## 2 Eye-tracking technology

Eye-tracking is a nearly 100-year-old technology, but has only recently been used for research into dictionary use. It is the process of measuring either the point of gaze (where someone is looking) or the motion of an eye relative to the head; an eye-tracker is a device for measuring eye positions and eye movement. In the context of usability studies, eye-trackers provide valuable insight into which features on a website are the most eye-catching, which features cause confusion and which ones are ignored altogether. In the process of eye-tracking, two basic parameters are measured: *saccades* and *fixations*.

“Saccades are rapid eye movements used in repositioning the fovea to a new location in the visual environment.” (Duchowski, 2007, p. 42)

“Fixations are eye movements that stabilize the retina over a stationary object of interest.” (Duchowski, 2007, p. 46)

As the eye is considered a “window to the mind”,<sup>2</sup> “gaze behaviour is usually interpreted as reflecting perception” (Lew et al., 2013, 230), based on the eye-mind assumption by Just and Carpenter (1980).

“The important eye-mind assumption proposed by Just/Carpenter (1980) also needs to be defined. The eye-mind assumption is based on the widely recognized assumption that there is a high correlation between long fixation durations and effortful processing in the user’s brain. [...] eye fixation and gaze time data reflect cognitive processes in the user’s brain.” (Simonsen, 2011, p. 76)

Among the advantages of eye-tracking studies are that:

- an analysis of the duration and number of fixations and saccades makes it possible to find out if users are focusing on the content, e. g. if they are reading a text carefully or only briefly scanning the screen;
- eye-tracking identifies areas on the screen that receive special attention;
- it is an online method, i.e. actual behaviour is recorded.

However, some of the disadvantages are that:

- people sometimes fix things with their eyes without actually perceiving them; whether this is the case or not cannot be confirmed through the use of eye-tracking systems (the opposite is beyond dispute: what is not seen is not perceived);
- information at the periphery of the visual field can reach the cognitive system and be processed; eye-tracking provides no data or analysis in relation to this;

---

<sup>2</sup> <http://ni.www.techfak.uni-bielefeld.de/research> (last accessed 13 July 2013).

- the eye-tracking method is limited to quantitative assessment; the mere statement of the fact that someone looked first at the header of a certain screen page allows no qualitative conclusion as to why this is the case.

For further insights into the method, see the article by Lew et al. (2013) mentioned above, and the chapter in this volume by Kemmer, who also provides detailed explanations of the eye-tracking method (Kemmer, this volume).

The method is, according to Lew et al., a promising avenue for research into dictionary use:

“Overall, eye-tracking technology proves to be a highly fitting and fruitful approach for examining what happens in dictionary consultation, and should be used more widely.” (Lew et al., 2013, 253)

In addition to the study by Lew, Simonsen and Tono have conducted eye-tracking studies (Simonsen, 2009, 2011; Tono, 2011); as those studies are reviewed in Lew et al. (2013, 230-32), they will not be commented upon here. With regard to dictionary portals, we are not aware of any user studies which concentrate on the web design of a portal.

### 3 A brief description of OWID

The *Online-Wortschatz-Informationssystem Deutsch* (OWID; Online German Lexical Information System) is a lexicographic Internet portal for various electronic dictionary resources that are being compiled at the IDS (cf. Müller-Spitzer, 2010). The main emphasis of OWID is on academic lexicographic resources of contemporary German. The dictionaries included in OWID range from a general monolingual dictionary (elexiko, cf. Klosa/Kopenig/Töpel, this volume) to a dictionary of neologisms, discourse dictionaries, a dictionary of proverbs and fixed multiword expressions, and a dictionary of German communication verbs.<sup>3</sup> OWID is a typical example of a dictionary net (in the sense of (Engelberg & Müller-Spitzer, forthcoming), as it provides inner, outer and external access to the included dictionaries, inter-dictionary cross-references and an integrated layout of portal and individual dictionaries. OWID is a constantly growing resource for academic lexicographic work in the German language.

In 2010, we planned to relaunch the website. The guiding principles for the new design were on the one hand to increase the visibility of the individual dictionaries in OWID, i.e. to strengthen the character of OWID as a dictionary portal and, on the

---

<sup>3</sup> See [www.owid.de](http://www.owid.de) (last accessed 13 July 2013).

other hand, to make the layout clearer and less cluttered. Specifically, the following new design elements that are relevant for the study were introduced:

- The screen (for the entry view) was divided into three parts: a navigation bar on the left with the keyword list or other navigational elements, the centre for the entry itself and a new bar on the right-hand side of the screen in which the different dictionaries from OWID are listed vertically. This latter bar is always visible, even if entries are displayed. (In the old design, the various dictionaries were listed on the homepage, but they were not present when an entry was displayed.)
- To identify the individual dictionaries, a new colour scheme was introduced. In the old layout, the key words themselves were highlighted in coloured type, but this led to a very problematic and cluttered visual representation. In the new layout, the dictionaries and the keywords are only preceded by a coloured box (cf. Figure 1). Internally in OWID, there was much discussion about whether this identification was sufficient to assign the entries to a dictionary.
- In the new web design, OWID provides two main outer access possibilities: the main search box and an alphabetical register. The latter, with the corresponding “go to” box, provides the user with a kind of fast outer access like leafing through a book and stopping at a certain place in the alphabet. Entering, for example, “defiz” in this search box, does not lead to a search result, but rather it initiates a look-up in the OWID headword list and immediately displays the best matching entry with the corresponding part of the headword list (in this case “Defizit”, cf. Figure 1). This headword list is a distinctive feature of OWID, because it is a merging of all headwords of all the dictionaries included in the portal. This access option combines with an option for including or excluding individual dictionaries, but the default setting is the inclusion of all dictionaries.<sup>4</sup>
- When you choose a single dictionary from the dictionary bar on the right-hand side of the screen, only the keywords from that dictionary are displayed in the headword list on the left; a view which we call internally the ‘ODO-view’ (‘one-dictionary-only’). Here, too, there was much debate about whether this was easy for users to understand.
- Lastly, the design of the entry itself was changed. We tried to present the information in a more uncluttered way, to divide the lexicographic content more clearly from the labels which classify the items and present comments on items (which occur very often especially in *elexiko*) in a more subtle way. The latter in particular was the subject of heated discussion between

---

<sup>4</sup> In Figure 1, the default setting of the toolbar for filtering the headword list is changed by excluding the non-elaborated entries of *elexiko*.

the web designer and the lexicographers (whether comments should still be presented in boxes and also whether the lighter boxes still attracted too much attention). In addition, the links to the senses of a headword, which consisted of short labels for each sense, were supplemented with the definition in order to provide the user with more information on the first screen (cf. Figure 1).

In the eye-tracking study described below, we have tried to evaluate some of these new design decisions.

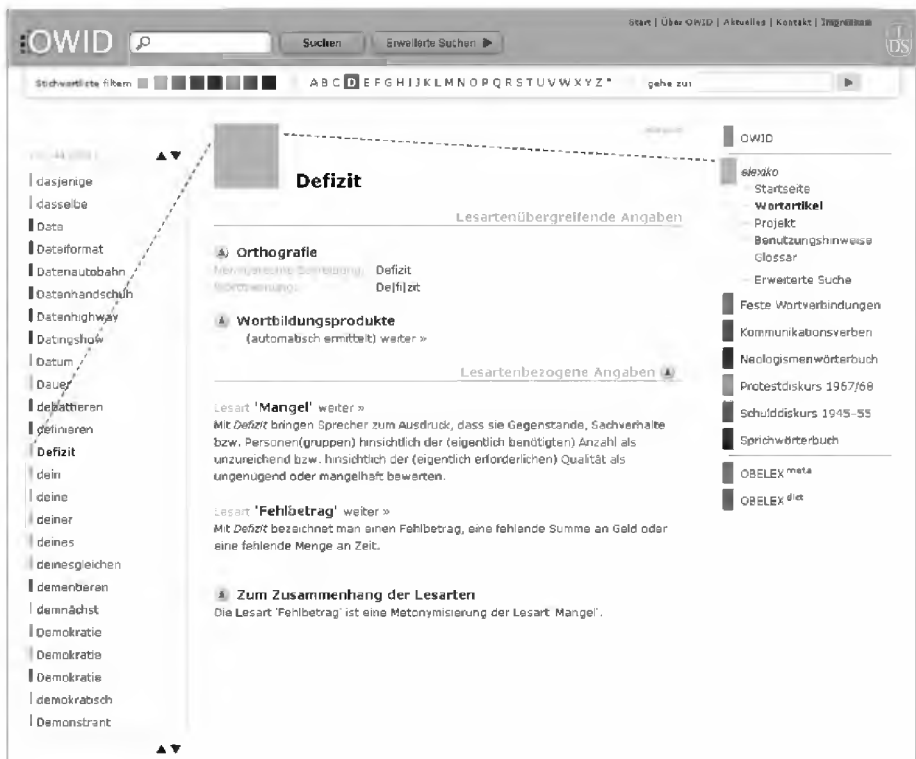


Fig. 1: Entry “Defizit” (‘deficit’) from elexiko with emphasis on the colour scheme for identifying the correlation between dictionary, headword and headword list.

## 4 Study design

### 4.1 Aim

As highlighted at the beginning of this chapter, the main aim of this study was the evaluation of the eye-tracking method with regard to future research projects. A secondary aim was to evaluate different design solutions for the new web design of OWID before we released the new online version. To summarize, we wanted to address the following questions in the study:

- Is it easy to see that OWID is a dictionary portal, i. e. that different dictionaries are integrated into OWID?
- Does the colour scheme work for the identification of the individual dictionaries, i. e. is it easy to assign keywords to the individual dictionaries by the coloured boxes?
- How are new elements of the inner access structure evaluated? In particular, are items easy to locate due to the less cluttered screen layout, and do the participants understand the simultaneous presentation of the sense-label and the definition in *elexiko*?
- Finally, a question about the layout: do the comments in boxes distract users from the items themselves?

### 4.2 Procedure and apparatus

We conducted the study in cooperation with the Mannheim Eye Lab (Uni Mannheim, chair Tracy).<sup>5</sup> This lab offers four computer stations, e. g., for reaction time experiments and two for eye-tracking, equipped with one High-Speed Tracker (SMI Hi-Speed 500) for reading and language processing research and a Remote Eye Tracker (SMI RED) for the study of language-view communication. We used the latter for our study. The setting was highly comfortable and naturalistic for the participants (cf. Figure 2), as also reported by Lew et al. for a similar device:

“Thanks to these features, the Tobii T60 has high ecological validity, offering participants the look and feel of a regular computer screen, thus a highly naturalistic setting for students accustomed to working at the computer.” (Lew et al., 2013, 236)

---

<sup>5</sup> See [http://master.phil.uni-mannheim.de/masterstudiengaenge/master\\_sprache\\_und\\_kommunikation/experimentallabor\\_und\\_mai\\_lab/index.html](http://master.phil.uni-mannheim.de/masterstudiengaenge/master_sprache_und_kommunikation/experimentallabor_und_mai_lab/index.html) (last accessed 13 July 2013).

In contrast, however, Tono states:

“While the eye mark recorder is a powerful tool, the setting inevitably becomes artificial. In order to calculate gaze points accurately, it was necessary to fix the subjects’ head onto the chinrest and ask them to look at the PC monitor.” (Tono, 2011, p. 151)



Fig. 2: Promotional image of the SMI Eyetrackers<sup>6</sup>.

The tasks in our study mostly consisted of two recurring blocks, each divided into three parts:

1. The participants received an instruction, such as: “Please take a look at the screenshot of the OWID homepage and try to gain an initial overview.”
2. The screenshot was presented (and gaze patterns were tracked).
3. In most cases, a question was asked afterwards in order to check whether the requested information had been found.

In the second block:

4. The participants again received an instruction, such as: “In the following, you can again see the OWID homepage . Please try to find out what dictionaries are included in OWID.”
5. The same screenshot (as above in 2) was presented (and gaze patterns were tracked).

<sup>6</sup> See <http://www.gizmag.com/smi-red500-500hz-remote-eye-tracker/16957/picture/124519/> (last accessed 13 July 2013).



6. A question was asked afterwards.<sup>7</sup>

The use of screenshots corresponds to other studies in which dictionary entries were shown in isolated form (Lew et al., 2013; Tono, 2011). However, in the case of OWID, we are dealing with an online portal, where this procedure may be considered to be even more problematic, because the test scenario is very different from a real usage situation, which would involve clicking and browsing (unlike a reading experiment in which the reading takes place on screen instead of on paper, but the test scenario is not fundamentally different from the actual task). Therefore, in terms of a natural setting, it would have been best to use a live version of the portal. However, due to technical limitations and due to the high demands of setting up a usability study with a live system using eye-tracking, this was not possible in our case.

### 4.3 Participants

38 people participated in our study, all of them students aged 20-30. They received €10 as reward for participation. The number of participants is very high for an eye-tracking study, as compared to 10 subjects in Lew et al., 8 subjects in Tono and 6 participants in the study by Simonsen (Lew et al., 2013; Simonsen, 2009; Tono, 2011). We wanted to have a high number of subjects so that we had the option of including randomization tasks and similar things in an appropriate way.

## 5 Results and discussion

### 5.1 Identifying OWID as a dictionary portal

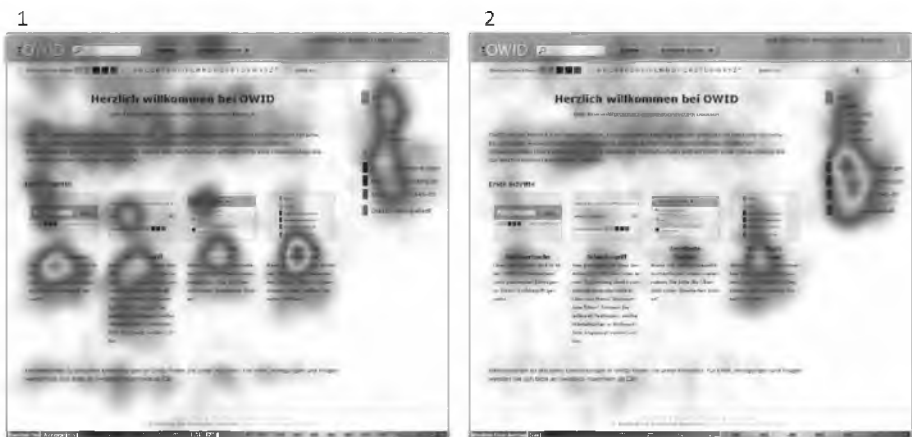
As stated in Section 3, the aim of the new web design was to strengthen the character of OWID as a dictionary portal (as opposed to a single online dictionary). Therefore, we wanted to check whether the participants recognized that the names listed in the right-hand bar were labels of single dictionaries. To examine this question, we gave our participants two instructions, each followed by a screenshot of the OWID homepage .

---

<sup>7</sup> Cf. for a similar approach (various instructions, same picture) an eye tracking study on a painting from the 60s: [http://commons.wikimedia.org/wiki/File:Yarbus\\_The\\_Visitor.jpg](http://commons.wikimedia.org/wiki/File:Yarbus_The_Visitor.jpg) (last accessed 13 July 2013).

1. ‘Please take a look at the screenshot of the OWID homepage and try to gain an initial overview.’ (*Bitte betrachten Sie auf der nächsten Seite einen Screenshot der Startseite von OWID und versuchen Sie, sich dabei kurz zu orientieren.*)
2. ‘In the following, you can again see the OWID homepage. Please try to find out what dictionaries are included in OWID’ (*Sie sehen im Folgenden erneut die Startseite von OWID. Versuchen Sie bitte herauszufinden, welche Wörterbücher in OWID enthalten sind.*)

The results are presented in Figure 3. The cumulative fixation count of all 38 participants is summarized here in the form of a heat map.<sup>8</sup>



**Fig. 3:** Heat map for all participants; (1) instruction: gain an overview of the homepage, (2) instruction: find out what dictionaries are included.

It is clear from the heat map that, after reading the first instruction, participants looked at all the pictures and texts on the OWID homepage. The results for after the participants read the second instruction are different: here, most concentrated on the right-hand bar (cf. Figure 3). This can be interpreted as a confirmation that participants recognized correctly that the included dictionaries are listed on the right-hand side of the screen. However, to make this interpretation reliable, we should have added more screens in which the dictionaries were listed in another position and/or with other elements (not names of dictionaries) on the right (cf. Section 6).

<sup>8</sup> A heat map is a “static representation, mainly used for the agglomerated analysis of the visual exploration patterns in a group of users [...]. In these representations, the ‘hot’ zones or zones with higher density designate where the users focused their gazes with a higher frequency.” ([http://en.wikipedia.org/wiki/Eye\\_tracking](http://en.wikipedia.org/wiki/Eye_tracking)) (last accessed 13 July 2013).

Without telling the participants in advance that we were going to do so, we asked them afterwards whether the following dictionaries were included in OWID:

- Neologismenwörterbuch
- Woxikon-Synonym-Wörterbuch
- PONS Deutsche Rechtschreibung
- elexiko
- Schulddiskurs 1945-55
- Feste Wortverbindungen
- OBELEX Bibliografie
- Weiß nicht / keine Angabe.

Only 32% of the participants chose all the right answers (cf. Table 1). They obviously concentrated their views on the right-hand bar of the screen (and therefore on the correct list), but did not remember all the items in this list correctly. However, we have to admit that the dictionaries in OWID have very unusual titles and they were combined in the response items with very popular German online dictionaries. Therefore, a clear interpretation of this latter result is difficult.

	Freq	Percent	Cum.
incorrect	26	68.42	68.42
correct	12	31.58	100.00
<b>Total</b>	<b>38</b>	<b>100.00</b>	

Tab. 1: Percentages of correct responses to the question ‘Which dictionaries are included in OWID?’.

## 5.2 Assignment of headwords to a dictionary

Another question we wanted to examine was whether it is possible for the participants to assign keywords to a dictionary only via the preceding coloured box, because the question of whether this little box was sufficient for this purpose was the subject of much discussion. In order to evaluate this, we chose the entry “auf ein gesundes Maß reduzieren” (‘reduce to a healthy level’) from “Feste Wortverbindungen” (‘Collocations online’) and gave the participants the following two instructions, each of them followed by a screenshot and questions.

1. ‘You are looking at a dictionary entry. Which headword is being described?’ (*Sie sehen gleich einen Wörterbuchartikel. Welches Stichwort wird beschrieben?*)
2. ‘You are now looking at the same dictionary entry again. Which dictionary is the entry from?’ (*Sie sehen jetzt noch einmal den gleichen Wörterbuchartikel. Aus welchem Wörterbuch stammt der Artikel?*)

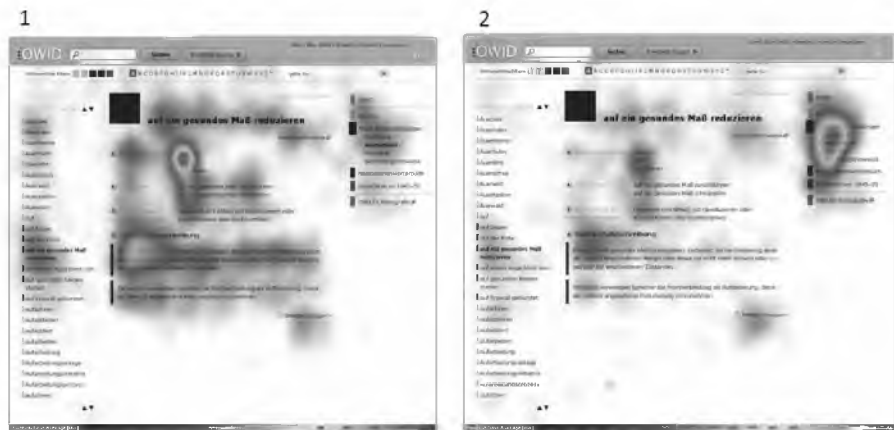


Fig. 4: Heat map for all participants; (1) instruction: which headword, (2) instruction: headword from which dictionary.

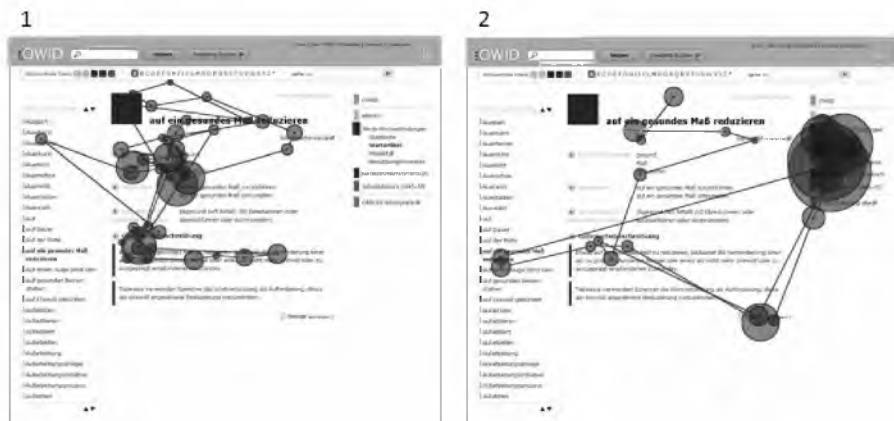


Fig. 5: Scan paths of two participants; (1) instruction: which headword, (2) instruction: headword from which dictionary.

The results are presented in Figures 4 and 5. Figure 4 shows that, while after the first instruction, participants concentrated on the middle of the screen, i. e. the entry itself, their fixation moved after the second instruction to the right-hand side where the name of the dictionary is presented. This is also illustrated in Figure 5, where the scan paths of two participants are presented. It appears that the connection between headword and dictionary is clear for our subjects. The connection to the same-

coloured words in the keyword list on the right can also be recognized in the scan path (Figure 5, screenshot 2). We repeated the same task with the entry “Ärztin” ([female] doctor); the results are comparable. However, we did not check these results against other screens with the same content, but different layouts.

### 5.3 Assignment of headwords from the headword list to a dictionary

In the OWID headword list, all the headwords of all the dictionaries integrated into OWID are merged together. A small coloured box preceding the headword indicates which dictionary it belongs to. Depending on how you set the function ‘Stichwortliste filtern’, or whether a dictionary is accessed via the dictionary list on the right-hand side, the keyword list changes, e. g., to the ‘ODO-view’ (one-dictionary-only). In our study, we wanted to examine whether participants recognized this difference, more specifically whether they could see that the headwords in the keyword list belonged to different dictionaries. We tried to approach this question by giving the following instructions:

1. ‘On the next page, you will see an entry from the dictionary of neologisms. Neologisms are new words or new meanings of established words which have entered the German language. Please look at the screenshot and try to familiarize yourself with it.’ (*Sie sehen auf der nächsten Seite einen Wortartikel aus dem Neologismenwörterbuch. Neologismen sind neue Wörter oder neue Bedeutungen etablierter Wörter, die in die deutsche Sprache eingegangen sind. Bitte betrachten Sie den Screenshot und versuchen Sie, sich dabei kurz zu orientieren.*)
2. ‘On the next page, you will again see the same entry from the dictionary of neologisms. Can you find more headwords from the dictionary of neologisms on the OWID page?’ (*Sie sehen auf der nächsten Seite noch mal den gleichen Wortartikel aus dem Neologismenwörterbuch. Finden Sie auf der auf der abgebildeten Seite von OWID weitere Stichwörter aus dem Neologismenwörterbuch?*)

In this case, we were not interested in the difference between the participants’ gaze patterns after the first and second instruction. Rather, after the second instruction, we formed (at random) two groups to which we presented two different screens: the first with a list on the left-hand side with headwords from different OWID dictionaries (headwords from the neologism dictionary positioned in the centre of the list), the second with headwords only from the neologism dictionary. Then, we wanted to see whether participants had different gaze patterns depending on the different content of the headword list (after reading the second instruction). The results are presented in Figure 6.

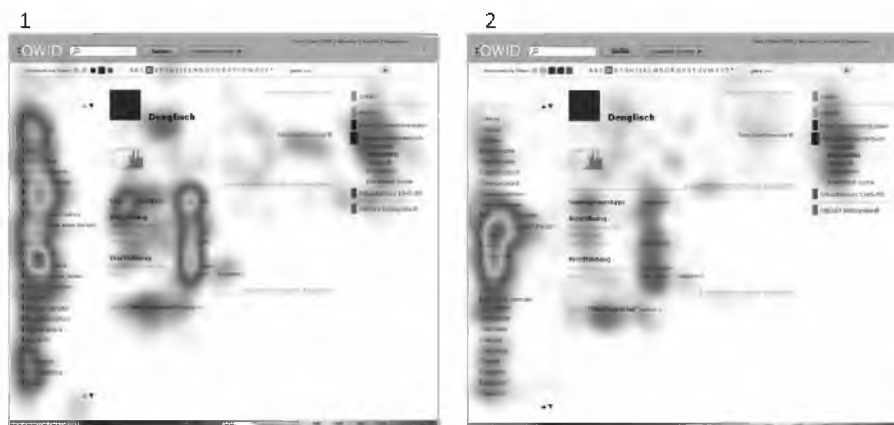


Fig. 6: Heat map for all participants; (1) ODO-view (neologisms only), (2) headwords from different dictionaries (neologisms in the centre).

The gaze patterns suggest that participants understood that only the headwords preceded by the small blue box were headwords from the neologism dictionary, i. e. that the presentation of the headword list in the new layout works well. It could be argued that maybe participants did not identify the neologisms by the blue coloured box, but by understanding neologisms as a special type of lexeme. This interpretation, in turn, is not supported by the response behaviour. After the screenshots had been presented, both groups had to answer the following question: ‘Please name more headwords which in your opinion are from the dictionary you have just seen. Please click on all the options you think are correct.’ (*Bitte nennen Sie weitere Stichwörter die Ihrer Meinung nach aus dem eben gesehenen Wörterbuch stammen. Bitte klicken Sie alle Alternativen an, die Ihrer Meinung nach richtig sind.*):

- denkbar
- Demokratie
- den Ball flach halten
- der ganz normale Wahnsinn
- Weiß nicht / keine Angabe

Only 12% of the participants chose all the correct answers (cf. Table 2). Therefore, the interpretation that the coloured box was the guiding element for recognition seems to be more plausible.

	Freq	Percent	Cum.
incorrect	26	68.42	68.42
correct	12	31.58	100.00
Total	38	100.00	

**Tab. 2:** Percentages of correct responses to the question ‘Do the keywords presented belong to the neologism dictionary?’.

## 5.4 Some questions on inner access structures

### 5.4.1 Navigation to sense-related items in *ellexiko*

In *ellexiko* (cf. Klosa et al., this volume), sense-independent items such as orthographic information or (in most cases) information on word formation is presented on the first screen if you open an entry. Sense-relevant information follows on a second screen when clicking on short labels for the individual senses. In the old layout, only these short labels were presented on the first screen. In our study, we wanted to look at how participants recognize this information (label and definition). Essentially, we wanted to explore what the gaze patterns of the subjects looked like when we ask them questions about individual meanings (e. g., Could they find the individual meanings? Did they read or scan all the labels first and only then look at the definition? Or is this – although this seems rather implausible – a linear reading process?). For sense navigation, especially in printed dictionaries, studies in the field of research into dictionary use already exist (Lew et al., 2013, for a summary of the results of different studies see 230-32; e.g., Lew & Tokarek, 2010; Lew, 2010; Nesi & Tan, 2011; Tono, 2001, 2011). However, our study design is rather different, with no specific research question, and is therefore not comparable with previous results.

First, we instructed the participants to see if the entry ‘horse’ (Pferd) had a sense like ‘apparatus used in gymnastics’ (Turngerät): ‘On the next page, you will see an entry from *ellexiko*. Please try to find out whether the headword can have the meaning ‘apparatus used in gymnastics’ (*Sie sehen auf der nächsten Seite einen Wortartikel aus ellexiko. Bitte versuchen Sie herauszufinden, ob das Stichwort eine Bedeutung/Lesart im Sinne von ‘Turngerät’ hat.*)

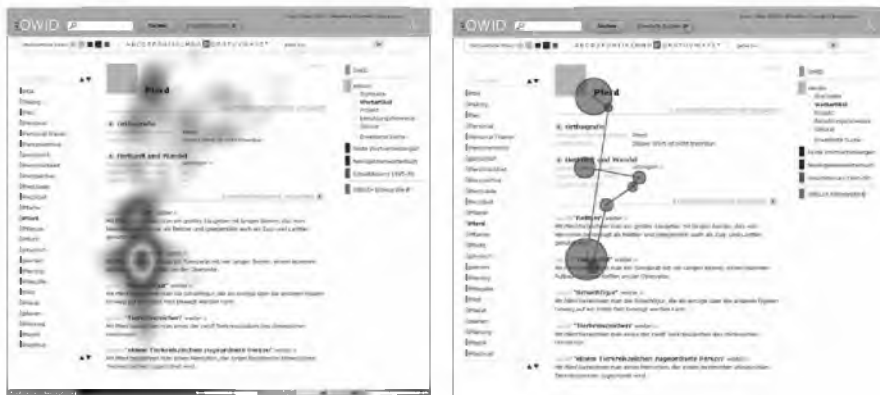


Fig. 7: Heat map of all participants, scan path of one participant (finding the sense labelled 'Turngerät').

The results are presented in Figure 7. It appears that the fixation is clearly focused on the requested sense; the scan path of one participant also illustrates this. It could be that the task was too simple and clear, and therefore a different result would have been very unlikely. Secondly, we asked participants to find a sense in the entry 'team' (Mannschaft) which is described as 'members of a group of people acting on behalf of an organization' ('Please try to find out whether in the following word entry, there is a meaning which is explained as 'members of a group of people who work for an organization'. If so, which is it? 'Bitte versuchen Sie herauszufinden, ob es im folgenden Wortartikel eine Bedeutung gibt, die erläutert ist mit 'Mitglieder einer für eine Organisation tätige Gruppe von Menschen'. Wenn ja, welche?'). The corresponding results are presented in Figure 8.

The interesting thing here is that participants obviously first scan all the labels very quickly and then turn to the content of the definition, even though the instruction clearly draws attention to the content. This suggests that the labels attract significant attention. For our online presentation, however, this is not an adverse effect. All in all, we can state that in the study, our participants found the appropriate meaning by firstly scanning the labels and then reading the definition, if necessary.



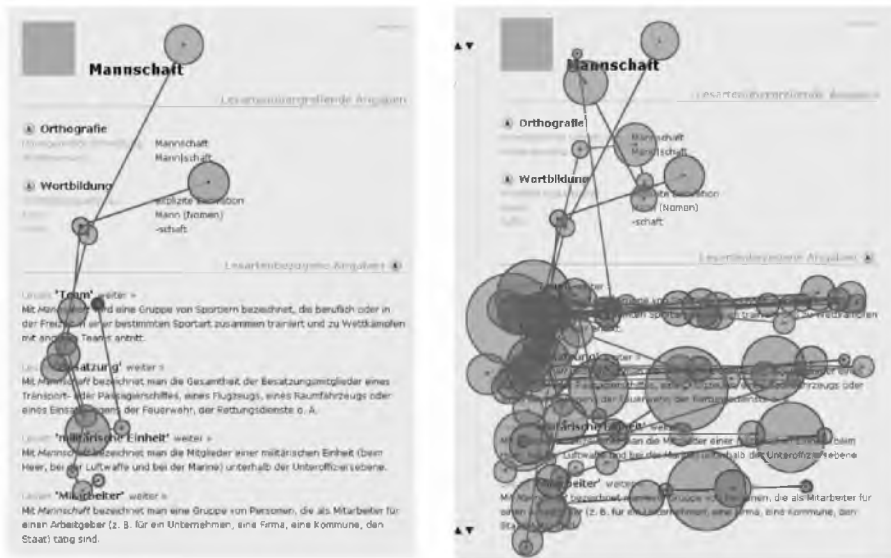


Fig. 8: Scan paths of two participants (recorded as a film), one snapshot at 00:01 sec (left) and the second at 00:14sec (right).

#### 5.4.2 Access via search paths to specific items

With the new OWID design, we tried to present the information in a more uncluttered way, and in particular to divide the lexicographic content more clearly from the labels used to classify the items. The aim of this design decision (as well as creating an uncluttered look) was to increase the internal access time to specific items, i. e. to ensure that it was easy to find the required information in another entry after an initial orientation. To check whether we had put this into practice successfully, we incorporated two screenshots of entries from the neologism dictionary ('Antimatschtomate' and 'angefressen') into our study, both containing items on style level. Here, we wanted to examine whether access to the items on style level was significantly faster in the second task than in the first. The following instructions preceded the screenshots:

1. Entry ‚Antimatschtomate‘: ‚In dictionaries, individual headwords are often assigned to particular styles, such as colloquial, elevated, slang, etc. To which style does the following headword belong, according to the dictionary of neologisms?‘ (*Häufig werden in Wörterbüchern einzelne Stichwörter bestimmten Stilebenen zugeordnet wie umgangssprachlich, gehoben, salopp etc. Zu welcher Stilebene gehört das folgende Stichwort nach Angabe des Neologismenwörterbuchs?*)

2. Entry ‚angefressen‘: ‘To which style does the following headword belong, according to the dictionary of neologisms?’ (*Zu welcher Stilebene gehört das folgende Stichwort nach Angabe des Neologismenwörterbuchs?*)

In the result, we can see a learning curve: while the subjects still needed an average of 5 seconds to find the information in the first task, less than half this time was needed for the second task (2.3 s). This faster access is also clearly visible in the example of the scan paths of two subjects, where in the second task, a clearly more focused search path can be seen (cf. Figure 9). The learning effect seems to confirm that the new layout is useful and clear and allows quick access to the information required. Again, in retrospect, it must be said that the learning effect itself is not surprising and that we should also have integrated the old layout as a control test to be in a position to say more reliably that the new layout is the key factor here. The question asked afterwards, about which style level the entries belong to, was answered correctly by most participants, i. e. the items were perceived correctly (cf. Table 3).

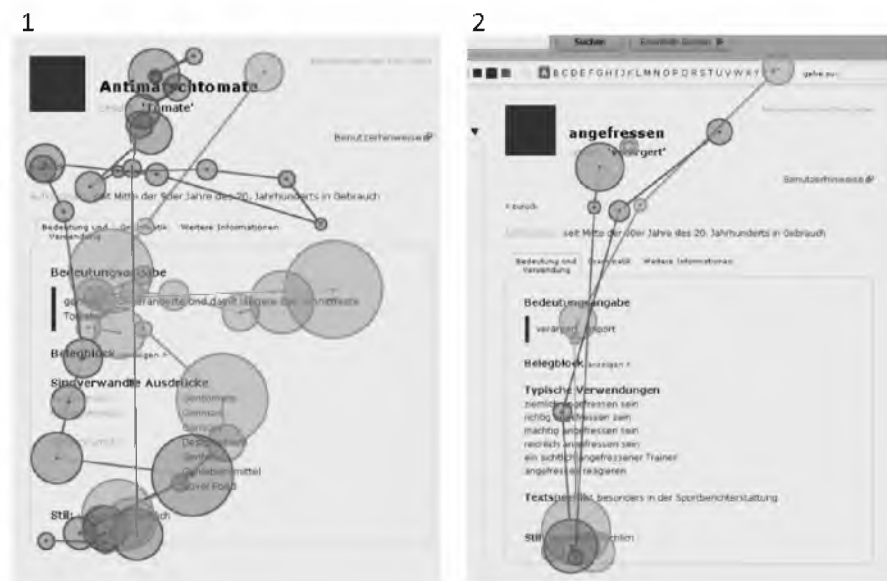


Fig. 9: Scan paths of two participants (searching for items on style level).

	<i>Antimatschtomate</i>			<i>angefressen</i>		
	Freq	Percent	Cum.	Freq	Percent	Cum.
<b>correct</b>	36	94.74	94.74	37	97.37	97.37
<b>incorrect</b>	1	2.63	97.37			
<b>don't know/ no answer</b>	1	2.63	100.00	1	2.63	100.00
<b>Total</b>	38	100.00		38	100.00	

**Tab. 3:** Percentages of correct responses to the question regarding the style level of ‘Antimatschtomate’ and ‘angefressen’.

### 5.4.3 Potential distraction by comments

One important part of the conception of *lexiko* was to provide the users with a lot of comments on lexicographic items whenever it might be helpful. These comments contain additional information which may be interesting to some users. Therefore, they should be visible, but it would prove counterproductive if these comments distracted the user’s attention from the items themselves. For this reason, we decided to present comments on items in the new layout in a more subtle way (e. g., in lighter boxes). As mentioned in Section 3, finding an appropriate form of presentation was the subject of much discussion between the web designer and the lexicographers. The main question was whether the presentation in boxes (even if they are lighter) is still too eye-catching and draws the user’s attention first of all to the comment instead of to the item. To evaluate this in our eye-tracking study, we presented two entries with items on word formation, the first without a comment and the second with an additional comment, both with the same instruction: ‘Please ascertain which components the following word is made up of.’ (*Bitte ermitteln Sie, aus welchen Bestandteilen das folgende Wort gebildet wird.*). Here, we wanted to see whether the gaze patterns showed a different focus in the second entry in contrast to the first one.

The results in the form of a heat map are presented in Figure 10. The gaze patterns do not show that the comment attracted much attention. Also, the scan paths of several participants confirmed that the subjects did not look at the box first (Figure 11). The comments as they are presented in the new layout seem not to distract users from the items. A limiting effect here could be that, during the course of the study, the subjects got used to us asking questions after each screenshot and therefore their attention was drawn to the item requested in the instruction.



Fig. 10: Heat map of all participants: items on word formation (one with a comment on word formation, one without).

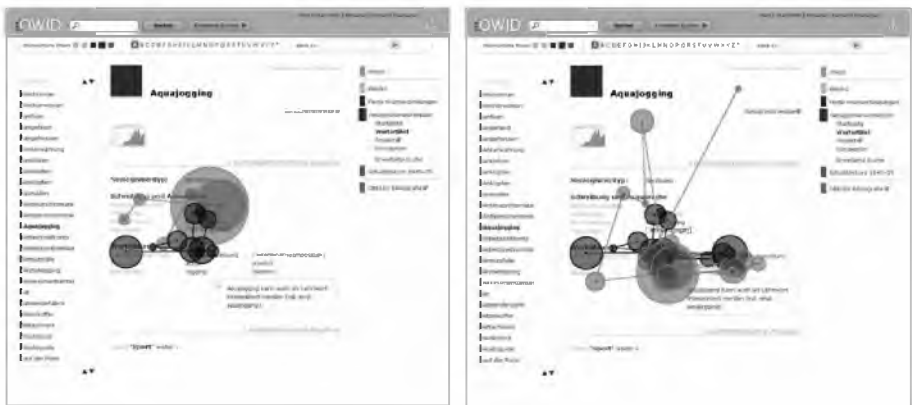


Fig. 11: Scan paths of four participants looking at the screenshot of the entry 'Aqua jogging'.

## 6 Critical comments on the study

In retrospect, we must say that the results presented here are of limited validity, because the study suffers from some methodological shortcomings measured against the standards of good empirical studies. Due to these methodological shortcomings and the vague interpretation of the results, there might be good reasons not to publish the results. Nevertheless, we wanted to provide the results here, because – as also emphasized by Lew et al. – there are few eye-tracking studies in the area of research into dictionary use and therefore other researchers planning an eye-tracking study might benefit from our experience.

With the experience we have now, we consider the following to be the main shortcomings:

- The research question was not the guiding element in designing the study. The specific research question should also be planned more carefully than we did in this study (Lew, 2011, 228-29).
- The questions were not tailored enough to eye-tracking as a method of data collection; they should have been more focused on what works particularly well with this technology. For example, we should have integrated many more comparative views in order to check the positive impression against other kinds of layout. With our study design, we cannot exclude the possibility that the old layout may have performed as well as the new one in the study.
- The questions asked in retrospect can also cause problematic effects as also pointed out by Lew et al. Although, in our study design, the subjects did not have to look away from the screen, but were able to answer the questions on the screen, the questions might have caused a guidance effect because the subjects knew that they were going to get ‘test’ questions afterwards, and their gaze patterns might have been influenced by this (although eye movements are difficult to control).

“We had rejected the option of asking the participants to write down the answers themselves, as this would have made them look away from the monitor and might have disrupted the gaze recording. We did not want to ask them to give the sense number itself, as this might have made them too aware of the sense selection aspect.” (Lew et al., 2013, 237)

Although this study had clear methodological shortcomings, we have learned a lot. This is not surprising, because to gain experience is (also) to learn by making mistakes. We hope, therefore, that it is useful to make these results and experiences available in this open manner.

## Bibliography

- Duchowski, A. (2007). *Eye Tracking Methodology: Theory and Practice*. London: Springer-Verlag London Limited. Retrieved from <http://dx.doi.org/10.1007/978-1-84628-609-4>
- Engelberg, S., & Müller-Spitzer, C. (forthcoming). Dictionary Portals. In R. H. Gouws, U. Heid, W. Schweickard, & H. E. Wiegand (Eds.), *Dictionaries. An international encyclopedia of lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlin/New York: De Gruyter.
- Just, M. A., & Carpenter, P. A. (1980). A Theory of Reading: From Eye Fixations to Comprehension, *87*(4), 329–354.
- Lew, R. (2010). Users Take Shortcuts: Navigating Dictionary Entries. In A. Dykstra & T. Schoonheim (Eds.), *Proceedings of the XIV Euralex International Congress* (pp. 1121–1132). Ljouwert: Afûk.
- Lew, R. (2011). User studies: Opportunities and limitations. In K. Akasu & U. Satoru (Eds.), *ASIALEX2011 Proceedings Lexicography: Theoretical and practical perspectives* (pp. 7–16). Kyoto: Asian Association for Lexicography.
- Lew, R., Grzelak, M., & Leszkowicz, M. (2013). How Dictionary Users Choose Senses in Bilingual Dictionary Entries: An Eye-Tracking Study, *Lexikos* 23, 228–254.
- Lew, R., & Tokarek, P. (2010). Entry menus in bilingual electronic dictionaries. *eLexicography in the 21st century: New challenges, new applications*. Louvain-la-Neuve: Cahiers du CENTAL, 145–146.
- Müller-Spitzer, C. (2010). OWID – A dictionary net for corpus-based lexicography of contemporary German. In A. Dykstra & T. Schoonheim (Eds.), *Proceedings of the XIV Euralex International Congress* (pp. 445–452). Leeuwarden/Ljouwert: Fryske Akademy. Retrieved from [http://www.euralex.org/elx\\_proceedings/Euralex2010/026\\_Euralex\\_2010\\_1\\_MULLER-SPITZER\\_OWID\\_A%20dictionary%20net%20for%20corpus-based%20lexicography%20of%20contemporary%20German.pdf](http://www.euralex.org/elx_proceedings/Euralex2010/026_Euralex_2010_1_MULLER-SPITZER_OWID_A%20dictionary%20net%20for%20corpus-based%20lexicography%20of%20contemporary%20German.pdf)
- Nesi, H., & Tan, K. H. (2011). The Effect Of Menus And Signposting On The Speed And Accuracy Of Sense Selection. *International Journal of Lexicography*, 24(1), 79.
- Simonsen, H. K. (2009). *Vertical or Horizontal? That is the Question: An Eye-Track Study of Data Presentation in Internet Dictionaries*. Copenhagen: Copenhagen Business School.
- Simonsen, H. K. (2011). User Consultation Behaviour in Internet Dictionaries: An Eye-Tracking Study. *Hermes. Journal of Language and Communication Studies*, 46, 75–101.
- Tono, Y. (2001). *Research on dictionary use in the context of foreign language learning: Focus on reading comprehension*. Tübingen: Max Niemeyer Verlag.
- Tono, Y. (2011). Application of Eye-Tracking in EFL Learners. *Dictionary Look-up Process Research*. *International Journal of Lexicography*, 23.