

# **The dynamics outside the paper: user contributions to online dictionaries**

**Andrea Abel<sup>1</sup>, Christian M. Meyer<sup>2</sup>**

<sup>1</sup>EURAC, Drususallee 1, Bolzano, Italy

<sup>2</sup> Ubiquitous Knowledge Processing Lab (UKP-TUDA),  
Technische Universität Darmstadt, Germany

E-mail: andrea.abel@eurac.edu, meyer@ukp.informatik.tu-darmstadt.de

## **Abstract**

Online dictionaries rely increasingly on their users and leverage methods for facilitating user contributions at basically any step of the lexicographic process. In this paper, we propose a novel classification of the different types of user contributions, which have not been systematically studied so far. With the help of many practical examples, we discuss three major types of user contributions and discuss multiple forms and implementations of them: (i) Direct user contributions, which comprise dictionary articles written entirely or partly by users in a collaborative effort; (ii) Indirect user contributions, which occur in different forms of explicit feedback (e.g., by e-mail or web forms) and implicit feedback through log file analysis or external user-generated content; (iii) Accessory user contributions, which go beyond the dictionary content by initiating an exchange either between the dictionary makers and their users or among the users themselves. We argue that the ease of communication and collaboration between dictionary makers and users has enormous potential, not only for keeping the dictionary up to date and of high quality, but also for developing improved, user-adapted views of, and access to, the contents of a dictionary. Studying the different types of user contribution is crucial for effectively planning online dictionaries and for future research on electronic lexicography.

**Keywords:** internet lexicography; online dictionaries; user contributions; collaborative lexicography

## **1. Motivation**

The World Wide Web offers various possibilities for users to contribute to dictionaries. These range from giving feedback or correcting errors to creating new dictionary articles and discussing language-related issues beyond the explicitly encoded knowledge. The ease of communication and collaboration between dictionary makers and users has enormous potential, not only for keeping the dictionary up to date and of high quality, but also for developing improved, user-adapted views of, and access to, the contents of a dictionary.

The discussion on user contributions in lexicography is mainly linked to online dictionaries, but is not new as even print dictionaries may be strongly based on collaboration with the public. The *Oxford English Dictionary*, for example, conducted reading programs right from its inception in the 19<sup>th</sup> century to collect quotations illustrating how words are used (cf. Thier, 2013).

However, the dynamics outside the paper are obviously different as they facilitate a greater variety of user contributions as well as immediate publication and timely feedback. With the rise of social media technologies (e.g., blogs, wikis, social networks) and the Web 2.0, users can actively participate in the compilation of a dictionary. In fact, we face a new kind of lexicographical process in which the formerly clear distinction between dictionary editors and dictionary users becomes increasingly blurred. This is also captured by the neologism *prosumer*, a blend of *producer* and *consumer* (cf. Lew, 2013). Carr (1997) describes this change of lexicographic paradigms as *bottom-up lexicography* according to which dictionaries are “evolving upward from readers” — as opposed to *top-down lexicography* “from editors, through publishers, to readers”.

For the first time, we systematically study the different types of user contribution backed by multiple practical examples found in existing online dictionaries. Our analysis takes into account both individual dictionaries (e.g., the *Oxford English Dictionary*, *Duden online*) and dictionary portals, such as *LEO*, *dict.cc*, and *canoonet* (cf. Storrer, 2010; Engelberg & Müller-Spitzer, in print). As a result of our work, we propose a classification for describing the dynamics induced by user contributions. At the top level, we distinguish the following three types of user contribution:

- (i) Direct user contributions
- (ii) Indirect user contributions
- (iii) Accessory user contributions

Obviously, a single dictionary project may utilize different types of user contributions at the same time. Therefore, we provide a general, dictionary-independent classification instead of focusing on a specific project. In the paper, we first discuss related work in this area and then describe each of the three types of user contribution in detail. Table 1 shows an overview of our proposed classification. We conclude the paper with a final discussion and a summary of our findings.

<b>Direct user contributions</b>	<b>Indirect user contributions</b>	<b>Accessory user contributions</b>
<ul style="list-style-type: none"> <li>• Contributions to open-collaborative dictionaries</li> <li>• Contributions to collaborative-institutional dictionaries</li> <li>• Contributions to semi-collaborative dictionaries</li> </ul>	<ul style="list-style-type: none"> <li>• Explicit feedback                             <ul style="list-style-type: none"> <li>– form-based feedback</li> <li>– free form feedback</li> </ul> </li> <li>• Implicit feedback                             <ul style="list-style-type: none"> <li>– log file analysis</li> <li>– external user-generated content</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Exchange between dictionary makers and dictionary users                             <ul style="list-style-type: none"> <li>– unidirectional communication</li> <li>– bidirectional communication</li> </ul> </li> <li>• Exchange among dictionary users</li> </ul>

Table 1: Overview of our functional classification of user contributions to online dictionaries

## 2. Related work

The earliest descriptions of user contributions to electronic dictionaries date back to the mid 1990s. In his well-known article, Carr (1997) introduces the terms *bottom-up lexicography* and *collaborative lexicography*, without further differentiating between them. Although Carr predominantly addresses the submission of dictionary articles or additions by e-mail, both expressions serve nowadays as umbrella terms for different types of user contributions. This also applies to other expressions that are, more or less, synonymously used to describe any type of user contribution, including *user involvement* (Lew, 2011), and contributions based on *user-generated content* (Lew, 2013).

Storrer (1998) distinguishes different types of *user participation* targeted at (i) correcting errors, (ii) identifying gaps, (iii) obtaining expert contributions on certain topics, and (iv) collecting contributions by laypeople in an entertaining and playful setting. In subsequent work, Storrer (2010) focuses on the distinction between dictionaries allowing for user contributions controlled by professional editors and dictionaries created by the users themselves in a collaborative effort.

Køhler Simonsen (2005) describes the evolution from lexicographic products to lexicographic services, which raises an increasing need for involving the users in every stage of the lexicographic process. To this end, he proposes two principles to facilitate user contributions in a specialized dictionary, and he associates each phase of the lexicographic process with the corresponding principles and objectives. By *active user involvement*, he refers to feedback on the design and the development of a dictionary by means of surveys or test groups. On the other hand, *lexicographic democracy* describes feedback on the dictionary articles and the quality of the lexicographic descriptions (e.g., submitting error corrections). The proposed classification is, however, limited to indirect user contributions, as Køhler Simonsen (2005) explicitly excludes the possibility of modifying the dictionary articles directly, as is the case, for example, in collaborative dictionaries. He argues in particular that each user contribution should be subject to editorial control.

Thus, Køhler Simonsen's definition of democracy is not to be confused with the use of *democratization* elsewhere. Fuertes-Olivera (2009), for instance, considers democratization as a result of *collective free multiple-language internet reference works* such as *Wikipedia* and *Wiktionary*, which are entirely compiled by users – without editorial control. He distinguishes them from *institutional internet reference works* that are offered by professional publishers.

A similar distinction is made by Lew (2011), who additionally introduces *collaborative-institutional dictionaries*, which, according to him, lie in between collective-free and institutional dictionaries. This type of dictionary is offered by professional publishers, but allows for direct user contributions.

Lew (2013) discusses multiple dictionary projects along the dimension of their degree of *user-generated content*. This ranges from lexicographic works that entirely consist of user-generated content (*collaborative dictionaries*) to a combination of user-generated content and professional content (comparable to concepts such as *semi-collaborative* [Melchior, 2012], or *user participation* [Storrer, 2010]), and works in which professional content dominates. Lew (2013), in line with Rundell (2012), sees potential in the combination of user-generated and professional content – especially for certain vocabulary types.

Melchior (2012; 2013) introduces the term *semi-collaborative* for his analysis of the *LEO* dictionary portal. He defines a semi-collaborative dictionary as being *supported by users* rather than *generated by users*. Thus, Melchior's use of the term relates to improving and extending existing content, as well as expanding and developing the dictionary project as a whole.

Though it is mostly discussed in the context of the quality of lexicographic products, *simultaneous feedback* (De Schryver & Joffe, 2004; De Schryver & Prinsloo, 2000) represents an important concept when thinking about user contributions, because it initiates a large amount of feedback implicitly and explicitly uttered by users. For printed dictionaries, this means releasing small-scale dictionaries, which are used to collect suggestions for a main dictionary that is being compiled in parallel (De Schryver & Prinsloo, 2001). For the electronic adaptation *fuzzy simultaneous feedback*, De Schryver & Joffe (2004) replace the traditional means of getting feedback (e.g., using questionnaires) with the generation of free implicit feedback, based on log file analysis. From the perspective of user contributions, (fuzzy) simultaneous feedback is similar to the proposal by Køhler Simonsen (2005) introduced above, in the sense that feedback occurs during different phases of the lexicographic process (cf. De Schryver & Prinsloo, 2000).

Recent studies of user contributions have become increasingly detailed. However, a comprehensive and systematic classification is still missing. Rather, there has been a variety of ambiguous and partly overlapping terms, which hampers the effective planning of forms of user contributions for new and established dictionaries. A particular problem is that most previous works are focused on one specific type of user contribution, for example, focusing on the degree of editorial control or discussing different types of feedback.

In his analysis of 88 online dictionaries according to various criteria, Mann (2010) lists three possible types of user contribution. First, *direct contributions* to the dictionary, including the compilation as well as the modification of articles. Second, *indirect contributions*, including the option to give feedback by means of forms, contact addresses, etc., which inherently implies a form of editorial control. Third, the *exchange with other dictionary users* by means of online forums. This classification comprises both collaborative approaches and user contributions based on feedback.

However, Mann (2010) provides little detail of the individual types of user contribution and omits, for instance, the forms of discourse between the dictionary makers and users, which we discuss in section 5. The goal of our contribution is therefore to classify the previously discussed dimensions of user contributions and close the gaps between existing classifications. We use the three types of user contribution proposed by Mann (2010) as a starting point.

### 3. Direct user contributions

By *direct user contributions* we refer to additions, modifications, and deletions of dictionary articles or parts of them performed by a dictionary user. We can distinguish between direct user contributions to open-collaborative, collaborative-institutional, and semi-collaborative dictionaries.

*Contributions to open-collaborative dictionaries* are neither constituted nor controlled by a predefined group of experts. Rather, the descriptions in the corresponding dictionaries are completely built by the users themselves. The open-collaborative approach has become particularly popular with the rise of the free online encyclopedia *Wikipedia*, in which users write and edit encyclopedic articles that are immediately published on the Web. Instead of expert knowledge, these user contributions are backed by the collective intelligence of a large number of authors, which has often been described as the “wisdom of crowds” (Surowiecki, 2005). According to Malone et al. (2010), the motivation for contributing to open-collaborative projects can be characterized by *money* (including any type of economic benefit and the training of personal skills), *love* (enjoyment, altruism, socializing with others), and *glory* (receiving recognition from peers).

Most open-collaborative dictionaries are based on fixed lexicographic instructions and a predefined article microstructure. The *Urban Dictionary* is one example of this, as the scope of the dictionary is made clear (i.e., slang, jargon, nonce words, and the like) and contributions are organized in a fixed web form asking for the word, a definition, example usages, and a number of keywords. Many dictionaries of this type focus on translations, for example, *bab.la* or *Glosbe*, as they are easy to model and usually only require fields for the term in the source and the target languages. Multilingual dictionaries particularly benefit from direct user contributions because of the broad diversity of the language pairs of contributing users (cf. Meyer & Gurevych, 2012).

More complex open-collaborative dictionaries that aim at compiling a general language dictionary, such as the *Kamusi project*, require extensive user interfaces to represent all encoded information types. While the majority of these dictionaries provide a dictionary-specific user interface, some of them are based on the wiki technology, such as *Wiktionary* or the *Rap Dictionary*. Wiki-based dictionaries are usually not based on fixed lexicographic instructions and a predefined microstructure.

They rather define a markup language with which the microstructure can be defined individually for each dictionary article (e.g., using bold face for encoding parts of speech). Matuschek et al. (2013) compare user contributions to a dictionary with a fixed microstructure (*OmegaWiki*) and with a loosely defined microstructure (*Wiktionary*). They find that a fixed microstructure limits expressiveness, because complex information types such as verb argument structures or hierarchically-organized word senses are often not modeled and are too complicated to add later on. The structural openness of *Wiktionary*, however, yields inconsistencies in the layout of the articles, and this hampers the fast and efficient use of the dictionary.

Since user contributions to open-collaborative dictionaries are not moderated by professional editors, they are subject to two types of quality-related flaws: (i) spam and vandalism, and (ii) unspecific, incorrect, outdated, oversimplified, or overcomplicated descriptions. In larger projects, there is hence a need for quality assurance measures. *Wiktionary*, for instance, recently introduced the *flagged revisions* feature for some of its language editions. A flagged revision marks a certain version of an article as having accomplished a basic quality standard. Permission to indicate an article as a flagged revision is only granted to active contributors after having edited at least 200 articles. So far, the flagged revisions indicate that an article is at least free of spam (type (i) flaws), but the feature also generally enables a distinction between a *sighted flag* (type (i) flaws) and a *quality flag* (type (ii) flaws).<sup>1</sup>

In addition to that, *requests* are another quality assurance measure in *Wiktionary*. If a contributor notices a quality flaw, which (s)he cannot resolve immediately, a colored “request” banner may be added to the article stating a need for verification (e.g., the addition of sources), extension (e.g., the addition of an example sentence), clean up (in terms of content and format), or deletion of an article.

A second type of direct user contribution is *contributions to collaborative-institutional dictionaries* (cf. Lew, 2011). These dictionaries are provided by major dictionary publishers, for example, the *Merriam-Webster Open Dictionary*. The motivation for a company to publish a collaborative-institutional dictionary is to collect evidence and suggestions for improving editorial dictionaries and to keep dictionary users interested in the publisher’s activities and products. Contributions to collaborative-institutional dictionaries may address arbitrary vocabulary as in the *Macmillan Open Dictionary*, or focus on a narrower scope, such as Duden’s *Szenesprachenwiki* for neologisms.

Typically, contributions are in the form of full dictionary articles, which are checked for spam, personal offense or defamation before being published. They are, however, not edited on a large scale, as is the case for semi-collaborative and indirect user

<sup>1</sup> <http://meta.wikimedia.org/w/index.php?oldid=5434621> (27 April 2013)

contributions (see below). Unlike contributions to open-collaborative dictionaries, the users cannot directly modify or delete other user contributions, but are limited to submitting new articles.

In contrast, *contributions to semi-collaborative dictionaries* are carefully examined by professional editors before they are incorporated into the dictionary. One example for this is the *TechDictionary*, which asks for submissions of technology- and computer-related dictionary articles. Naber (2005) found for the semi-collaborative synonym dictionary, *OpenThesaurus*, that only a fraction of the registered users actively contribute to the project. Although user contributions are not limited to additions, he found that most of them merely add new synonyms.

Direct user contributions are also the backbone of the *LEO* project, a collection of eight semi-collaborative bilingual dictionaries. Direct user contributions have been encouraged since the launch of the project in the mid 1990s. Melchior (2013) describes different user contributions to *LEO*, which comprise multiple types of contributions according to our classification system. What we define as contributions to semi-collaborative dictionaries are the submission of new entries, which can be discussed with other users in a forum, as well as the donation of entire word lists and glossaries. After these submissions have been checked by the *LEO* editors for correctness, they are usually directly added to the actual dictionary.

#### **4. Indirect user contributions**

*Indirect user contributions* are suggestions, corrections, supplementary material, comments, external content, and usage data provided by users as feedback to the dictionary makers. The users do not have the possibility to directly modify dictionary articles. We distinguish between *explicit* and *implicit feedback*.

*Explicit feedback* refers to suggestions, wishes, and error corrections explicitly submitted by the users. Thus, users contribute to the dictionary through their feedback on existing content, by providing supplementary material for single articles (e.g., illustrative usage examples and citations), submitting corrections (e.g., spotting erroneous entries, indicating unclear definitions), or commenting on the dictionary as a whole, for example in terms of the presentation of the dictionary articles. Feedback may also include suggesting new content, e.g. in order to fill lemma gaps.

In this context, we can make a further distinction: dictionaries and dictionary portals allowing for *form-based feedback* by providing templates with predefined fields, and those allowing for *free form feedback*, where any text can be submitted, for instance using e-mail or open text fields. There can also be combinations of both types of explicit feedback.

The *LEO* dictionaries provide, for example, separate web forms for reporting errors, such as typos or imprecise translations.<sup>2</sup> The web forms in *LEO* are characterized by providing only a few fields, which are, however, obligatory. Users can also comment on the dictionary as a whole. Melchior (2012) discusses the conflicting opinions of different types of users regarding the content of the dictionary. Some users argue, for example, in favor of adding newly-coined terms even though they might be used only for a very short period of time. This conflicts with other users who complain about confusing and overloaded search results. In addition to that, the users may test beta versions of the dictionary and give feedback by e-mail or forum posts on the overall layout, the presentation of specific data, and new features, such as the presentation of inflection tables (cf. Melchior, 2013).

The *Oxford English Dictionary* provides a very detailed web form with mandatory and optional fields, allowing the users to suggest any improvements at any time.<sup>3</sup> Aside from this kind of feedback, the editors also react to informal messages in the form of letters or e-mails. The *Oxford English Dictionary* particularly fosters initiatives to get in contact with its users, such as the search for *Science fiction citations* recording the first use of an expression. Although participants can submit their citations in an open format e-mail, they are requested to follow strict rules on what kind of information is required.<sup>4</sup> In the project *Wordhunt*, the *Oxford English Dictionary* cooperated with the BBC to collect verifiable evidence of the first use of a word.<sup>5</sup> Thier (2013) gives a detailed overview of these efforts.

These two examples show that there is a smooth transition between direct contributions to semi-collaborative dictionaries and indirect contributions in the form of explicit feedback. While the submission of a new translation to *LEO* (a contribution to a semi-collaborative dictionary) is directly published as part of the dictionary (if the editors agree on it), the citations sent to the *Oxford English Dictionary* (i.e., explicit feedback) represent supplementary material that requires critical verification and selection. The contributions often do not represent a separate dictionary article, but rather a specific piece of evidence that is incorporated into the actual dictionary article. The latter also holds for error corrections that are reported to the dictionary editors.

Rautmann (2013) describes that users of *Duden online* have the possibility to suggest missing lemmas and submit extensions or error corrections by clicking on a button *Wortvorschlag* (i.e., *lemma suggestion*) available at the top of each entry and leading

<sup>2</sup> [http://dict.leo.org/pages/collaboration/ende/reportError\\_en.html](http://dict.leo.org/pages/collaboration/ende/reportError_en.html) (4 June 2013)

<sup>3</sup> <http://www.oup.com/uk/oedsubform/> (4 June 2013)

<sup>4</sup> [http://www.jessesword.com/sf/how\\_to\\_cite](http://www.jessesword.com/sf/how_to_cite) (4 June 2013)

<sup>5</sup> <http://public.oed.com/resources/for-students-and-teachers/balderdash-and-piffle> (4 June 2013)

to a web form.<sup>6</sup> Like the *Oxford English Dictionary*, *Duden online* reacts to e-mails containing propositions and suggestions. The user feedback is considered a valuable resource for the editors to help them improve the dictionary content (Rautmann, 2013).

A different type of explicit feedback is the request for quality assessment. Under the heading “Contribute!”, *dict.cc* asks its users to improve the dictionary by rating a translation as “YES (100% correct)” or “NO / MAYBE”.<sup>7</sup> The task is described in a set of guidelines and designed similarly to the increasingly popular human intelligence tasks on common crowdsourcing platforms, such as Amazon Mechanical Turk,<sup>8</sup> which are frequently used for user studies in marketing, social sciences, or artificial intelligence.

The second type of feedback we define is *implicit feedback*, which is provided by users through their usage of the dictionary. This kind of feedback does not require the users to make any effort, and often they do not even realize that they are contributing.

The way a website is used and accessed by a user is often logged in *webserver log files*. Through the use of visualization tools such as *Google Analytics*, dictionary publishers are able to analyze their users and the way their dictionaries are used. *Duden online* identifies, for example, the most frequently accessed articles and lists them in a sidebar. A publisher can also analyze the search terms supplied by the users and spot lemma gaps in the dictionary. Furthermore, this kind of analysis facilitates the analysis of lookup strategies. It turned out that *Duden online* users often entered multiword expressions, such as “im Folgenden” (“hereafter”) or “des Weiteren” (“in addition”), in the search window. Thus, the editors decided to add frequently-searched multiword expressions as separate lemmas rather than treating them as subentries of one of their constituents (cf. Rautmann, 2013).

Apart from specific tools, the analysis of log files is often suggested as a means of revealing a user’s needs and improving the dictionary (cf. De Schryver & Joffe, 2004). In *Elektronisches Lernerwörterbuch Deutsch–Italienisch* (cf. Abel et al., 2003) the analysis of log files has been characterized by a user model recording the actual use of the dictionary individually for each user (e.g., the number of words looked up per visit, the type of lemma and data categories, etc.). Because of this, users have to register by creating a user account and log in before accessing the dictionary. A similar analysis has been done for the *Base lexicale du français* in order to record not only the words and word combinations used as search terms, but the whole lookup behavior of the users (cf. Verlinde & Binon, 2010).

<sup>6</sup> remark of the authors (4 June 2013): function temporarily disabled

<sup>7</sup> <http://contribute.dict.cc/?action=wizard> (4 June 2013)

<sup>8</sup> <https://www.mturk.com> (4 June 2013)

However, the use of log files has also been criticized as yielding limited, superficial conclusions (cf. Möhrs & Müller-Spitzer, 2008; Verlinde & Binon, 2010). A particular problem is the noise introduced by robots and scripts that automatically browse through the dictionary and thus yield imprecise results. Relevant literature in this field lacks methods for properly cleaning the log files.

Many dictionaries or dictionary portals, such as *Merriam-Webster Online* or *Dictionary.com*, allow their users to sign up for a personal account. Once logged in, a user can, for instance, select their favorite articles or organize the dictionary articles in multiple word lists. Although these features are primarily intended for organizing a user's work, the publisher can utilize this information to learn about frequently-used articles or articles that are organized in the same word list and thus might benefit from being cross-referenced. *Wordnik* publishes those word lists and hence makes them part of the dictionary (McKean, 2011).

Finally, the use of *external user-generated content* is another type of implicit feedback. *Wordnik*, for instance, also includes a great deal of user-generated content from external sources, including images uploaded by users from Flickr and short text messages from Twitter. The users of these external services implicitly contribute with their content to the dictionary. An important consideration when using external user-generated content is the method of dealing with inappropriate content. Lew (2013) discusses, for instance, the use of embarrassing images in the *Google Dictionary*. The vast amount of user-generated content usually impedes checking the contents manually. The dictionaries rather rely on disclaimers, collaborative filtering (cf. Terveen & Hill, 2001), or natural language processing systems.

## 5. Accessory user contributions

*Accessory user contributions* go beyond the dictionary content by initiating an exchange either between the dictionary makers and their users or among the users themselves.

Many dictionary publishers provide blogs reporting interesting or funny facts about language use and the dictionary. The *Macmillan Dictionary Blog*<sup>9</sup> features, for example, the regular series "Language tip of the week", targeted at improving the language proficiency of learners, as well as the "Stories behind Words" series, in which they invite scholars to write about their personal meaning of a certain word. The blog posts usually contain hyperlinks to dictionary articles and thus serve the purpose of promoting the publisher's products and encouraging customers to return.

We consider blogs as a form of *unidirectional communication* for initiating an *exchange between dictionary makers and dictionary users*. Similar measures

<sup>9</sup> <http://www.macmillandictionaryblog.com> (4 June 2013)

include using newsletters, social networks, or microblogging services to distribute news to the dictionary users. Thier (2013), for example, gives an overview of unidirectional communication in the context of the *Oxford English Dictionary*.

A notable type of offer is online language games. Schoonheim et al. (2012) describe, for instance, the “*Het Verloren Woord*” (The Lost Word) game of the *Algemeen Nederlands Woordenboek*. As part of this game, users receive cryptic descriptions of a ‘lost’ word and are asked to exchange ideas and submit their solution. The game attracted a large number of players and the authors mention that it serves an educational and a dictionary-didactic purpose, in addition to mere publicity.

If the users, in turn, contribute to this form of communication by commenting on or rating the posts, they can contribute to defining interesting topics and hence shape the publisher’s offer. We consider this as bidirectional communication, since it results in a mutual exchange between the dictionary makers and users.

The language blog “*Fragen Sie Dr. Bopp!*” (“Ask Dr. Bopp!”) by *canoonet* evokes another type of bidirectional communication: In keeping with the motto ‘there are no stupid questions; each question will be answered’, a user can submit a language-related question and receives an answer by a language expert. Such offers provide useful insight into the information needs of users and help in improving the dictionary. In addition, the answer to a question usually refers to dictionary articles and hence is another way of promoting the dictionary.

Accessory user contributions are not limited to communication between experts and laypeople. The technologies of the Web 2.0 also yield increasing possibilities for initiating an *exchange among the dictionary users* themselves.

A well-known example of this type of accessory contribution is the forum of the *LEO* online dictionaries. Consider the German compound *Nutzerbindung* (customer retention). At the time of writing, there is no English translation encoded in the *LEO* dictionary. However, there is an entry in the forum, in which a user seeks a translation for this term.<sup>10</sup> The user briefly defines the term in German and proposes the literal translation *user binding* (which is obviously wrong). Answers to the forum post propose the phrases “to build a loyal customer base” and “to get repeat business (or customers)”. This example shows that accessory user contributions are an important addition to the dictionary itself, because the users can react to the specific context of a language-related question.

Other means for initiating this kind of discourse include user comments and discussion pages. *Wordnik*, for example, provides a function for commenting on the dictionary articles; this may be used to ask questions or simply to share one’s own

<sup>10</sup> <http://dict.leo.org/forum/viewUnsolvedquery.php?idThread=88976> (7 August 2013)

opinion on a word. Discussion pages are present in *Wiktionary* allowing users to discuss each dictionary article on a separate page. Unlike the commenting function and the forum posts, user contributions to discussion pages are not bound to a linear order. Instead, utterances can be contributed at any position of the discussion page, which makes it possible to discuss multiple issues at the same time.

Accessory user contributions raise a similar issue regarding the inclusion of user-generated content: inappropriate comments are to be removed. In small projects, this can be achieved by checking each contribution manually. Larger projects make use of automatic systems such as spam filters or rely on manual checking in a collaborative effort. *Wordnik*, for instance, displays a link for reporting comments that contain spam.

## 6. Conclusion

Drawing on the relevant literature on user contributions to dictionaries and previous approaches to classifying them, we argue that the existing classifications are insufficient to capture the broad variety of user contributions in a comprehensive way. This is why we propose a new classification distinguishing three main types of user contributions and multiple subdivisions:

- (i) Direct user contributions comprise collaborative efforts in open-collaborative, collaborative-institutional, and semi-collaborative dictionaries. This type of user contribution is targeted towards insertions, modifications, and deletions that directly affect the dictionary articles.
- (ii) Indirect user contributions are subdivided into explicit feedback based on e-mail or web forms and implicit feedback through log file analysis or external user-generated content. Thereby, the users have only indirect means of changing a dictionary article.
- (iii) Accessory user contributions go beyond the dictionary content as they include communication either between the dictionary makers and their users in a unidirectional or bidirectional way or among the users themselves.

We described each type of user contribution with the aid of multiple practical examples relating both to individual dictionaries and to dictionary portals. We have particularly pointed out that a dictionary is not limited to a single type of user contribution. This becomes evident, for example, in the *LEO* dictionaries, which facilitate user contributions of all three main types that we distinguish.

Our proposed classification of user contributions is crucial for properly planning any online dictionary and for future research on user contributions. In this context, quality is a core aspect, which has not yet been exhaustively addressed, in particular

with regard to defining and evaluating quality (cf. Penta, 2011; Nesi, 2012). This is especially a problem if the dictionary function and target audience is not entirely clear, as is often the case with online dictionaries. This is certainly a desideratum for further research.

## 7. Acknowledgements

Christian M. Meyer has been supported by the Volkswagen Foundation as part of the Lichtenberg-Professorship Program under grant No. I/82806.

## 8. References

- Abel, A., Gamper, J., Knapp, J. & Weber, V. (2003). Formative Evaluation of the Web-based Learner's Dictionary ELDIT. In D. Lassner & C. McNaught (eds.) *Proceedings of Ed-Media 2003 World Conference on Educational Multimedia, Hypermedia & Telecommunications, June 23-28, 2003, Honolulu, Hawaii, USA*. Norfolk (USA), pp. 1210–1217.
- Algemeen Nederlands Woordenboek*. Accessed at: <http://anw.inl.nl>
- Bab.la*. Accessed at: <http://bab.la>
- Base lexicale du français (BLF)*. Accessed at: <http://ilt.kuleuven.be/blf>
- canoonet*. Accessed at: <http://www.canoo.net>
- Carr, M. (1997). Internet Dictionaries and Lexicography. *International Journal of Lexicography*, 10(3), pp. 209–230.
- De Schryver, G.-M. & Joffe, D. (2004). On How Electronic Dictionaries are Really Used. In G. Williams & S. Vessier (eds.) *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud*, pp. 187–196.
- De Schryver, G.-M. & Prinsloo, D.J. (2001). Fuzzy SF: Towards the ultimate customised dictionary. *Studies in Lexicography*, 11(1), pp. 97–111.
- De Schryver, G.-M. & Prinsloo, D.J. (2000). Dictionary-Making Process with 'Simultaneous Feedback' from Target Users to the Compilers. In U. Heid, St. Evert, E., Lehmann & Ch. Rohrer (eds.) *Proceedings of the Ninth Euralex International Congress. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart*, pp. 807–818.
- dict.cc*. Accessed at: <http://www.dict.cc>
- Dictionary.com*. Accessed at: <http://www.dictionary.com>
- Duden online (DO)*. Accessed at: <http://www.duden.de>
- Elektronisches Lernerwörterbuch Deutsch–Italienisch (ELDIT)*. Accessed at: <http://www.eurac.edu/eldit>

- Engelberg, St. & Müller-Spitzer, C. (in print). Dictionary Portals. In R.H. Gouws, U. Heid, W. Schweickhard & H.E. Wiegand (eds.) *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Special Focus on Computational Lexicography*. Berlin/New York: de Gruyter.
- Fuertes-Olivera, P.A. (2009). The Function Theory of Lexicography and Electronic Dictionaries: WIKTIONARY as a Prototype of Collective Free Multiple-Language Internet Lexicography. In H. Bergenholtz, S. Nielsen & S. Tarp (eds.) *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Bern: Peter Lang, pp. 99–134.
- Glosbe. Accessed at: <http://glosbe.com>
- Google Dictionary. Accessed at: <http://www.google.com/dictionary> [offline since 2011]
- The Kamusi Project. Accessed at: <http://kamusi.org>
- Köhler Simonsen, H. (2005). User Involvement in Corporate LSP Intranet Lexicography. In H. Gottlieb, J.E. Mogensen & A. Zettersten (eds.) *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2-4, 2002, at the University of Copenhagen*. Tübingen: Niemeyer, pp. 489–510.
- LEO. Accessed at: <http://dict.leo.org>
- Lew, R. (2011). Online dictionaries of English. In P.A. Fuertes-Olivera & H. Bergenholtz (eds.) *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum, pp. 230–250.
- Lew, R. (2013). User-generated content (UGC) in English online dictionaries. In A. Abel & A. Klosa (eds.) *Ihr Beitrag bitte! – Der Nutzerbeitrag im Wörterbuchprozess (OPAL – Online publizierte Arbeiten zur Linguistik)*. Mannheim: Institut für Deutsche Sprache, pp. 9–30.
- Macmillan Open Dictionary. Accessed at: <http://www.macmillandictionary.com/open-dictionary>
- Malone, T.W., Laubacher, R. & Dellarocas, C. (2010). *Harnessing Crowds: Mapping the Genome of Collective Intelligence*, MIT Sloan School Working Paper 4732–09. Accessed at: <http://ssrn.com/abstract=1381502>.
- Mann, M. (2010). Internet-Wörterbücher am Ende der „Nullerjahre“: Der Stand der Dinge. Eine vergleichende Untersuchung beliebter Angebote hinsichtlich formaler Kriterien. In R.H. Gouws, U. Heid, St.J. Schierholz, W. Schweickard & H.E. Wiegand (eds.) *Lexicographica 26*. Berlin/New York: de Gruyter, pp. 19–46.
- Matuschek, M., Meyer, C.M. & Gurevych, I. (2013). Multilingual Knowledge in Aligned Wiktionary and OmegaWiki for Translation Applications, *Translation: Computation, Corpora, Cognition – Special Issue on ‘Language Technology*

*for a Multilingual Europe*, 3(1), pp. 87–118.

McKean, E. (2011). Wordnik: Notes from an online dictionary project. In I. Kosem & K. Kosem (eds.) *Electronic lexicography in the 21st century: New applications for new users (eLex2011)*. Bled, Slovenia.

Melchior, L. (2012). Halbkollaborativität und Inline-Lexikographie. Ansätze und Überlegungen zu Wörterbuchredaktion und Wörterbuchforschung am Beispiel LEO Deutsch-Italienisch. In R.H. Gouws, U. Heid, St.J. Schierholz, W. Schweickard & H.E. Wiegand (eds.) *Lexicographica 28*. Berlin/New York: de Gruyter, pp. 337–372.

Melchior, L. (2013). Ansätze zu einer halbkollaborativen Lexikographie. In A. Abel & A. Klosa (eds.) *Ihr Beitrag bitte! – Der Nutzerbeitrag im Wörterbuchprozess (OPAL – Online publizierte Arbeiten zur Linguistik)*. Mannheim: Institut für Deutsche Sprache, pp. 31–52.

*Merriam-Webster Online*. Accessed at: <http://www.merriam-webster.com>

*Merriam-Webster Open Dictionary*. Accessed at:  
<http://nws.merriam-webster.com/opedictionary>

Meyer, Ch.M. & Gurevych, I. (2012). Wiktionary: a new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography. In: S. Granger & M. Paquot (eds.): *Electronic Lexicography*. Oxford: Oxford University Press, pp. 259–291.

Möhrs, Ch. & Müller-Spitzer, C. (2008). First ideas of user-adapted views of lexicographic data exemplified on OWID and elexiko. In M. Zock & C-R. Huang (eds.) *Coling 2008: Proceedings of the workshop on Cognitive Aspects on the Lexicon (COGALEX 2008)*. Manchester, August 2008, pp. 39-46.

Naber, D. (2005). OpenThesaurus: ein offenes deutsches Wortnetz. In: *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen: Beiträge zur GLDV-Tagung*. Frankfurt: Peter Lang, pp. 422–433.

Nesi, H. (2012). Alternative e-dictionaries: Uncovering dark practices. In S. Granger & M. Paquot (eds.) *Electronic Lexicography*. Oxford: Oxford University Press, pp. 363-378.

*OmegaWiki*. Accessed at: <http://www.omegawiki.org>

*OpenThesaurus*. Accessed at: <http://www.openthesaurus.de>

*Oxford English Dictionary (OED)*. Accessed at: <http://www.oed.com>

Penta, D. J. (2011). *The Wiki-fication of the dictionary: Definitng lexicography in the digital age*. Paper presented at 'Unstable platforms: the promise and peril of transition', 7th Media in Transition Conference, Massachusetts Institute of Technology, Cambridge, MA, 13-15 May, 2011.

*The Rap Dictionary*. Accessed at: <http://www.rapdict.org>

Rautmann, K. (2013). Duden online und seine Nutzer. In A. Abel & A. Klosa (eds.) *Ihr*

- Beitrag bitte! – Der Nutzerbeitrag im Wörterbuchprozess* (OPAL – Online publizierte Arbeiten zur Linguistik). Mannheim: Institut für Deutsche Sprache, pp. 53–66.
- Rundell, M. (2012). 'It works in practice but will it work in theory?' The uneasy relationship between lexicography and matters theoretical. In R.V. Fjeld & J.M. Torjusen (eds.) *Proceedings of the 15th EURALEX International Congress, Oslo*: Department of Linguistics and Scandinavian Studies, University of Oslo, pp. 47–92.
- Schoonheim, T., Tiberius, C., Niestadt, J. & Tempelaars, R. (2012). Dictionary Use and Language Games: Getting to Know the Dictionary as Part of the Game. In R.V. Fjeld & J.M. Torjusen (eds.) *Proceedings of the 15th EURALEX International Congress, Oslo*: Department of Linguistics and Scandinavian Studies, University of Oslo, pp. 974–979.
- Storrer, A. (1998). Hypermedia-Wörterbücher: Perspektiven für eine neue Generation elektronischer Wörterbücher. In H. E. Wiegand (ed.) *Wörterbücher in der Diskussion III*. Tübingen: Niemeyer, pp. 107–135.
- Storrer, A. (2010). Deutsche Internet-Wörterbücher: Ein Überblick. In R. H. Gouws, U. Heid, St. J. Schierholz, W. Schweickard & H. E. Wiegand (eds.) *Lexicographica 27*, Berlin/New York: de Gruyter, pp. 155–164.
- Surowiecki, J. (2005). *The Wisdom of Crowds*, New York: Anchor Books.
- Szenesprachenwiki*. Accessed at: <http://szenesprachenwiki.de>
- TechDictionary*. Accessed at: <http://www.techdictionary.com>
- Terveen, L. & Hill, W. (2001). Beyond Recommender Systems: Helping People Help Each Other. In Carroll, J.M. (ed.) *Human-Computer Interaction in the New Millennium*, Boston: Addison-Wesley.
- Thier, K. (2013). Das Oxford English Dictionary und seine Nutzer. In A. Abel & A. Klosa (eds.) *Ihr Beitrag bitte! – Der Nutzerbeitrag im Wörterbuchprozess* (OPAL – Online publizierte Arbeiten zur Linguistik). Mannheim: Institut für Deutsche Sprache, pp. 67–74.
- Urban Dictionary*. Accessed at: <http://www.urbandictionary.com>
- 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: Fryske Akademy, pp. 1144–1151.
- Wikipedia*. Accessed at: <http://www.wikipedia.org>
- Wiktionary*. Accessed at: <http://www.wiktionary.org>
- Wordnik*. Accessed at: <http://www.wordnik.com>