How Many People Constitute a Crowd and What Do They Do? Quantitative Analyses of Revisions in the English and German Wiktionary Editions

Sascha Wolfer (*wolfer@ids-mannheim.de*)
and
Carolin Müller-Spitzer (*mueller-spitzer@ids-mannheim.de*)
*Institute for the German Language (Institut für Deutsche Sprache),
Mannheim, Germany*

Abstract: Wiktionary is increasingly gaining influence in a wide variety of linguistic fields such as NLP and lexicography, and has great potential to become a serious competitor for publisher-based and academic dictionaries. However, little is known about the "crowd" that is responsible for the content of Wiktionary. In this article, we want to shed some light on selected questions concerning large-scale cooperative work in online dictionaries. To this end, we use quantitative analyses of the complete edit history files of the English and German Wiktionary language editions. Concerning the distribution of revisions over users, we show that — compared to the overall user base — only very few authors are responsible for the vast majority of revisions in the two Wiktionary editions. In the next step, we compare this distribution to the distribution of revisions over all the articles. The articles are subsequently analysed in terms of rigour and diversity, typical revision patterns through time, and novelty (the time since the last revision). We close with an examination of the relationship between corpus frequencies of headwords in articles, the number of article visits, and the number of revisions made to articles.

Keywords: USER-GENERATED CONTENT, ONLINE DICTIONARY, WIKTIONARY, REVISION, EDIT, FREQUENCY, COLLABORATION, WISDOM OF THE CROWD

Zusammenfassung: Wie viele Menschen sind in einer 'Crowd', und was tun sie? Quantitative Analysen der Revisionen im englischen und deutschen Wiktionary. Wiktionary gewinnt immer mehr an Einfluss in vielen linguistischen Bereichen wie bspw. NLP und Lexikographie. Es hat das größte Potential, ein ernsthafter Wettbewerber für die Vertragslexikographie und akademische Lexikographie zu werden. Allerdings wissen wir wenig über die "Crowd", die für den Inhalt von Wiktionary verantwortlich zeichnet. Im vorliegenden Artikel wollen wir einige ausgewählte Fragen bearbeiten, die sich auf groß angelegte Kooperationsarbeit an Online-Wörterbüchern beziehen. Wir verfolgen dabei einen quantitativen Ansatz und verwenden die kompletten Historien des englischen und deutschen Wiktionarys als Datenbasis. Wir zeigen, dass — im Vergleich zur kompletten Autorenbasis des Wiktionarys — nur sehr

wenige Autoren für die überwältigende Mehrheit der Revisionen in beiden Wiktionarys verant-wortlich sind. Im Folgenden vergleichen wir diese Verteilung mit der Verteilung der Revisionen über alle Artikel. Dann werden die Artikel hinsichtlich Gründlichkeit und Diversität, typischen Revisionsmustern in der Zeit sowie der Neuigkeit (Zeit seit der letzten Revision) untersucht. Wir schließen mit einer Analyse des Zusammenhangs zwischen Korpusfrequenz des Stichworts, der Anzahl der Seitenaufrufe des Artikels und der Anzahl der Revisionen des Artikels.

Stichwörter: NUTZERGENERIERTE INHALTE, ONLINEWÖRTERBUCH, WIKTIONARY, REVISION, ÜBERARBEITUNG, FREQUENZ, KOLLABORATION, SCHWARMINTELLIGENZ

1.   Introduction

There is an on-going debate about whether collaboratively constructed dictionaries have the potential to become serious competitors for publisher-based and academic dictionaries (Hanks 2012, Meyer and Gurevych 2012, Rundell 2012). The most promising candidate currently available is Wiktionary — the dictionary project of the Wikimedia foundation. Wikimedia's main project, Wikipedia, has already proven its potential to cover large proportions of user needs in terms of encyclopaedic knowledge; at least if we use page view statistics as an indicator for user satisfaction: On 2016-05-23, the English Wikipedia alone registered over 5,600,000 page views per hour. Studies suggest that the Wikipedia community "takes issues of quality very seriously" (Stvilia et al. 2008)[1]. But Wikipedia's success does not necessarily imply that the same foundation's dictionary project is going to have a comparably major impact on the global dictionary landscape. But Wiktionary is obviously used by many people[2] for a wide array of linguistic needs. And, as we elaborate in the next section, Wiktionary content is also widely used as a scientific resource.

Our main focus is not on investigating the quality of Wiktionary content (see relevant literature in the next section) but more on the processes that shape Wiktionary. It is, in our opinion, essential to get to know the crowd behind Wiktionary a little better in order to gain insights into its composition and processes. This information can help us paint a more detailed picture of "the crowd", which in turn will help us to research Wiktionary and its implications for lexicography as a whole. A good starting point is the revision (or edit) history that determines the state of Wiktionary. Keep in mind, though, that even as you are reading this article, Wiktionary is changing. We can only look at a specific snapshot at a specific point in time. However, even if Wiktionary might look different now than it did at the time of writing this article, we are confident that general principles regarding crowd composition and behaviour that are not subject to sudden change but evolve over much longer periods of time can be deduced. With data files supplied by the Wikimedia foundation (which will be described later in the article), we can consult the complete edit history of all available Wiktionary language editions. And we might identify some general underlying principles regarding *how many people* revise articles *when*

and *how*.

Many of the Wiktionary processes are run automatically using so-called bots. These revisions can also be found in our dataset. We are, however, primarily interested in non-automatic and non-minor revisions. As we will elaborate later, these revisions primarily shape the dictionary and might involve editorial choices.

A note on how we refer to the individual elements of "the crowd": Usually, the term "users of a dictionary" is reserved for the recipients of a dictionary. This terminology contrasts the users and the lexicographers as authors of a dictionary. However, this distinction is not that easy for dictionaries containing user-generated content such as the Wiki-based Wiktionary. Meyer and Gurevych (2012: 271-272) mainly use the term "users" for the authors of Wiktionary, the "Wiktionarians". Lew (2014) no longer considers users of the Web 2.0 (which Wiktionary is as an example of) as "passive recipients of packaged content". Rather, "they actively contribute to the creation and provision of self-made content. This double capacity of newly empowered users can be aptly captured in the neologism *prosumer*, which is a blend of *pro*ducer and *con*sumer." (Lew 2014: 1). Every individual who accesses a page in Wiktionary can choose to contribute to the dictionary at any given time by clicking the "edit" button in the upper right corner of the browser. Therefore, we acknowledge that "web users' social roles become blurred" (ibid.). However, to be as clear as possible, we will use the term "author" for people involved in revision processes and "user" to refer to people who looked something up in Wiktionary.

The remainder of the article is structured as follows: In the next section, we will introduce some related work about the quality of Wiktionary. We will also introduce several scientific applications of Wiktionary content. Section 3 will deal with data preparation and pre-processing of the history files. Also, we will introduce some basic statistics related to revisions, especially the relationship between automatic and minor revisions within our datasets. The main section of the article is Section 4. Initially, we will investigate the number of authors and, more importantly, the distribution of revisions among authors (4.1). In Section 4.2, we will highlight some of the core editing processes in the two language editions. Here, several questions are of interest: How are revisions distributed over entries (and, in the same vein, the relationship between rigour and diversity)? When is the crowd most active? Are there typical chronological revision patterns? How old are the entries (i.e., how long are the phases during which no-one revises articles)? Is there a relationship between revision frequency, number of visits and corpus frequency of the headword? In Section 5, we will provide a summary including some closing remarks.

2.   Related work

In terms of content quality of in Wiktionary, there is disagreement. Meyer and Gurevych (2012), amongst others, refer to the "wisdom of the crowd" (Surowiecki

2005) phenomenon and express their hope that it makes up for the potential "lack of lexicographic experience" (Meyer and Gurevych 2012: 271). They further state that "[c]ollaboratively constructed lexicons are continually updated by their community" which "yields a steeply increasing coverage of words and word senses. […] An important characteristic of collaborative lexicography is that the large number of authors has the ability to express the actual use of language […]." (Meyer and Gurevych 2012: 259). Hanks (2012) is not that optimistic. While acknowledging positive aspects like "imaginative use […] of multimedia hypertext" (p. 81) which he sees as a "model for the electronic dictionary of the future" (p. 82), he also states that "[i]n the English Wiktionary, the etymologies are taken from or based on those in older dictionaries; as are definitions, which are extremely old-fashioned and derivative, taking no account of recent research in either cognitive linguistics or corpus linguistics" (p. 78). As stated above, we will not compare the quality of Wiktionary entries to professionally edited dictionaries in this article. Instead, we want to gain a more detailed insight into the revision processes in two Wiktionary editions: The English language edition, which is the largest in terms of number of pages/ articles (at the time of writing this article, the English Wiktionary had over 4,250,000 entries), and the German language edition, which is rather small in comparison (the German Wiktionary had approx. 430,000 entries at the time this article was written, which is roughly a tenth of the English Wiktionary, and ranks 14 among all language editions)[3]. As previously mentioned, it is essential to shed some light on the crowd that is behind Wiktionary to discuss the position of Wiktionary in the lexicographic landscape. Another reason why we want to get to know the crowd is the wide-spread application of Wiktionary as a data source for a range of scientific applications, including works in the field of natural language processing (NLP) like sense definitions (Henrich, Hinrichs and Vodolazova 2011), semantic relatedness (Zesch et al. 2008), synonymy networks (Navarro et al. 2009), pronunciation extraction (Schlippe et al. 2010), idiom identification (Muzny and Zettlemoyer 2013) and many more. Other areas of application are sentiment analysis (Chesley et al. 2006) and analyses of vocabulary difficulty (Medero and Ostendorf 2009). In addition, some parts of Wiktionary entries are integrated into other lexicographic resources, for example in bilingual dictionaries (Lindemann 2014). As can be seen from this (doubtlessly incomplete) list, Wiktionary has become an increasingly important resource, which contrasts with the little knowledge we actually have of the processes that create(d) it. We hope that this article can contribute to a better understanding of these processes.

Some questions we want to answer using Wiktionary revision histories have been similarly dealt with by other researchers, but mainly with respect to the Wikimedia foundation's largest project, Wikipedia (Greenstein and Zhu 2012, Poderi 2009, Stein and Hess 2008, Wilkinson and Huberman 2007). Since Wiktionary uses the same platform, we can transfer some of the ideas that have been previously applied to Wikipedia. In Section 4.2, we refer to an idea by Lih (2004)

who sees Wikipedia as "the largest example of participatory journalism to date" and compares a set of benchmark articles to Wikipedia articles that have been cited in the press. Although this question is not directly relevant for the topic of this article, Lih introduces the notions of *diversity* and *rigour*[4]. We will define and use these concepts in Section 4.2 to compare entries within one Wiktionary language edition and also to compare the two language editions to one another.

In Section 5, we will summarise our findings and discuss some implications for the lexicographic landscape.

3.   Data preparation

3.1   Downloading and pre-processing Wikimedia history files

Revision data were extracted from the edit history data dumps available from the Wikimedia Foundation[5] on 2015-08-07. The files are available in XML format and have to be parsed with a SAX parser due to their size. We used an R (R Core Team 2015) script and the XML package (Lang 2013) to implement the parser. The complete edit history files were converted to CSV files with the following information associated with each revision: (1) the title of the revised page, (2) the revision timestamp, (3) the name of the author who carried out the revision, (4) whether the revision was flagged as minor, and (5) the comment the author added to the revision. It is not trivial to decide whether a revision was made automatically (i.e., by a bot) or by a human. After consultation with contributors to the German Wiktionary, we used two lists of user names provided by the English[6] and German[7] Wiktionary language edition. These lists contain all users that are flagged as bots, which is the standard procedure to identify users as bots in Wikimedia products. The consultation with the contributors showed that this is — if not perfect — the most reliable way to identify automatic revisions. In the next step, we also tagged all revisions as automatic that were associated with a comment containing the strings "autoedit" or "clean up". Results concerning the distribution of automatic and non-automatic revisions are summarized, amongst others, in Table 1.

In this article, we only look into the revision history of the English and German Wiktionary language editions. Since it takes time and computational effort to parse the history files, we compiled CSV files for eight language editions of Wiktionary and provide them to the scientific community to conduct analyses for other language editions and/or expand the presented set of analyses. CSV revision history files for English, Malagasy, French, Russian, Polish, German, Chinese, and Spanish are available under http://dx.doi.org/ 10.7910/DVN/TYLQBN. The languages are ordered according to the number of revisions. All files contain the revision history as of August 2015. If readers are interested in creating CSV files for another language edition and/or newer history files, the respective R scripts are also available under http://dx.doi.org/

10.7910/DVN/TYLQBN. We would greatly appreciate if any output generated by these scripts were also to be made available to the scientific public.

3.2   Minor and automatic revisions

Whenever authors revise a page in Wiktionary, they can flag their revision as minor. This is a signal to others "that only superficial differences exist between the current and previous versions [of the page]. Examples include typographical corrections, formatting and presentational changes, and rearrangements of text without modification of its content. A minor edit is one that the editor believes requires no review and could never be the subject of a dispute" (https://en.wikipedia.org/wiki/Help:Minor_edit, last access on 2015-11-13). Although it is not perfectly safe to rely on authors checking the respective box if they only made minor adjustments to the page, we use this information in our analyses, simply because it is the most reliable source of information available. The help page cited above is quite clear about when a revision should be flagged as minor, and we believe that, since both the English and German Wiktionary have existed for more than ten years, authors are most likely acquainted with the use of the check box now.

Table 1: Date of first revision of a non-meta page, number of total revisions, revisions of meta pages as well as automatic and minor revisions for the English and German Wiktionary edit history

	English Wiktionary	German Wiktionary
Date of first revision of non-meta page	2001-09-27 ("zen")	2002-04-24 ("Metamorphose")
Total revisions	32,767,004	4,362,860
Revisions of meta pages	2,343,380 (7.15%)	398,930 (9.14%)
Automatic revisions	21,693,943 (66.2%)	2,831,858 (64.9%)
Minor revisions	23,736,184 (72.4%)	2,991,574 (68.6%)

Table 1 shows some key figures of minor, automatic and meta page revisions in both Wiktionary editions. Meta pages are all pages that are not "normal" entries. We identified meta pages by a colon in the page name. Note that there is not only one type of meta page. There may be talk pages associated with each article, user pages for each user, and user talk pages that are associated with each user page. Also, there are several meta pages that are not associated with other pages, such as help and question pages, several special pages and so on. Actually, the term "meta page" is not used for these kinds of pages in the Wiktionary. In the context of this article, meta page can be translated as "not an entry page".

Meta pages are frequently revised during the early stages of a new Wiki-media project so as to set up all relevant pages. This is reflected by the devel-

opment of shares of meta page revisions. In the English Wiktionary, for example, 29.5% of the first 10,000 revisions were meta page revisions. Compared to the overall value of 7.15% (see third row in Table 1), this share is quite high. The first non-meta revision of the English Wiktionary edition was approximately half a year earlier than in the German Wiktionary.

In the English Wiktionary, 7.5 times more revisions were made than in the German Wiktionary. Later in this article, we will also deal with the number of authors who contribute to the two Wiktionary editions. There, we will also see that many more authors work on the English Wiktionary than on the German Wiktionary.

An important piece of information missing from Table 1 is the cross-combination of automatic/human and minor/non-minor revisions. This information is available in Table 2. In both language editions, minor non-automatic revisions constitute around two thirds of all revisions. Automating tedious repetitive processes with bots obviously only makes sense if a lot of revisions can be performed with the bot program. So, this high share of automatic revisions is no surprise. The revisions we will be primarily interested in in this article are the ones which fall into the lower right cells of the contingency tables: non-minor, human revisions. In both language editions, these revisions account for around a fifth of all revisions. The large values for χ^2 (both above 1 million and hence not included here) and φ suggest that the relationships between the automation of revisions and the fact that they are flagged as minor in Table 2 are strong and did not occur by chance.

Table 2: Frequencies, percentages, and marginal sums for automatic/human vs. minor/non-minor revisions of non-meta pages in the English and German Wiktionary editions. Association measure for English: $\varphi = .671$, for German $\varphi = .658$

English Wiktionary				German Wiktionary			
	minor	non-minor	total		minor	non-minor	total
automatic	20,034,791 65.9%	1,187,568 3.90%	21,222,359 69.8%	automatic	2,518,446 63.5%	240,180 6.06%	2,758,626 69.6%
human	2,900,133 9.53%	6,301,132 20.7%	9,201,265 30.2%	human	324,032 8.17%	881,272 22.2%	1,205,304 30.4%
total	22,934,924 75.4%	7,488,700 24.6%	30,423,624 100%	total	2,842,478 71.7%	1,121,452 28.3%	3,963,930 100%

4.   Results and discussion

4.1   How many people constitute a crowd?

The OED editorial staff comprises around 125 members[8]. This is without a doubt a very high number of lexicographic experts working on a group of dic-

tionaries and presumably the biggest editorial dictionary worldwide. Just for comparison: the staff at Duden, the most prominent publisher's dictionary for German, consists of eight permanent lexicographers[9]. It is safe to consider these people lexicographic experts as well. Nonetheless, we have to ask ourselves whether it might also be a good idea to use a (freely available) dictionary that can be potentially edited by *all* speakers of a language with access to the Internet? We should consider another question before making a judgement on this matter: How many people actually constitute the crowd behind Wiktionary? Meyer and Gurevych (2012: 272) already refer to the Zipf-like nature of the distribution of revisions over authors. This means that the vast majority of authors make very few revisions and that a small group of authors (compared to the overall number of registered authors) are responsible for the majority of revisions (just like the most frequent word types in a language accumulate the vast majority of word tokens). This statement is certainly correct as we will see in a moment. However, we want to clarify what this means in terms of numbers. In this section, we will only consider non-automatic, non-minor revisions of non-meta pages, because these are the types of revisions that really shape the dictionary and add the main content to entries. For the English Wiktionary, this applies to 6,301,132 revisions. The German Wiktionary history file contains 881,272 such revisions.

Both Wiktionary language editions allow visitors to revise articles without registration. The revision is then associated with the current IP address of the respective visitor. In the revision history of the English Wiktionary, 750,055 (11.9%[10]) revisions were made by unregistered authors. Unregistered authors made 144,002 revisions (16.3%) in the German Wiktionary. One might object that these are too many revisions to exclude from the analyses. However, there are certain problems associated with the analysis of IP-based data. The most crucial problem is that, for IP addresses, we cannot be sure if one person always edits with only one IP and, on the other hand, whether one IP always identifies one user[11].

We consider data from 36,958 registered authors for the English Wiktionary and 6,111 authors for the German Wiktionary. An author needed at least one non-minor revision to a non-meta page to enter the analysis. As can be seen in Table 3, the total number of non-minor revisions is distributed very unevenly among the authors. In the English Wiktionary, 33 authors are responsible for over half of all non-minor revisions to non-meta pages. In the German Wiktionary, only 14 authors made the majority of revisions. This obviously means that the tails of these distributions are very long, i.e. there are many authors with only few revisions. In the English Wiktionary, almost half of all registered authors (44.3%) only made one revision. It is similar in the German Wiktionary: There, 42.3% of all registered authors only made one revision.

Table 4 is a transformation of Table 3, i.e. the number of revisions is used as the starting point. It includes shares for at least 2, 5, 10, 100, and 1000 revisions respectively. In the English Wiktionary, an author needs to have made

1,346 revisions to be included in the top 1% of authors in terms of revisions[12]. In the German Wiktionary, this threshold lies at 1,571 revisions. Given that the maximum number of revisions an author made is 237,600 for the English Wiktionary (user "Equinox"[13]) and 57,432 for the German Wiktionary ("Dr. Karl-Heinz Best"), this is a rather low value and shows how extreme this distribution behaves when it approaches its extremes.

Table 3: Number of top ranked authors in terms of non-minor revisions to non-meta pages and the associated shares of all those revisions. Included ranks differ after 100 for the different language editions to accommodate for the different numbers of authors. Also included are the minimum numbers of authors needed to reach over 50 and 95 percent of all revisions

Number of top ranked authors, English Wiktionary	Share of all revisions in the English Wiktionary	Number of top ranked authors, German Wikt.	Share of all revisions in the German Wiktionary
10	26.8%	10	42.8%
33	50.2%	14	50.2%
50	58.3%	50	80.6%
100	71.1%	100	89.2%
1000	94.9%	262	95.0%
1026	95.0%	500	97.1%
5000	98.6%	1000	98.3%
10000	99.2%	2000	99.1%

Table 4: Share of authors with at least n revisions in the English and German Wiktionary

Number of revisions n	Share of authors in the English Wiktionary	Share of authors in the German Wiktionary
at least 2	55.7%	58.0%
at least 5	25.5%	28.4%
at least 10	15.8%	17.4%
at least 100	4.27%	4.63%
at least 1000	1.20%	1.37%

The figures we presented in this section are informative if we want to get an impression of the revision distribution over users. We can, however, take another perspective by conducting a thought experiment. We compare the number of users to the sizes of the editorial staff of two dictionary publishers in the respective countries. The OED is one of the most well-known publishers for the English language and the Duden for German respectively. What would it mean for the revisions in Wiktionary if we transfer the respective staff sizes to

the number of Wiktionary authors?

We can compare the OED's staff size by checking the number of revisions of the English Wiktionary for the top 125 authors. Or to put it differently: We check, what percentage of all revisions the top 125 authors made. The value is 74.8%. This could be considered an impressive figure. However, also note that we would lose 1,400,313 human, non-minor revisions in the English Wiktionary if we were to exclude all authors not in the top 125 authors. At Duden, eight permanent employees work on entries. The top 8 authors contributing to the German Wiktionary made 38.1% of all revisions. If we would exclude all revisions from rank 8 downwards, we would lose 456,725 revisions.

4.2   Revision processes

Distribution of revisions over entries

As we saw in the previous section, revisions are very unevenly distributed among authors. From the article perspective, we can check how revisions are distributed among the entries in the two Wiktionary language editions accordingly. We are primarily interested in pages related to entries, so we excluded all meta pages from this analysis. Also, we only consider human, non-minor revisions to keep analyses from the article perspective as comparable as possible to the analyses from the author perspective. This leads to a data set with 1,983,023 article pages from the English Wiktionary and 228,869 article pages from the German Wiktionary.

Table 5: Number of top ranked articles in terms of non-minor revisions to non-meta pages and the associated shares of all revisions. Included ranks differ after 100 for the different language editions to accommodate for the different numbers of articles. Also included is the minimum numbers of articles needed to reach over 50 percent of all revisions

Top ranked articles in the English Wiktionary	Share of all revisions in the English Wiktionary	Top ranked articles in the German Wiktionary	Share of all revisions in the German Wiktionary
10	0.166%	10	0.310%
50	0.532%	50	0.990%
100	0.891%	100	1.60%
1,000	4.60%	1,000	7.61%
10,000	17.0%	10,000	30.8%
100,000	42.5%	26,203	50.0%
163,072	50.0%	100,000	82.6%
500,000	71.0%	200,000	96.7%
1,000,000	84.4%		
1,500,000	92.3%		

Table 5 shows the respective distribution. It is obviously less extreme than the distribution over authors. Let us simply compare the share of revisions the 100 most edited articles received: In both Wiktionary editions, the top 100 edited articles received less than 2% of all revisions. For authors (see Table 3), this is a very different picture: The top 100 authors in the English Wiktionary are responsible for over 70% of all revisions, in the German Wiktionary, the respective value is almost 90%.

One might argue that the distributions over authors and articles cannot be compared in such a way because there are far more articles than authors in both Wiktionaries. A way to work around this is presented in Figure 1. On the *x*-axis, we included more and more percentages of articles or authors. On the *y*-axis, we calculated the share of revisions these authors made or these articles received for each step. The grey line shows us how the graph would look if revisions were perfectly evenly distributed over articles and authors (e.g. that 50% percent of authors made exactly 50% of all revisions). We can see that all distributions deviate clearly from the grey line and that there are slight differences between the language editions.
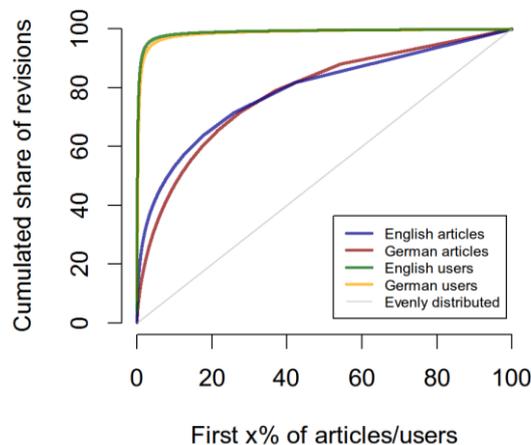

Figure 1: Top x% of articles/authors in the English and German Wiktionary editions and their associated cumulated share of revisions. The straight grey line symbolises a hypothetical distribution where each article and author receives the same share of revisions

The clearest difference is between the distributions for articles and authors, though. We need to include a much larger share of articles to reach a specific amount of revisions. For example, to register 80% of all revisions, we only need to include 0.474% of all authors (which corresponds to 175 authors) but we need to include 39.3% of all articles (which corresponds to 780,185 articles). In summary, we have to acknowledge that revisions are distributed very un-

evenly over authors **and** articles, but the distribution among authors is more extreme than for articles.

From a lexicographic point of view, this might be interpreted as a good sign. Although only few authors are responsible for the vast majority of revisions in both Wiktionaries, this does not imply that only very few articles receive preferential treatment. These few authors distribute their work in the dictionary over a wide range of articles. On the other hand, this may also support the impression that not many articles are very sophisticated. We will elaborate on this question in the following section by introducing the terms rigour and diversity.

Rigour and diversity

As briefly mentioned in the introduction, Lih (2004) used the terms rigour and diversity in a study of Wikipedia articles. Rigour is the variable we analysed in the previous subsection: the total number of edits for an article. "The assumption is that more editing cycles on an article provides [sic] for a deeper treatment of the subject or more scrutiny of the content" (Lih 2004: 8). We already saw that some articles are edited more rigorously (in this sense) than others. The idea behind the concept of diversity is that "[w]ith more editors, there are more voices and different points of view for a given subject" (ibid.). Given this operationalisation, the more unique authors contributed to an article, the more diverse it should be. Following Lih (2004), both concepts can be interpreted as the higher, the better. Although the concepts were originally applied to Wikipedia articles, they can just as easily be applied to our Wiktionary datasets. Again, we are only taking non-minor revisions to non-meta pages from registered authors into consideration. Given the definitions of diversity and rigour, the value for diversity may never exceed the value for rigour. If 100 different authors make 100 revisions to an article, diversity and rigour are equal (100). As soon as at least two of the 100 revisions are made by the same author, the value for diversity decreases while the value for rigour remains constant. The more interesting question then is by how much the data points for the different articles deviate from the baseline, which is defined as rigour and diversity being equal. In Figure 2, all articles (non-meta pages) from both Wiktionaries are plotted. The two language editions are colour-coded. The location of an article is defined by its diversity (x-axis) and rigour (y-axis).

Several conclusions can be drawn from the graph: (1) Articles in the English Wiktionary are more diversified, both in terms of diversity and rigour. This might simply be due to the fact that there is much more potential for an article to get higher values of diversity and rigour in the English Wiktionary, because there are 7.16 times more revisions and 6.05 times more authors in the English than in the German Wiktionary. However, there are also 8.66 times more articles in the English than in the German Wiktionary. So one might argue that there is also much more "ground to cover" by revisions and authors in the Eng-

lish Wiktionary. (2) The bi-variate distribution of articles in terms of diversity and rigour clearly differs from the baseline. This suggests that more than one author normally contributes to an article. (3) The "upward bend" of the overall pattern suggests that the relationship between diversity and rigour is not strictly linear. The more the rigour of an article increases, the less pronounced the increase in diversity. The comparison of two regression models predicting rigour by diversity suggests that the quadratic trend explains additional variance compared to only the linear trend[14]. This means that the relationship "more rigour means more diversity" is weaker for articles with very high rigour values.
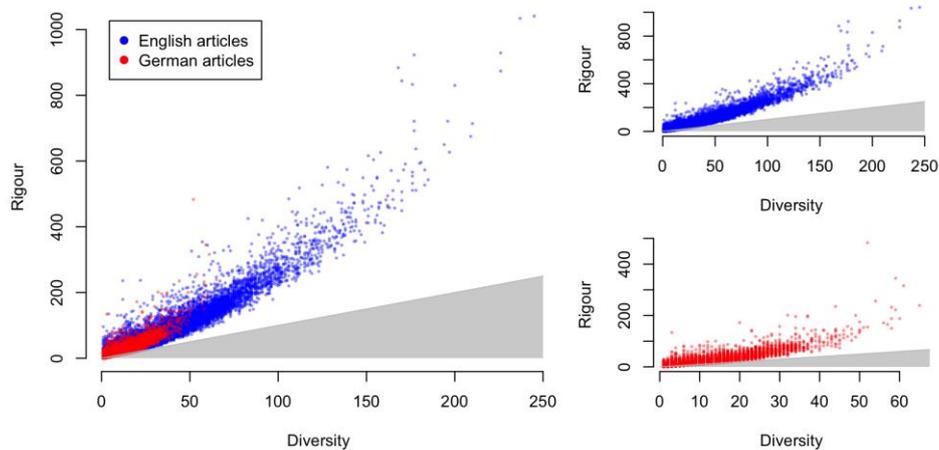

Figure 2: Diversity (number of unique registered and contributing authors) and rigour (number of revisions) of the articles in the English and German Wiktionary language editions (both plotted on top of each other in the left panel and separate plots for both languages on the right). The grey area symbolizes the area of the graphs where rigour is lower than diversity. No data point may fall in this area. The data point for "water" (English Wiktionary) was excluded from the graph for better readability. Its coordinates are (2284, 231)

Although Figure 2 spans a broad range of diversity and rigour values, we also have to point out that the majority of articles in both Wiktionaries score rather low on these scales. Visually, this is represented by the very dense data point cluster in the lower left region of the graph. The mean diversity of articles is 2.09 (English) and 2.59 (German). Mean rigour ratings are 3.18 (English) and 3.85 (German). Due to the skewness of the distributions, the respective medians[15] are slightly lower. Which conclusions can we draw for the revision processes in Wiktionary given these results? As might not be apparent in Figure 2, the majority of articles in both Wiktionaries were edited by a single author

only[16] (English: 67.3%, German: 53.0% of all articles). Would we apply a four (or more) eyes principle to ensure the quality of Wiktionary articles, these articles might not be sufficient. However, to evaluate this question further, we could also take discussion pages into consideration. It might well be the case that potential problems in articles are expressed on the accompanying discussion page and corrected by the one author who edited the associated article.

Typical revision patterns over time per article

We are dealing with over 13 years of history for two online dictionaries. So, it would be a shame to not take time into consideration as well. Two potentially interesting questions are: Is there a "typical" revision pattern over time for articles? And how long are typical "idle periods" of articles, meaning the time spans in which no-one works on the article (as measured by non-minor, non-automatic revisions)? The first question is not easy to answer because it is not clear how to operationalise the critical concepts "typical" and "revision pattern". We propose a visualisation to gain a first insight into revision histories of individual articles. On the x-axis, we plot the time since article creation (technically, article creation is the first revision). On the y-axis, we plot the cumulated share of revisions of this article on the specific day. The path one line takes through time can be called the "revision trajectory" of an article. Whenever long horizontal lines appear in the revision trajectory, no-one worked on the article for a long time. Vertical lines within such a revision trajectory indicate a period when the article was revised repeatedly in rapid succession. To keep the plot visually manageable, we only include the top 1000 revised articles from the English Wiktionary.

Figure 3 shows these revision trajectories. There is an accumulation of trajectories on the middle diagonal of the graph. This represents a steady revision process with revisions coming in quite regularly. The randomly chosen article "chicken", symbolised by the green line in Figure 3, is an example for this. There are no clear phases when "chicken" receives many revisions (which would lead to steeper lines) or when it receives very few revisions (which would lead to a more horizontal trajectory). The trajectories for "water" and "cool" are quite similar: Both revision histories start off slowly. Then, both articles receive many revisions in rapid succession. The "revision sprint" sets in later for "water" than for "cool"[17]. After that, the slope of the trajectory of "water" is slightly higher than for "cool" and there is also a minor revision sprint at around 4000 days since article creation. A few other, more extreme trajectories are also visible in the upper left region of the graph. These are articles that received a lot of revisions early in their history and are then only revised very infrequently — supposedly because the community thinks they are finished or no-one takes care of them anymore. So, we think the quantitative view of this data shows no intriguing patterns. It might leave room for further research to

examine groups of articles that share a similar revision pattern more qualitatively in order to find out whether these groups of headwords share particular (linguistic) properties.
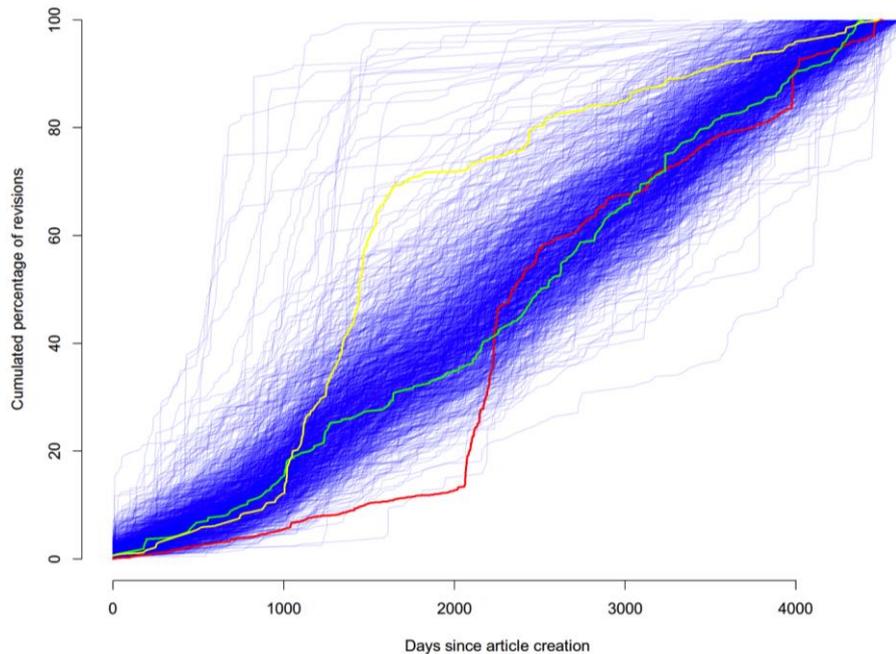

Figure 3: Cumulated percentage of revisions in relation to the days since article creation for the 1000 most revised articles from the English Wiktionary. Each line ("revision trajectory") represents one article. Three revision trajectories are highlighted. The red line symbolises "water", the yellow line "cool" and the green line "chicken". The latter two were chosen randomly. "Water" is the article that received the most revisions in the English Wiktionary

How old are the entries?

The advantage online dictionaries have over printed dictionaries is the flexibility to adapt an entry within a very short period of time because the authors do not have to wait for the next edition of the printed book. This advantage should be especially relevant for Wiktionary as well as all other Wikimedia products. After all, "wiki" is Hawaiian for "fast". So, we would expect that articles are regularly revised and we do not expect a large number of articles that remain unrevised for a very long time. A fast and easy way to assess the typical idle

period of articles is to take the current state of articles and calculate how long ago the last revision was concluded. We can then calculate the means or medians of these time spans for both Wiktionaries. The mean (median) time span during which an article is not revised in the German Wiktionary is 257 (143) days. For the English Wiktionary this time span is longer: 441 (386) days.

Another way to approach this is to start from a given number of days x (e.g. 183 days, half a year). We then count backwards starting from 2015-08-07, the day the history dump was created, until we arrive at the day x before 2015-08-07 (in our example, this would be 2015-02-05). We then extract the number of articles n_x that have not been revised within this time span. Again, we only consider human, non-minor revisions. If we divide n_x by the total number of articles, we arrive at the percentage of all articles that have not been revised in the last x days, starting from 2015-08-07. In the example, 1,443,141 articles have not been revised in the time between 2015-02-05 and 2015-08-07 (n_{183} = 1,443,141). These are 72.8% of all articles. In the German Wiktionary, n_{183} is only 106,189. However, there are also fewer articles in the German Wiktionary. Nevertheless, these are only 46.4% of all articles in the German Wiktionary. If we now vary x from 1 to 4134 days (which is the longest time during which an article was not revised) and calculate each n_x and the respective share of articles, we can plot x and the share in one graph. Figure 4 is the result of this process. The further we go back in time, the fewer articles were not revised within this time span. There are certain differences between the two language editions, though.

The values decrease faster for the German Wiktionary and also get to the "bottom line" much faster than the values for the English Wiktionary. On the one hand, this is not that surprising because there are fewer articles in the German Wiktionary. So, it should be easier to cover a large share of these articles with revisions. However, one should also consider the number of authors in this argumentation. Given that the community shaping the English Wiktionary is considerably larger than that of the German Wiktionary, we would have expected patterns here that are more similar. The rather steep decreases, especially in the German Wiktionary are another interesting observation found in Figure 4. For example, the line drops sharply between 500 and 600 days, i.e. many articles were revised in a comparably short time. We do not have a final explanation for this. It could be due to the German Wiktionary's official 10th "birthday" on 2014-05-01[18] that motivated authors to revise a lot of articles. The pattern visualised in Figure 4 is also reflected by the values in Table 6. There, we change perspective from non-revision to revision: How many articles have been revised in the last x years? In this way, we see that the majority of articles from the German Wiktionary were revised in the last 6 months while for the English Wiktionary, only roughly a quarter of the articles were revised in the same time span.
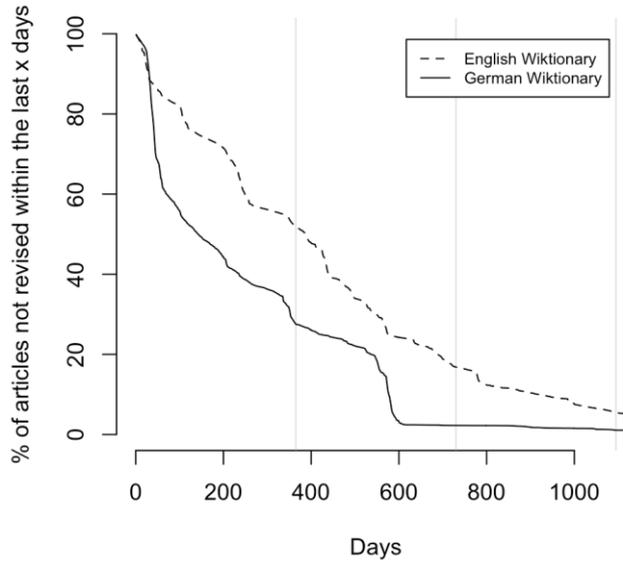
Figure 4: Share of articles that have not been revised in the last x days in relation to the time span x, starting from 2015-08-07 counting backwards in time. Vertical grey lines indicate one, two, and three years

Table 6: Share of articles that were revised within the last x years in the English and German Wiktionary language editions

Time span x	Share of articles in the English Wiktionary	Share of articles in the German Wiktionary
½ year	27.2%	53.5%
1 year	47.8%	72.4%
2 years	83.1%	97.7%
3 years	94.5%	98.9%

We can also look at a few of the articles that have not been revised in a very long time. In the German Wiktionary, for example, there are 1,758 articles that have not been revised in the last 5 years. These contain many articles for non-German words but also entries for German words. Interestingly, most of these articles are permanent redirections (e.g., "mit Mann und Maus untergehen" → "Mann und Maus"; "m.E." → "m. E."; "labio-dental" → "labiodental"; "jenes" → "jener") or very short articles only indicating alternative ("gewerbsmässig", "einigermassen") or old spellings ("deplaciert", "deplaziert"). All these articles have a "reference article" they are linked to (the redirection target, another spelling or the correct spelling) that can be updated. So, it makes sense that there are indeed a few articles that have not been revised in a very long time.

The relationship between revisions, number of visits and corpus frequency

As we have previously shown (Müller-Spitzer et al. 2015), articles with high frequency headwords are visited more frequently in the German Wiktionary. In this article, we introduce another variable to this relationship: the number of revisions an article receives during the history of Wiktionary. It would make sense for the authors of Wiktionary to revise articles that are relevant to a broad public. Or put differently, the question we want to answer is whether articles that are revised more frequently are also visited more frequently. To answer this question, we enrich the article-based data set with the number of visits based on the Wikimedia log files available from https://dumps.wikimedia. org/other/pagecounts-raw/ (last accessed on 2015-12-02). We use the number of visits during 2014 for this analysis. As corpus frequency measures for the German Wiktionary, we use word form frequencies based on the German Reference Corpus/Deutsches Referenzkorpus (DeReKo, Kupietz et al., 2010). For the English Wiktionary, we used a frequency list based on the Google Books 2012 unigram data for both American and British English[19]. We found 147,205 (64.3%) of all German Wiktionary headwords in the DeReKo word form frequency list. Keep in mind that there are also articles for non-German words in the German Wiktionary. These words are most likely not included in the frequency list and are thus excluded from this analysis. The same holds for the English Wiktionary. There, we found 594,075 (24.7%) headwords in the frequency list.

All three variables of interest (number of revisions, number of visits, corpus frequency of the headword) are correlated[20], so we have to find a way to "disentangle" the relationships between them. One way to achieve this is to divide the data into a given number of equal sized portions. Previously, we divided the data into ten equally sized portions in terms of frequency (Müller-Spitzer et al. 2015: 16). In statistical terminology, we are looking at frequency deciles. We can then look at each frequency decile and concentrate on the other two variables we are interested in: number of visits and number of revisions. Frequency deciles are aligned on the *x*-axes in Figure 5 (English Wiktionary) and Figure 6 (German Wiktionary). On the *y*-axes, the mean number of visits for the articles in this frequency decile and the respective revision category (colour of bars) is recorded. The revision category of an article is defined by the percentile rank the respective article takes in relation to all other articles. The revision category "top 5%" contains all articles that are among the top 5% in terms of revisions (127,316 articles in the English Wiktionary; 11,509 articles in the German Wiktionary). Accordingly, the revisions category "top 10%" contains all articles that are among the top 10% articles in terms of revisions, excluding those already associated with the "top 5%" category — so this category should more precisely be named "top 5-10% of articles" (169,089 articles in the English Wiktionary; 11,656 articles in the German Wiktionary). All other articles fall into the category "bottom 90%" (2,106,436 articles in the English Wiktionary and 205,704 articles in the German Wiktionary).
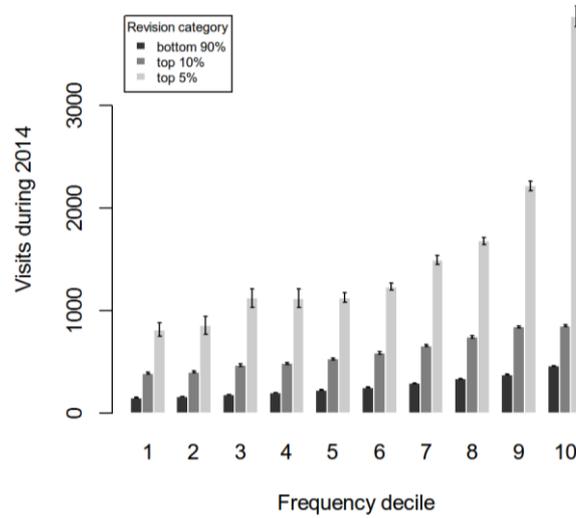
Figure 5: Visits to the English Wiktionary in 2014 (y-axis) in relation to corpus frequency decile of the headword (x-axis) and the revision category (colour of bar)
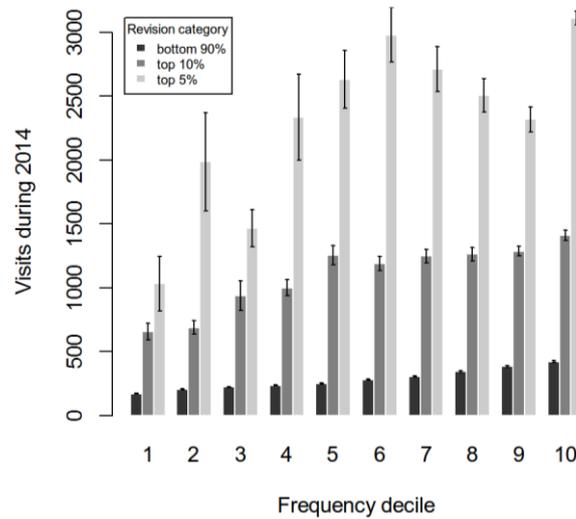

Figure 6: Visits to the German Wiktionary in 2014 (y-axis) in relation to corpus frequency decile of the headword (x-axis) and the revision category (colour of bar)

Both Figure 5 and Figure 6 exhibit a similar pattern. In all frequency deciles, articles that fall into the top 5% of articles in terms of revisions are also visited more often. This difference does not only hold true in comparison to the bottom 90%, but also in comparison to the 5% of articles below the top 5% (category name "top 10%"). The effect of corpus frequency on look-up frequency is visible for the English Wiktionary. Especially the top 5% of revised articles show a clear tendency to be visited much more often in the frequency deciles 6 to 10. This effect is not that clear in the German Wiktionary. While there is still an overall effect of corpus frequency on look-up frequency, this effect is not consistent for the top 5% of revised articles. However, the effect we are primarily interested in holds true: Articles that are revised more frequently are also looked up more frequently. One could object that an article page has to be accessed to be edited. So, each time someone revises an article, she or he also has to visit this article. It could be tested whether this fact is responsible for the observed pattern by subtracting the number of revisions in 2014 from the number of visits in 2014 and repeat the above analysis. We do not report the respective figures because they look almost the same as Figure 5 and Figure 6. The only visible change is in the absolute number of visits (y-axis). The overall pattern, however, remains stable.

5.   Summary

Traditionally, dictionaries are compiled by publishers or by academic projects financed by the public. Such dictionaries are written by lexicographic experts. Wiktionary, as a collaborative, non-profit dictionary is based on voluntary work. The idea behind such collaborative joint activities is that the professionality of some is substituted by the collective intelligence of many. In this paper, we attempted to evaluate the question whether there are really many people working on Wiktionary and whether it is therefore the right wording to speak of a 'wisdom of crowds' phenomenon.

We saw that the distribution of lexicographic effort, as operationalised by the number of revisions, is heavily biased for both authors and articles. Few authors do the majority of the work and the majority of revisions are distributed among few articles, with the latter distribution being less extreme. These distributions are also reflected by the bivariate distribution of diversity and rigour, where we saw that a few articles rank high on both scales while the vast majority of articles, however, score rather low on both scales. In terms of quality, this is problematic. In an ideal world, most articles would have high values for rigour and diversity. We would recommend that the Wiktionary community focuses their efforts on low-ranking articles in terms of rigour and diversity. Maybe, articles with very high numbers of visits would be a good starting point.

Although both author participation and article revisions are so unevenly distributed, most of the articles in Wiktionary are not very old in terms of the

most current revision. This might be interpreted as a good sign, but it also means that there does not seem to be an effective organisation of revisiting older articles to check whether they might need updating. Rather, we can observe certain points in time during which "revision sprints" seem to take place — both on the level of the whole dictionary, but also for individual articles.

Finally, we saw that frequency in language-use (as measured by corpus frequency of the headword), consultation frequency and revision frequency are heavily interdependent. The most frequent words in language are visited and revised more often than words that play a more marginal role in every-day language-use. This might be seen as a form of consumer-orientation (work on those words that people are looking up). However, this practice (which is most likely not a conscious one) is also associated with the risk of losing track of less frequent words.

On the basis of this data, there are, in our view, two possible answers to the question whether many people are working on Wiktionary:

No, there are not many people: Anyone involved in dictionary projects knows that 30 people working irregularly and part-time are not enough to write a good and comprehensive general dictionary. If a dictionary team were that small, either most of the sense disambiguation information, paraphrases, collocations and typical phrases would be copied from other lexicographic resources and only a small part of the vocabulary would be described from scratch, or only a small part of the vocabulary would be elaborated in an innovative way and, correspondingly, the dictionary would only contain this part of the vocabulary. As Wiktionary contains as many headwords as the big general dictionaries, the common strategy seems to be to integrate copyright-free data from older dictionaries and to complement them whenever necessary. From that point of view, Wiktionary is in no way comparable to Wikipedia. On the basis of the revision data, we can surely say that Wiktionary will not surpass previous lexicographic works on a wide basis in terms of comprehension of vocabulary, state-of-the-art semantic description or innovative forms of presentation. This would require a bigger crowd working on a dictionary. Accordingly, Wiktionary will not replace publishers' and academic dictionaries in terms of content and quality.

Yes, there are many people: When Wiktionary was launched there was considerable doubt that the project would attract any kind of community. The line of reasoning was that it clearly may be a pleasure to share own knowledge about a field of research, a kind of sport or a historical person with others and to use the opportunity to spread this knowledge via Wikipedia. However, who wants to be an expert, for example, on the *word* 'marmalade' or 'tree' and spend time elaborating dictionary entries? At most neologisms or special parts of the vocabulary seemed to be a promising field to attract voluntary work. These expectations were not fulfilled. The revision data clearly shows that 'normal' highly frequent words are being revised, not only neologisms etc. And there

are at least 30-50 people who regularly work on the English resp. the German Wiktionary.

What does that mean for the relation between Wiktionary and academic or publishers' lexicography? In our view, the best relation would be not that of competitors but of two actors in the field that benefit from cross-pollination of ideas and to see Wiktionary — from the point of academic and publishers' lexicography — as one more possibility to disseminate lexicographic data. However, we are very aware of the fact that this idea ignores the problem of financing lexicographic enterprises. If the user base of Wiktionary grows partially through the integration of established lexicographic work (see Rundell 2015, for elaboration on this issue) that was financed by others, it is hard to convince an enterprise or a public institution to keep playing the game. However, one fact is clear from our analyses of the revision data: We cannot expect Wiktionary to become a better dictionary on a wide basis than established dictionaries. In consequence, if no professional work is conducted on dictionaries, Wiktionary will be no long-term compensation. On the other hand, it is a pleasure to see that there is a language-interested community that works on dictionaries voluntarily. Is this not also a sign for the relevance of dictionaries?

Finally, we would like to stress that Wiktionary is a resource of major interest for usage research because all the lexicographic content, the user/author statistics and revision data is freely available at all times. It is very hard to receive such comprehensive data from any other (especially commercially oriented) dictionary projects.

6.   Endnotes

1.   However, there are also studies suggesting that "the number of active contributors in Wikipedia has been declining steadily for years" (Halfaker et al. 2013: 664). The authors of the study argue that "several changes the Wikipedia community made to manage quality and consistency in the face of massive growth in participation have ironically crippled the very growth they were designed to manage" (ibid.). Note that this refers to the contributors of Wikipedia and not to the recipients.

2.   The English language edition of Wiktionary was visited approx. 980 million times in 2014 and 600 million times in 2015. The German language edition was visited approx. 156 million times in 2014 and 97 million times in 2015. Please note that a new page count definition came into effect after April 2015. What might look like a decrease from 2014 to 2015 is an effect of this new definition. Extensive statistics on all Wikimedia projects are available from https://stats.wikimedia.org/wiktionary/EN/Sitemap.htm [last accessed on 2016-02-04].

3.   Note that the German Wiktionary is rank 6 in terms of number of revisions (4,710,263 revisions compared to 40,266,646 revisions in the English Wiktionary). This information is available from https://meta.wikimedia.org/wiki/Wiktionary/Table [last accessed on 2015-11-12].

4.   In British English, rigor and rigour have different meanings. In BE, "rigour" has the meaning we want to imply, rigor has a medical meaning. In contrast to Lih (2004), we are using "rigour" to avoid confusion.

5.   The most actual data dumps are available under https://dumps.wikimedia.org/backup-index.html [last accessed at 2016-06-08].

6.   The list of users with the bot flag for the English Wiktionary is available under https://en.wiktionary.org/wiki/Special:ListUsers/bot [last accessed at 2015-11-16].

7.   The list of users with the bot flag for the German Wiktionary is available under https://de.wiktionary.org/w/index.php?title=Spezial%3ABenutzer&username=&group=bot [last accessed at 2015-11-16].

8.   Information taken from http://public.oed.com/the-oed-today/staff-of-the-oxford-english-dictionary/ [last accessed on 2015-11-11].

9.   Personal e-mail communication with a Duden editorial staff employee.

10. Percentages apply to non-automatic, non-minor revisions of non-meta pages.

11. Concerning this argument, there might be objections that we cannot be sure about registered authors, either. For example, the user account "Wamito" in the German Wiktionary is used by (at least) two persons as indicated on the associated user page. However, we are confident that this is an exception. We are not aware of any way to assess this automatically for all user accounts.

12. The top 1% of authors in the English Wiktionary consists of 369 authors and 61 authors in the German Wiktionary.

13. As an anonymous reviewer points out, this "sounds as [sic] an incredibly high number of revisions for one single user". However, the author's page in Wiktionary (https://en.wiktionary.org/wiki/User:Equinox, last accessed on 2016-05-23) does not suggest that she or he makes heavy use of automated revision processes.

14. The model comparison tests whether the residual sum of squares decreases significantly if the quadratic term is included into the predictor structure. Both F and X^2 tests detect highly significant improvements of model fits for both the English and German Wiktionary if the quadratic trend is included.

15. The median is the value which divides a distribution into two equally sized halves (i.e. there is an equal number of data points below (or equal) and above the median). Compared to the mean, the median value is more robust in presence of extreme values ("outliers").

16. Note that we excluded non-registered authors from the analysis which might be considered "unfair" in this context. Consequently, our measurements represent the lower bound of diversity and rigour values.

17. Note that "later" has to be interpreted relative to the creation date of the respective article. In Figure 3, $x = 0$ could be a different absolute point in time for each of the thousand articles we plotted.

18. There are revisions in the German Wiktionary history file that are over two years older than the "official" birthday on 2004-05-01. We decided to keep these earlier revisions because they also contribute to the current state of the German Wiktionary.

19. Google Books frequency data is available from http://storage.googleapis.com/books/ngrams/books/datasetsv2.html [last accessed on 2015-12-02].

20. Spearman correlations for the English Wiktionary are: corpus frequency vs. revisions: $\rho = .381$; corpus frequency vs. visits: $\rho = .477$; visits vs. revisions: $\rho = .273$. Spearman correlations for the German Wiktionary are: corpus frequency vs. revisions: $\rho = .313$; corpus frequency vs. visits: $\rho = .331$; visits vs. revisions: $\rho = .616$.

7.   References

Chesley, P., B. Vincent, L. Xu and R.K. Srihari. 2006. Using Verbs and Adjectives to Automatically Classify Blog Sentiment. Nicolov, N., F. Salvetti, M. Liberman and J.H. Martin (Eds.). 2006. *Computational Approaches to Analyzing Weblogs: Papers from the 2006 Spring Symposium*: 27-29. (Technical Report SS-06-03.)Menlo Park, CA: AAAI Press.

Greenstein, S. and F. Zhu. 2012. *Collective Intelligence and Neutral Point of View: The Case of Wikipedia*. Working Paper, National Bureau of Economic Research. http://www.nber.org/papers/w18167.

Halfaker, A., R.S. Geiger, J.T. Morgan and J. Riedl. 2013. The Rise and Decline of an Open Collaboration System. How Wikipedia's Reaction to Popularity is Causing its Decline. *American Behavioral Scientist* 57(5): 664-688.

Hanks, P. 2012. Corpus Evidence and Electronic Lexicography. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 57-82. Oxford: Oxford University Press.

Henrich, V., E. Hinrichs and Vodolazova, T. 2011. Semi-Automatic Extension of GermaNet with Sense Definitons from Wiktionary. *Proceedings of the 5th Language and Technology Conference (LTC 2011), Poznań, Poland, November 25–27, 2011*: 126-130.

Lang, D.T. 2013. XML: Tools for Parsing and Generating XML within R and S-Plus. http://CRAN.R-project.org/package=XML.

Lew, R. 2014. User-Generated Content (UGC) in English Online Dictionaries. Abel, A. and A. Klosa (Eds.). 2014. *Der Nutzerbeitrag im Wörterbuchprozess. 3. Arbeitsbericht des wissenschaftlichen Netzwerks "Internetlexikografie"*: 9-30. Mannheim: Institut für Deutsche Sprache. (OPAL — Online publizierte Arbeiten zur Linguistik 4/2014.

Lih, A. 2004. Wikipedia as Participatory Journalism: Reliable Sources? Metrics for Evaluating Collaborative Media as a News Resource. *Proceedings of the 5th International Symposium on Online Journalism, April 16–17, 2004, University of Texas at Austin, USA*. https://online.journalism.utexas.edu/papers.php?year=2004.

Lindemann, D. 2014. Creating a German–Basque Electronic Dictionary for German Learners. *Lexikos* 24: 331-349.

Medero, J. and M. Ostendorf. 2009. Analysis of Vocabulary Difficulty Using Wiktionary. *Proceedings of the ISCA International Workshop on Speech and Language Technology in Education (SLaTE–2009)*. Warwickshire, England.

Meyer, C.M. and I. Gurevych. 2012. Wiktionary: A New Rival for Expert-built Lexicons? Exploring the Possibilities of Collaborative Lexicography. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 259-291. Oxford University Press.

Müller-Spitzer, C., S. Wolfer and A. Koplenig. 2015. Observing Online Dictionary Users: Studies Using Wiktionary Log Files. *International Journal of Lexicography* 28(1): 1-26. doi:10.1093/ijl/ecu029.

Muzny, G. and L. Zettlemoyer. 2013. Automatic Idiom Identification in Wiktionary. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP), 18–21 October 2013 Seattle, Washington, USA*: 1417-1421. Stroudsburg, USA: Association for Computational Linguistics.

Navarro, E., F. Sajous, B. Gaume, L. Prévot, H. ShuKai, K. Tzu-Yi, P. Magistry and H. Chu-Ren. 2009. Wiktionary and NLP: Improving Synonymy Networks. *ACL Workshop on The People's*

Web Meets NLP: Collaboratively Constructed Semantic Resources, August 2009, Singapore: 19-27. Stroudsburg, USA: Association for Computational Linguistics.

Poderi, G. 2009. Comparing Featured Article Groups and Revision Patterns Correlations in Wikipedia. *First Monday* 14(5). http://firstmonday.org/ojs/index.php/fm/article/view/2365/2182.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rundell, M. 2012. It Works in Practice but Will It Work in Theory? The Uneasy Relationship between Lexicography and Matters Theoretical. Fjeld, R.V. and J.M. Torjusen (Eds.). 2012. *Proceedings of the 15th EURALEX International Congress*, 7–11 August 2012, Oslo: 47-92. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo. http://www.euralex.org/elx_proceedings/Euralex2012/pp47-92%20Rundell.pdf.

Rundell, M. 2015. From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. *Lexikos* 25: 301-322.

Schlippe, T., S. Ochs and T. Schultz. 2010. Wiktionary as a Source for Automatic Pronunciation Extraction. Kobayashi, T., K. Hirose, S. Nakamura. 2010. *11th Annual Conference of the International Speech Communication Association (Interspeech 2010)*, *Makuhari, Chiba, Japan, September 26–30, 2010*: 2290-2293. International Speech Communication Association (ISCA). http://dblp2.uni-trier.de/db/conf/interspeech/interspeech2010.html

Stein, K. and C. Hess. 2008. Viele Autoren, gute Autoren? Eine Untersuchung ausgezeichneter Artikel in der deutschen Wikipedia. Alpar, P. and S. Blaschke (Eds.). 2008. *Web 2.0: Eine empirische Bestandsaufnahme*: 107-129. Wiesbaden: Vieweg+Teubner.

Stvilia, B., M.B. Twidale, L.C. Smith and L. Gasser. 2008. Information Quality Work Organization in Wikipedia. *Journal of the American Society for Information Science and Technology* 59(6): 983-1001.

Surowiecki, J. 2005. *The Wisdom of Crowds*. New York: Anchor Books.

Wilkinson, D.M. and B.A. Huberman. 2007. Cooperation and Quality in Wikipedia. *WikiSym '07: Proceedings of the 2007 International Symposium on Wikis, Montreal, Canada, October 21–23, 2007*: 157-164. New York: ACM Press.

Zesch, T., C. Müller and I. Gurevych. 2008. Using Wiktionary for Computing Semantic Relatedness. Fox, D. and C.P. Gomes (Eds.). 2008. *Proceedings of the 23rd AAAI Conference on Artificial Intelligence, Chicago, Illinois, July 13–17, 2008*: 861-867. Menlo Park, California: AAAI Press.