

# Overwriting knowledge: analyzing the dynamics of Wikipedia articles

Nathalie Mederake

Deutsches Wörterbuch, Göttingen Academy of Science and Humanities,  
Papendiek 14, 37073 Göttingen, Germany  
E-mail: nmedera@gwdg.de

## Abstract

The popularity of open collaborative content generation such as Wikipedia, while expanding the amount of available information, also poses particular challenges as its user-generated content changes constantly. This paper proposes to study the development of Wikipedia entries and to systematically measure and evaluate this type of user-generated dynamics. The applied approach is able to identify phases of the constant process of content generation. It takes into account the interrelations between dynamics of user contributions and article-related real-world events. A data set spanning article versions and associated discussion pages over two years was analysed. This allowed identifying trigger pulses that drive the articles' development both on qualitative and quantitative levels. For effective planning of online dictionaries that stress the involvement of users or intend to add collaborative components, it is crucial to consider such findings. The approach might also be transferrable to lexicography in terms of analysing the revisions of a collaborative dictionary entry as a signal indicative of lexical change. For that reason, I conclude with a discussion of the results and their relevance for expert lexicographic products.

**Keywords:** wiki; collaborative lexicography; content generation process

## 1. Introduction

With the rise of the Web 2.0, users can actively participate in the compilation of online reference works such as dictionaries and encyclopaedias. However, these works can be subdivided into different partial areas of lexicography (each with its own characteristic forms), as they are displayed by Wiegand et al. (2010: 125). Lexicographic products can investigate the respective language or their subjects “when the perspective of the comments is such that one can obtain answers about corresponding non-language objects” (ibid.). According to the distinction made by Wiegand et al., the largest available and fastest growing collaboratively constructed encyclopaedia project Wikipedia is to be defined as a non-scientific lexicographical reference work, predominantly fulfilling the mentioned purposes related to subjects.

Compared to editorial reference works, the collaborative lexicographic process shows significant differences in the steps and phases towards compilation. One of the peculiarities of a collaborative lexicographic process is the iterative writing process that yields multiple revisions of an entry (cf. Meyer, 2013: 53). These revisions can lead to continuous changes in the lexicographic product, for example, when a new article constituent is introduced. Hence, collaborative projects are revision-driven and

not directed to a final closing phase as might be the case with editorial reference works. Users write and edit articles in a collaborative manner and the outcome is published immediately on the web; also, feedback can be instantly given. One might consider it either a problem or actually a benefit that web contents are subject to constant change and that dictionaries or encyclopaedias thus will not remain the 'final products' they used to be for a long time. Of course, the traditional dictionaries or encyclopaedias are also not entirely "final" - there is a discrete number of successive editions representing the major development over longer time periods. In contrast, the fact that wiki entries are updated in a continuous manner, as often as needed or regarded useful, in principal by anyone who wishes to make a change, has made them an integral part of everyday life.

It is not surprising that methods of how to systematically measure or evaluate user-generated contents within the wiki-environment are developing. They are concerned e.g. with the evolution of discussion (Kaltenbrunner & Laniado, 2012), the understanding of the writing process (Kallass, 2015), and the investigation of look-up frequencies (Müller-Spitzer et al., 2015). The research of Stvilia et al. (2005a, b; 2008) and Stvilia & Gasser (2008) discusses the aspects and dynamics of information quality in Wikipedia and gives useful pointers on how the quality assessment and improvement process operates. Their model is concerned with changes in the field of information quality and can actually be used for reasoning about similar dynamics in different settings. In their study, they used the discussion page or talk page and other process-oriented pages within Wikipedia to determine indicators for information quality. Despite these advances, web dynamics continue to be an ongoing challenge for lexicographers (and linguists in general). In addition, lay users are still mostly unaware of the developments that happen in the background of collaborative projects such as wikis and of how contents are changed in the course of a revision.

In fact, since every user benefits from up-to-date content and is given the opportunity to reflect on how content has developed in the page history, it is important to set the starting point there: What changes have been made, which links have been replaced or which illustrations have been chosen at what time? In addition, less compressed forms of presentation, as available in the wiki-interface, result in longer, sometimes less structured articles<sup>1</sup>. But what is important or relevant for both the users and the producers in this reference work; what do they deal with, especially in a more narrative structure? I believe that answering these questions will also lead to fruitful findings for institutional or professional lexicography. The research of Müller-Spitzer et al. (2015) for example uses quantitative evaluations of log files to explore general patterns of look-up behaviour in Wikipedia's sibling, German Wiktionary, to

---

<sup>1</sup> As a side note: The absence of space restrictions in the digital environment altogether will, in the long run, lead to longer dictionary articles, or narrative article structures on word-related information in institutional lexicography as well, like in the examples of so-called Wortgruppenartikel (= entries referring to word group) in *lexiko* or Macmillan's *BuzzWord*.

understand the needs of users and the information they would like to have. Accordingly, I believe that we can only use search results derived from wikis for our own lexicographic products if we fully understand how the collaborative system works and what is important for the active user. I will therefore present a method of how to systematically study the development of Wikipedia entries. The analysis takes into consideration findings from the history page related to the respective article as well as the discussion pages, together with corresponding real-world events. Besides, some light will be shed on the following questions: what kind of information seems to be important for user-generated content in an online encyclopaedia and what are the underlying strategies of revision? I will conclude with findings on regularities in the dynamics induced by the collaborative environment and a discussion of the results within the field of lexicography.

## 2. Model and distinctive Features

The concept of Wikipedia has been popular for a long time, as has collaborative online editing in general. These processes are being widely used even by information professionals (Lih, 2004; Emigh & Herring, 2005) – and they have also found their way into daily language lexicographic routine. In fact, there seems to be a fruitful coexistence between Wikipedia and more traditional language dictionaries: institutional dictionary projects such as *Algemeen Nederlands Woordenboek* also offer links to Wikipedia in their search results<sup>2</sup>. Similarly, institutional language dictionaries are used as references in Wikipedia’s articles<sup>3</sup>. Taking the sister project Wiktionary into account, it becomes apparent that the German Wiktionary, for example, relies to a large extent on secondary sources such as Duden online, *Digitales Wörterbuch der deutschen Sprachen* or *Deutsches Wörterbuch von Jacob and Wilhelm Grimm* (cf. Meyer, 2013: 42). However, the variety within primary, secondary and tertiary sources, such as monographs, grammar etc. (cf. Wiegand, 2010: 133), tends to differ according to the specifications of each reference work and also depends on whether it is going to serve language or subject related lexicographic purposes. Likewise, it is argued that open-collaborative contributions (that by definition draw upon very diverse sources) have enormous potential in keeping the contents of a dictionary up to date and ensuring their high quality (cf. Abel & Meyer, 2013: 179), even if most of them are not constituted or controlled by a predefined group of experts. In fact, Wikipedia actually “gets better the more people use it, since more people can contribute more knowledge, or can correct details in existing knowledge for which they are experts” (Vossen & Hagemann, 2007: 47).

---

<sup>2</sup> E.g. <http://anw.inl.nl/article/peer>

<sup>3</sup> Compare references to Oxford English Dictionary and *Griechisches Etymologisches Wörterbuch* in the German Wikipedia article ‘Birne’:  
<https://de.wikipedia.org/wiki/Birnen#Quellen> (6/7/2015)

Therefore, a general model for the better understanding of the collaborative process will be presented. It refers to the Wikipedia system in particular and highlights its distinctive features. Following Bruns' (2008: 102) description of a wiki, “[w]ikis enable their users to create a network of knowledge that is structured ad hoc through multiple interlinkages between individual pieces of information in the knowledge base; they represent, in short, a rapidly changing microcosm of the structures of the wider Web beyond their own technological boundaries”. Based on this, circular movements in the contribution process and complex interactions of endogenous and exogenous factors can be specified (cf. Fig. 1). Such factors correspond to activity peaks that have been observed so far not only in Wikipedia (e.g. Kaltenbrunner & Laniado, 2012; Mayer, 2013: 123–143) but also in other social media platforms such as Youtube (Crane & Sornette, 2008) or Twitter (Lehmann et al., 2012).

One of the endogenous factors for a collaborative encyclopaedia is for example the software platform of Wikipedia, which is built upon a relational database with different search paths. The linking structure also allows for immediate cross-references – even to articles that do not yet exist. Additionally, wiki-based reference systems are usually neither based on fixed (lexicographic) instructions nor do they show a predefined microstructure. One of the main characteristics of wiki software is an extreme reduction of the costs of collaborative content creation, dissemination and upkeep. The structural openness obviously causes inconsistencies in the layout of the articles and their microstructure. But most importantly, users can and do directly modify contributions of other users. The process of production and using is ongoing and is never finished. In fact, the most important result of collaborative editing is a continuous process rather than a static product. This process can generate projects that are richer and more complex than those produced by individuals, which leads us to the most important exogenous factor: A wiki is nothing without its users.

Wikipedia still grows and develops its features, despite the known discrepancy in active and passive user behaviour, e.g. in German Wikipedia (cf. Busemann, 2013: 319). For example it has been shown (cf. Döring, 2010: 177) that passive usage (via page visits etc.) prompts further active participation. Additionally, search engine optimization has had a significant effect on the visibility (and in that, recognition) of web content. In this environment the concept of ‘prosumption’ (i.e. in the most general sense, the creation of products and services by the same people who will ultimately use them) seems to work better than an elaborate and refined product created by experts (such as expert lexicographers). The idea behind the prosumer commodity and thus that of user-generated content (Lew, 2014), and bottom-up-lexicography (Carr, 1997) is that the roles of producers and consumers blur and merge. It is also argued that criteria such as openness, sharing, peering and global outreach increase the value of prosumer participation. Facing the collaborative extension and editing of Wikipedia, Bruns coined the term ‘produsage’ to describe user-led content production within the Web 2.0 environment. He argues that “within the communities which engage in the collaborative creation and extension of information and knowledge [...] the role of

‘consumer’ and even that of ‘end user’ have long disappeared, and the distinctions between producers and users of content have faded into comparative insignificance” (Bruns, 2008: 2).

Therefore, boundaries become transient. The concept of article ownership does not apply as anyone can modify articles at any time. The collaborative process is intermittent and not systematic due to significant interactions. They are fostered on an object level, where article creation (and thus representation of knowledge) takes place, as well as on a meta-level, where the above mentioned concept of ‘produsage’ as well as events and developments over time affect every article. Such interactions also determine the dynamic character of content creation. Because of the ongoing “work in progress” situation the quality of every article can also only be expected to be fluid and transient. Here, the term ‘dynamic’ points to the fact that the articles’ contents and appearances change over time. But is there a pattern?

In their studies about dynamics in information quality, Stvilia et al. (2005b; 2008) and Stvilia & Gasser (2008) agree on the definition about information quality as being the assessment on information’s ‘fitness for use’ (cf. Juran, 1992; Wang & Strong, 1996) in a particular task system or activity system. Regarding information quality in Wikipedia, they observed a number of patterns in the development trajectories for featured articles that appeared to follow the life cycle of the underlying entities. However, besides the articles’ underlying entities or the context of its evaluation (e.g. degree of domain knowledge) and use (also in terms of sociotechnical structure) there is a significant link to the element I described as ‘produser’. In terms of quantification, this means: the number of edits an article may receive is affected by the attention drawn to the article’s entities. Ferron & Massa (2011a, b), Keegan et al. (2011) and Kallass (2015) have identified this kind of intensive participation in revisions and discussions on talk pages as event-related. Additionally, the analysis of Stvilia & Gasser (2008) showed that Wikipedia “would direct community resources to a particular article in anticipation of an event that could change the quality and/or criticality of the article” (ibid.).

This means that the triggering of an article’s development is caused by real-world changes related to its topic as well as by initiatives of the produser-element. Thus, “fitness for use” seems to resemble a negotiation process which is highly context sensitive: Coherence needs to be achieved in terms of the articles’ entities<sup>4</sup> and the potential contribution of the produser-element to this topic – in short, coherence between the interactions of endogenous and exogenous factors.

---

<sup>4</sup> Here, context sensitivity also relates to Wikipedia policies. E.g. in English Wikipedia the avoidance of *recentism*, that is editing an article without a long-term, historical view, and determining proper weight in depth of detail, quantity of text, prominence of placement, etc., belong to the content policies of Wikipedia:  
[https://en.wikipedia.org/wiki/Category:Wikipedia\\_content\\_selection](https://en.wikipedia.org/wiki/Category:Wikipedia_content_selection)

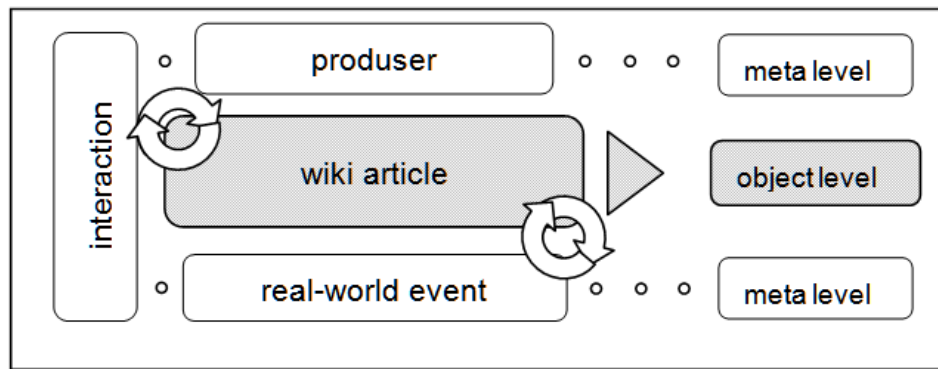


Figure 1: Activity model within a wiki (cf. Mederake, 2014: 239)

### 3. Data set and methodological approach

Wikis include mechanisms that allow us to follow visible changes made to pages over time, i.e. the display of the related data history<sup>5</sup>, as well as discussion pages or talk pages, which are tied to entries and where various content-related issues can be addressed. As these features are central to the Wikipedia quest in terms of information quality, I will make use of them to see what information is distributed and when.

Wikipedia articles describe or deal with different kinds of entities: people, places, events, concepts, or things. The data set for this study comprised the edit histories of two articles from the German Wikipedia: ‘Zitronenpresse’ (= lemon squeezer)<sup>6</sup> and ‘Eurokrise’ (= European debt crisis)<sup>7</sup>. Describing 1) a very general object and 2) a current event, these articles are typical examples of article topics in the German Wikipedia; the article ‘lemon squeezer’ also was awarded the label ‘worth reading’ until it was highlighted as ‘excellent’ during the survey and can therefore be qualified as a high-quality article.<sup>8</sup> Categories like ‘worth reading’ or ‘excellent’ denote article

<sup>5</sup> In the edit history, meta-data elements can be found containing the following information: data and time, name or IP of the user, comment to clarify the edit purpose. Edit histories are also a source for meta-information about the article (age, time of update, number of times the article has been edited, information about editors and edit type). Such elements of the data history can provide valuable information about the social structure and dynamics of the articles’ content creation. <http://de.wikipedia.org/wiki/Hilfe:Versionen>

<sup>6</sup> <http://de.wikipedia.org/wiki/Zitronenpresse>

<sup>7</sup> <http://de.wikipedia.org/wiki/Eurokrise>

<sup>8</sup> Articles are awarded featured article status after the community has achieved a consensus that the article meets the featured articles criteria (comparable to English Wikipedia; i.e. attributes as well-written, comprehensive, well-researched, neutral, stable, appropriate structure, consistent citation format and so forth). It can be judged that these are general quality dimensions based on respective cultural and social conventions, and characteristics specific to the encyclopedia article genre and the community of Wikipedia. Articles keep their featured article status, even if they get changed again, until they are demoted for lack of meeting the quality requirements. <http://de.wikipedia.org/wiki/Wikipedia:Bewertungen>

status in the German Wikipedia comparable to the ‘featured article’ status, which articles in the English Wikipedia can achieve (after a thorough review process). It should be noted that the objective of the featured article process is to encourage the writing process to evolve and improve, thus increasing quality within Wikipedia.

Over a period of more than two years, a data set of 20 article versions altogether was created, using monthly, bi-monthly or quarterly data points. The first version of every article topic marks the starting point of the survey. Additionally, I looked into the logs of the associated discussion pages or talk pages to allow a more in-depth content analysis of specific incidents within the articles’ development.

In order to observe which instances had been moved or added at what time during the articles’ development, findings in frame semantics (following Konerding, 1993) were applied in a coding procedure to develop a classification scheme. This scheme was then applied to all versions of an article. Coding was performed by using QDA software. Frame semantics<sup>9</sup> came into play in order to assess the current state of knowledge displayed in the articles’ content and to evaluate what was considered noteworthy at what time in the article. The additional analysis of real-world events (being located on the meta-level, see above) then helped to identify some of the trends and patterns in the articles’ development. Besides qualitative assessment, the focus had been set on data for statistical and quantitative analysis, which was recorded manually for additional results.

Konerding (1993) used findings in frame theory for a study with a lexicographic-lexicological approach. In his approach he redesigned frame theory to “a theory for knowledge representation/realization” (= Theorie der Wissensdarstellung/vergegenwärtigung; translated from Konerding, 1993: 92) and exemplified how linguistic frame analysis can be applied to a variety of purposes by employing frames empirically. In doing so, he developed a method to systematically characterize relevant slots of a frame by using a set of questions. He also invented a procedure called ‘hyperonym type reduction’ including a restricted set of highest hyperonyms to determine potential reference points or slots of any linguistic expression by retracing every one of them to such a highest-level hyperonym (cf. Ziem, 2014: 267). This procedure is used to identify the slots in a frame and is important for the implementation of frames as analytical instruments. As a result, only a relatively small set of German nouns occur as end elements in the reduction chain. In consequence, it is basically the slots in the frame that any lexeme (noun) evokes which correspond to the slots in the frame of a noun specified in Konerding’s approach. Nevertheless, the expression of these lexemes can be retraced via the procedure of hyperonym type reduction.

---

<sup>9</sup> I understand frames as conceptual knowledge units that linguistic expressions evoke. They group slots and fillers as structural constituents to define a stereotypical object.

Use of Koneading's approach has been popular in German language research in order to document how concepts (of knowledge) have developed and which aspects of slots are focused in different types of discourse (cf. Ziem, 2014: 16). Therefore, it has been exemplified in several studies how his proposal of linguistic frame analysis can be applied to a variety of purposes by employing frames empirically (ibid.). Due to the wide range of possible applications of frames, they serve as a tool kit in my study to analyse content development in Wikipedia entries. Lexical items, in this case the headword, provide access to a considerable amount of subject knowledge in the corresponding article and display how they have developed over time. For means of my analysis the hyperonyms "artefact" (for 'lemon squeezer') and "event" (for 'European debt crisis') were identified as well as the additional reference points in the frame system of each hyperonym according to Koneading (1993: 309–340). In combination with a systematic question-answer-advance (e.g. for an object-related article, "What are features and characteristics of a lemon squeezer?" "How did this artefact originate?"), an encoding paradigm was defined to study the development of an article with respect to its content. Here, I specifically focused on the use of hyperlinks and their immediate text environment as potential fillers or information units within the systematic frame approach. Hyperlinks do not only act as navigation tools in the network of knowledge unfolded by the articles' editors but also as salient features within a narrative article as they draw the user's attention to specific areas of the text. Additionally, and for the benefit of a more granular analysis of the articles' development, topics from the discussion pages and ongoing real-world events were taken into account.

#### **4. Analysis and discussion**

As stated above, frame analysis in combination with a systematic question-answer-approach was used for means of encoding the wiki data. The methodology allowed dissecting and reasoning about the articles' development in the German Wikipedia both conceptually and systematically.

Recurrently analysing the article versions by a code system provided a perspective on the content's diachronic development. Additionally, the hierarchical structure of every article version was taken into account. Connections with endogenous activities of Wikipedia and real-world events, or so-called exogenous activities, could then be traced in the articles' development. These events were called trigger pulses (see above). On a quantitative scale, developments in the article structure became visible whenever a trigger pulse had been identified on the meta-level. Due to the code structure, it was possible to retrace the movement of the information unit around the hyperlinks within the articles' structure. The number of dots in each cell denotes the quantity of fillers or information units per movement type.



date / movement	3/05	6/05	10/0 5	8/06	12/0 6	3/07	4/07	5/07	6/07	7/07	8/07
launch	••••	•				••••				•	
inactive		••••	••••	••••	••••		•••	•••	••••	••••	••••
displaced						••••				•••	
deleted							••	••			
<b>trigger pulses</b>	article launch					writing contest				featured article	

Table 1: Movement of information units in “Zitronenpresse”

The methodological approach and applied code system made it possible to locate selected fillers and allowed statements about changes (i.e. if and when they had been made). As the slot-filler-combination is not likely to change very much in an article that deals with an artefact (Table 1), it changes more likely in an event-related topic (Table 2). Trigger pulses can be identified here, too, but the constant relevance of the topic is noticeable as well in the recurrent launch of new information units. Furthermore, it can be observed how some parts of the article content become more inactive or stable for some time.

date / movement	2/10	5/10	8/10	11/1 0	2/11	5/11	8/11	11/11	2/12
launch	••••	••••	•••	••	•••	••	••	••	••
inactive		•	••	••••	••••	••••	••••	•••	••
displaced		••	••••	•	••	•		•	••••
deleted			•		•	••		••	•••
<b>trigger pulses</b>	media coverage/ sovereign default				Operations by the EFSF				fiscal compact

Table 2: Movement of information units in “Eurokrise”

Trajectories and patterns of an interconnection of endogenous and exogenous factors are, in fact, visible in a feedback loop, e.g. when activity rises due to a featured article process, or ongoing events. As mentioned above, real-world events do affect the number of edits performed on an article; along with these come qualitative changes, which can be qualified on different linguistic levels. Results show that the underlying concept of each article, according to the conceptual frame approach applicable to

either an artefact or an event, is likely to be revised in its components after a trigger pulse.

The development of the articles also showed that the encyclopaedic character of entries (i.e. by stressing information about geographical place-names or names of important persons) evolves only over time. The encoding paradigm helps to set the focus on entities (as can be seen, for example, in numerous references to significant events in time, relevant places, cultural or public figures). Numerous fillers have been identified here, but other reference points or slots were also considered in later versions of an article covering different fields of knowledge representation (Table 3).

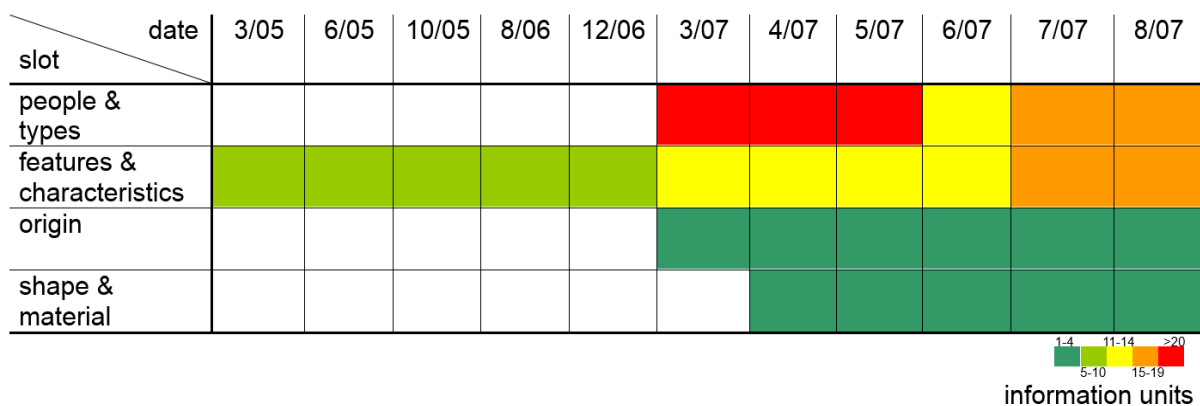


Table 3: Relevant areas of knowledge representation in “Zitronenpresse”

So far, the approach has proved to be useful in the identification of some indicators for interactions and activities within a wiki. However, to understand what information flows into the activated frame and what is relevant for an understanding of a ‘lemon squeezer’ or the ‘European debt crisis’, it can be helpful to enter deeper layers of the information units to identify key elements in the filler-slot structure. As already pointed out, ‘lemon squeezer’ refers to a frame around an artefact that is a kitchen utensil. In terms of this particular frame, high type and token frequencies over a significant period of time within this frame allow us to assume possible stable components or ‘entrenchments’ (cf. Ziem, 2014: 292–299). In fact, the slot ‘features & characteristics’ operated with different fillers: A lemon squeezer is used to make juice; is used for different citrus fruits; is designed to separate the pulp, etc. The phenomenon of a high type-frequency should be considered here as it underlines the importance of the slot ‘features & characteristics’ for the Zitronenpresse frame.

In the article ‘European debt crisis’ the consolidation of a token ‘Greek debt crisis’ was quite noticeable. In the German article versions the filler ‘Greek debt crisis’ could be placed in the slot ‘occurrence’ as Greece was one of the first countries to show a budget deficit. But the filler also matched the slots ‘correlations’ and ‘interference’ as budget crisis in Greece and beyond spread and bailout measures as well as Greece withdrawal

from the Eurozone were discussed. Also, a hyperlink ‘Greek debt crisis’ was recurrently used in the “see also” section as it relates to a topic similar to the discussed one in the article ‘European debt crisis’. However, a high token frequency consolidates the filler but weakens the slot. This means that the answer to a question “What is the European debt crisis?” may include the instance ‘Greek debt crisis’ as a sort of a default value. However, the exact description of this relationship remains open; at least in the examined article versions.

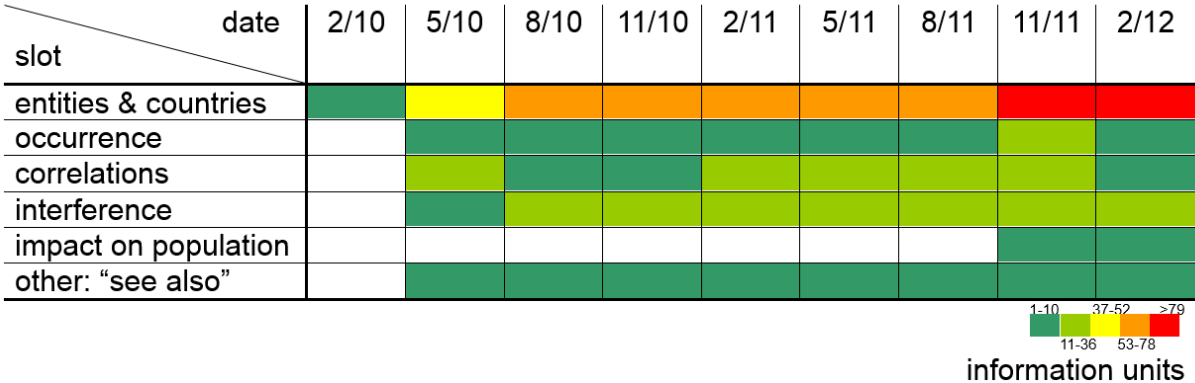


Table 4: Relevant areas of knowledge representation in “Eurokrise”

As we can see, the presented approach takes certain features of Wikipedia’s dynamics into account. Using this approach, I identified phases and interrelations, as well as some aspects of coherence in the interrelation of endogenous and exogenous factors when attention is drawn to the article and its development is triggered. The outlook on possible default values or the process of developing stable components is worth considering. Although the specific patterns of dynamic changes highlighted by this analysis will only be valid for a restricted period until the next edit, the principles derived from this approach should remain relevant and can be applied to other topics or information resources.

Of course, exploring the revisions of an entry is only one step in the multifaceted task of understanding what is important or relevant for both users and producers of a wiki. Certainly, this task needs a broad spectrum of research activities, for example dealing with general patterns of look-up behaviour (Müller-Spitzer et al., 2015), or classifying edits in collaboratively created articles (Daxenberger & Gurevych, 2013).

### 5. Relevance in lexicography

I previously pointed out that we can only use the results of so-called user generated content or bottom-up-lexicography for our own lexicographic products if we fully understand how the collaborative system works and what is important for the active user. So how can we use the understanding of the collaborative process that we have so far in institutional or professional lexicography? I want to emphasize three possible benefits of analysing Wikipedia dynamics:

- (1) Learning about the collaborative process.
- (2) Using an already existing collaborative product for expert lexicographic purposes.
- (3) Incorporating the collaborative process into an expert lexicographic product (or combining the two).

While pointing out some patterns and trajectories in the life cycle of an article within Wikipedia, we learned that trigger pulses as well as context sensitivity and coherence drive the development. Practical consequences of this are that information flow and information build up is subject to change according to the described factors. The more extensive but also potentially less stable contributions will be associated with whatever seems currently relevant. Thus, relevance has both positive and negative aspects for expert lexicographic purposes. However, in the long run, contributions that are highly contested or on the fringe of a topic will only have a short lifespan and will eventually be ‘overwritten’ during the article’s development, while more general, consensual information will remain. Such facets of the collaborative process should also be taken into account when using it for an expert lexicographic product. As Bon & Nowak (2013) emphasized, a procedure supplying entries with encyclopaedic or world knowledge can support text comprehension as well as (in my point of view) discourse comprehension.

Finally, a combination of direct user contribution via the collaborative process, e.g. in a semi-collaborative dictionary related either to object or language issues, and an expert lexicographic product should point out both more static and more dynamic views on the same topic. Effectively, as Lew (2011: 237) states, the opposition institutional versus collective dictionary may no longer be a sharp one. The discussed examples of Merriam-Webster’s Open Dictionary and the Macmillan Open Dictionary in Lew’s overview on English online dictionaries shows, however, that user-added entries do not meet the criteria for inclusion in the regular edition. However, Lew (2014: 25) and Taganova (2013) might agree on the point that “[t]he cooperation of readers and editors can turn beneficial for the dictionary compilers, as representatives of different interest groups and subcultures can make contribution to the Open Dictionary projects, indicating the words that lexicographers might miss out” (Taganova, 2013: 111). The lexical description of entire vocabularies, however, is a job better suited for language professionals (cf. Lew 2014: 17). A potential outlook is also to transfer the given approach to lexicography, by analysing the revisions of a collaborative dictionary entry as indicative of lexical change. However, additional work needs to be done in order to apply the insights gained from analysing dynamics of encyclopaedic-style Wikipedia entries to environments concerned with information on word-meaning and language comprehension. In any case, with recent advancements in user-generated environments, different views on language become available and may get users more actively interested in lexicographic work in general.

## 6. Conclusions

In this overview, I presented a short study on the developments in two articles from the German Wikipedia. By means of time series data, a certain pattern was observed which pointed to trajectories between endogenous and exogenous factors within Wikipedia's activity to produce and enhance articles. This pattern appeared to follow a life cycle with regard to the articles' entities.

I believe that this study, in particular the clarification of development patterns within articles, can contribute to a better understanding of collaborative induced dynamics. These results can be utilized when using Wikipedia entries or articles from other wikis for a different lexicographic product. The proposed model can also be used to predict the developments, thus facilitating the use of collaborative products in institutional lexicography. The model may also provide pointers to what is worth taking into account when using user-generated content. Finally, a combination of expert and collaborative knowledge should be considered when thinking about new lexicographic products.

## 7. Acknowledgement

The author would like to thank the reviewers for their valuable comments and helpful suggestions.

## 8. References

- Abel, A. & Meyer, Ch. (2013). The dynamics outside the paper: User Contributions to Online Dictionaries. *Proceedings of the 3rd Biennial Conference on Electronic Lexicography* (elex 2013). Ljubljana: Trojina, Institute for Applied Slovene Studies/Tallinn: Eesti Keele Instituut, pp. 179–194.
- Barrett, D. (2009). *MediaWiki*. Beijing: O'Reilly.
- Bon, B. & Nowak, K. (2013). Wiki Lexicographica. Linking Medieval Latin Dictionaries with Semantic MediaWiki. *Proceedings of the 3rd Biennial Conference on Electronic Lexicography* (elex 2013). Ljubljana: Trojina, Institute for Applied Slovene Studies/Tallinn: Eesti Keele Instituut, pp. 407–420.
- Bruns, A. (2008). *Blogs, Wikipedia, Second Life, and beyond*. From production to produsage. New York: Lang.
- Busemann, K. (2013). Wer nutzt was im Social Web? *Media Perspektiven*, 7-8, pp. 391–399.
- Carr, M. (1997). Internet Dictionaries and Lexicography. *International Journal of Lexicography*, 10(3), pp. 209–230.
- Crane, R. & Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. *PNAS* 105(41), pp. 15649–15653.
- Daxenberger, J. & Gurevych, I. (2013). Automatically classifying edit categories in Wikipedia revisions. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 578–589.
- Döring, N. (2010). Sozialkontakte online: Identitäten, Beziehungen, Gemeinschaften. In W. Schweiger & K. Beck (eds.) *Handbuch Online-Kommunikation*.

- Wiesbaden: Springer, pp. 159-183.
- Emigh, W. & Herring, S. (2005). Collaborative authoring on the Web: a genre analysis of online encyclopedias. *Proceedings of the 39th Hawaii International Conference on System Sciences* (HICSS). Track 4. Volume 04. Los Alamitos: IEEE Press.
- Ferron, M. & Massa, P. (2011a). Collective memory building in Wikipedia: the case of north African uprisings. *Proceedings of the 7<sup>th</sup> International Symposium on Wikis and Open Collaboration*, pp. 114–123.
- Ferron, M. & Massa, P. (2011b). Studying collective memories in Wikipedia. *Journal of Social Theory*, 3(4), pp. 449–466.
- Juran, J. (1992). *Juran on quality by design. The new steps for planning quality into goods and services*. New York: Free Press.
- Kallass, K. (2015). *Schreiben in der Wikipedia: Prozesse und Produkte gemeinschaftlicher Textgenese*. Wiesbaden: Springer.
- Kaltenbrunner, A. & Laniado, D. (2012). There is no deadline: time evolution of Wikipedia discussions. *Proceedings of the 8th Annual International Symposium on Wikis and Open Collaboration*, A8.
- Keegan, B. et al. (2011). Hot off the wiki: dynamics, practices, and structures in Wikipedia's coverage of the Tōhoku catastrophes. *Proceedings of the 7<sup>th</sup> International Symposium on Wikis and Open Collaboration*, pp. 105–113.
- Konerding, K.-P. (1993). *Frames und lexikalisches Bedeutungswissen*. Tübingen: Niemeyer.
- Lehmann, J. et al. (2012). Dynamical classes of collective attention in Twitter. *Proceedings of the 21st international conference on World Wide Web*, pp. 251–260.
- 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. (2014). User-generated content (UGC) in online English dictionaries. *OPAL - Online publizierte Arbeiten zur Linguistik* 2014.4, pp. 8–26.
- Lih, A. (2004). Wikipedia as Participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource. *Paper for the 5th International Symposium on Online Journalism*. University of Texas at Austin.
- Mayer, F. (2013). *Erfolgsfaktoren von Social Media: Wie „funktionieren“ Wikis?* Berlin: LIT.
- Mederake, N. (2014). Artikel der Wikipedia aus lexikografischer und textlinguistischer Perspektive. In M. Mann (ed.) *Digitale Lexikographie*. Hildesheim: Olms, pp. 229–249.
- Meyer, Ch. (2013). *Wiktionary: The Metalexigraphic and the Natural Language Processing Perspective*. Darmstadt.
- Müller-Spitzer, C. et al. (2015). Observing online dictionary users: studies using Wiktionary log files. *International Journal of Lexicography*, 28(1), pp. 1–26.
- Stivilia, B. & Gasser, L. (2008). An activity theoretic model for information quality change. *First Monday*, 13(4). Available at:

<http://firstmonday.org/article/view/2126/1951>.

- Stvilia, B. et al. (2005a). Information quality discussions in Wikipedia. *Technical Report ISRN UIUCLIS--2005/2+CSCW*.
- Stvilia, B. et al. (2005b). Assessing information quality of a community-based encyclopedia, In F. Naumann, M. Gertz, & S. Mednick (eds.). *Proceedings of the International Conference on Information Quality (ICIQ 2005)*. Cambridge, Mass.: MIT, pp. 442–454.
- Stvilia, B. et al. (2008). Information quality work organization in Wikipedia. *JASIST*, 59(6), pp. 983–1001.
- Taganova, T. (2013). New Words in Contemporary Dictionaries of the English Language: Are Words Invented by the Society or is the Society Changed by Words? In O. Karpova & F. Kartashkova (eds.) *Multi-disciplinary Lexicography: Traditions and Challenges of the XXIst Century*. Newcastle upon Tyne: Cambridge Scholars, pp. 103–113.
- Vossen, G. & Hagemann, S. (2007). From Version 1.0 to Version 2.0: A Brief History of the Web. In J. Becker et al. (eds.). *ERCIS Working Papers*, Vol. 4. Available at: [https://www.ercis.org/sites/www.ercis.org/files/pages/research/ercis-working-papers/ercis\\_wp\\_04.pdf](https://www.ercis.org/sites/www.ercis.org/files/pages/research/ercis-working-papers/ercis_wp_04.pdf).
- Wang, R. & Strong, D. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), pp. 5–33.
- Wiegand, H. E. et al. (2010). *Wörterbuch zur Lexikographie und Wörterbuchforschung*. Berlin: De Gruyter.
- Ziem, A. (2014). *Frames of understanding in text and discourse: theoretical foundations and descriptive applications*. Amsterdam: Benjamins.

### Websites:

- Algemeen Nederlands Woordenboek*. Accessed at: <http://anw.inl.nl/> (6 July 2015)
- BuzzWords*. <http://www.macmillandictionary.com/buzzword/> (22 May 2015)
- elexiko*. <http://www.owid.de/wb/elexiko/gruppen/index.html>. (22 May 2015)
- Macmillan's Open Dictionary*. Accessed at: <http://www.macmillandictionary.com/open-dictionary/>. (22 May 2015)
- Merriam-Webster's Open Dictionary*. Accessed at: <http://nws.merriam-webster.com/opendictionary/>. (22 May 2015)
- Wikipedia*. Accessed at <https://de.wikipedia.org/>. (March 2005 – February 2012)

This work is licensed under the Creative Commons Attribution ShareAlike 4.0 International License.

<http://creativecommons.org/licenses/by-sa/4.0/>

