Mapping the Morphing Meanings: Identification and Representation of Semantic Shifts in Lexicographical Works

Seda Hovhannisyan, Utpal Pandey
European Master in Lexicography students,
University of Erlangen-Nuremberg, Germany
sed.hovhannisyan456@gmail.com, utpalpandey20@gmail.com

Abstract

The constant evolution of language involves the frequent emergence of novel meanings within established words, facilitating the extension and adaptation of vocabulary. A scrutiny of lemmas in the *Neologismenwörterbuch* (Dictionary of Neologisms) of *Leibniz-Institut für Deutsche Sprache* (IDS)[1] reveals that over 8% of neologisms in German fall under the category of semantic neologisms. This paper explores the landscape of semantic neology and proposes efficient strategies for lexicographers to swiftly incorporate contemporary semantic shifts into dictionaries, particularly on online platforms.

This article pursues two main objectives: firstly, the identification of semantic neology. To demonstrate this, three words have been retrospectively selected from the corona wordlist of IDS.[2] To achieve this the paper investigates the current state of computational methods for meaning detection and their potential integration into lexicographic practices; secondly, it suggests a user-friendly feature for online dictionaries that facilitate the rapid retrieval of new meanings. In an effort to distinguish between processes and outcomes of semantic shift, the paper also briefly discusses the existing literature on semantic neology, primarily drawing on the works of Bechmann (2013, 2016) and Keller (2003).

Keywords: semantic neology; new meanings in dictionaries; meaning detection; distributional semantics; corpus lexicography

1 Out of 2480 lemmas in the Dictionary of Neologisms by Leibniz Institute for German Language 201 lemmas have acquired a new meaning. https://www.owid.de/docs/neo/start.jsp (Retrieved on January 25, 2024).

2 The corona wordlist compiled by IDS can be accessed through the following link, titled Neuer Wortschatz rund um die Coronapandemie (New Vocabulary Related to the Coronavirus Pandemic): https://www.owid.de/docs/neo/listen/corona.jsp (Retrieved on January 25, 2024).

1. Introduction

A natural language is not a natural phenomenon, not a natural organism […]. A natural language is not an artefact. Like inflation or a traffic jam, it is the unintended cumulative consequence of a countless number of intentional communicative acts by countless people. It is one of those phenomena which are indeed the result of human action, but not the execution of any human design […]. (Keller 1989: 115)

Although individuals do not have a holistic control on a language, they can still wield it as an organon, a tool to attain their objectives in the extralinguistic world, going beyond a mere exchange of factual information about it. This pursuit involves the strategic use of appropriate signs, intricately complemented by the nuances of extralinguistic settings. Sometimes these strategies result in a semantic shift.

Bechmann (2013) recognizes that semantic shift follows very clear pragmatic principles, which can be structurally understood through the displacement and incorporation of parameters of meaning. Semantic change is the main type of lexical change, which does not result in formal neologisms (cf. Grzega 2004: 63) and manifests itself as a special case of language change at the word level (cf. Bechmann 2013: 90). This does not cause coinage of a new form, as indicated above, but association of an existing lexical form with a previously unregistered meaning. In other words, it is an innovation or a new association on the signified side (cf. Marečková 2011: 56).

The treatment of new meanings presents a peculiar challenge in neologism lexicography. Kinne (1996: 345) describes the process of development of new meanings as much more protracted and difficile and therefore often more difficult to identify than the emergence of new lexemes. Lemnitzer (2010: 72; translation: UP) makes a similar claim in the context of digital neologism lexicography:

A new meaning of an existing form cannot be identified with modern language technology. The detection of word meanings based on the distribution of the word is the subject of research in computational linguistics. However, since a new meaning of a particular word is always initially based on a small number of occurrences – the new meaning is only just beginning to establish itself – the distributional method for detecting a new meaning is unlikely to be very successful.

However, in recent years there has been a surge in studies on computational methods of semantic shift detection. Thanks to the availability of large text corpora these methods have shown significant improvement. A universally accepted procedure for automatically identifying novel senses through corpora, however, has not yet been established (Renau 2023; Schlechtweg et al. 2022).

Thus, the primary aim of this paper is to trace the advancements in automatic semantic shift detection since Lemnitzer's bold prediction in 2010 and assess their relevance to the field of lexicography. A secondary aim is to propose a feature for online lexicographical products that facilitates the swift inclusion and retrieval of new meanings.

2. Semantic Shift: Its Processes and Outcomes

"The term shift implies a time interval $t_0 – t_1$ and either the substitution of elements by others [A (t_0) – B (t_1)] or the change of properties of elements [X $(A)t_0$ – Y $(A)t_1$] in the

time interval in question" Lüdtke (1984: 731). In the context of semantic shift, it means that there has been change or shift in signified of the sings. "[...] This change does not take place abruptly, but also not completely continuously, it includes a series of small steps, but with widespread limitation on cognitive-pragmatic parameters" (cf. Blank 2005: 1325). These steps are needed because of the various extralinguistic and linguistic factors. There is no clear answer as to how this pragma-semantic creativity of individuals at the micro level is accepted by the language community. However, the processes and outcomes of semantic shift can be identified and described.

2.1 Processes of Semantic Shift

"Semantic shift is [...] the result of a relation, coupled with parameters of meaning, between the choice of linguistic device and the intended purpose of the speaker" (Bechmann 2013: 327). When a word is assigned a different meaning by means of a linguistic device, this new meaning might become conventionalized over time, i.e. it gets detached from the context. This process of detachment of a meaning from its context is called the process of semantic shift. Within a language community, speakers employ diverse methods to introduce a word into a novel context, thereby altering the usage norms of the word. (1) processes based on similarity, such as metaphorization and cohyponymic transfer; (2) processes based on contiguity, such as metonymization and synecdoche; (3) processes based on contrast, such as euphemismization, antiphrasis, and auto-antonymy; and (4) processes based on linguistic economy, such as ellipses and implicators.

2.2 Outcomes of Semantic Shift

The processes of semantic change not only alter the semantic properties of words but also impact the language level. The outcomes of semantic shift at the word level can be classified into two groups: (1) quantitative changes and (2) qualitative changes. The quantitative changes include broadening and narrowing and the qualitative changes include amelioration and pejoration.

The processes of semantic change can have various consequences at language level. The semantic change gives rise to various semantic relations, including ambiguity and polysemy at the language level. When a new meaning emerges, it does not immediately replace the existing one. The coexistence of old and new meanings, and thus at least temporary polysemy in the language's vocabulary, is always the norm in the process of semantic change (cf. Bechmann 2016: 254). When a single lexeme carries multiple meanings, it can, in specific contexts, lead to misunderstandings. Additionally, societal changes can render certain meanings obsolete. Consequently, there may occur a reversal of polysemy, where older meanings fade away. Keller (2003) highlights examples such as *ängstlich* (anxious) and *edel* (noble), where the erstwhile meanings of "fear-inducing" and "aristocrat" have disappeared.

Semantic change can also cause the grammaticalization of the lexeme. The transition from verbs to auxiliary verbs is evidence of such development. In these cases, a linguistic unit gradually loses its lexical meaning and is increasingly used as a morphosyntactic marker. Bechmann (cf. 2016) proposes two types of grammaticalization: grammatical paradigmaticization (e.g., the development of *brauchen* (need) as a modal verb) and diathesis

alternation (e.g., the change in the direction of action in verbs like *entschuldigen* (to apologize) and *erschrecken* (to frighten).

3. Computational Methods of Meaning Detection

For the last 15 years, semantic change has also been a topic of interest amongst computational linguists (Navigli 2009; Lau et. al. 2012). The foundation for all methods in this field lies in the well-known Firthian distributional hypothesis: "You shall know a word by the company it keeps" (Firth 1962: 11). "In other words, difference of meaning correlates with difference of distribution" (Harris 1954, 156). Thus, in all the computational linguistics models, regardless of the fact whether they use pure co-occurrence computing or contextualized embedding methods, a word's meaning or senses rely on the context in which they appear in a written or spoken corpus.

The detection of semantic neology through computational tools involves at least three steps. First of all, the potential candidates need to be identified, which might have undergone semantic change. It is followed by the representation of the meaning of the selected words that is word embeddings using a computational model of choice at two points in time and in the end, these two representations are compared to discern any changes in the meaning. In essence, the accuracy of detecting new meanings with computational models is intricately tied to the precision of the word embedding representation achieved by the chosen model.

Existing approaches for lexical semantic change detection are mainly based on three types of meaning representations: (i) semantic vector spaces, (ii) topic distributions, and (iii) sense clusters. In semantic vector spaces, each word is represented as two vectors reflecting its co-occurrence statistics at different periods of time. Semantic change is typically measured by the cosine distance (or some alternative metric) between the two vectors, or by differences in contextual dispersion between the two vectors. In semantic change detection, cosine distance is a measure used to quantify the similarity or dissimilarity between word embeddings or vectors representing the meanings of words. It calculates the cosine of the angle between two vectors, providing a numerical value that indicates how similar or different the meanings of the words are in the given context. Lower cosine distances typically imply greater similarity between word meanings, while higher distances suggest greater dissimilarity.

The topic models infer a probability distribution for each word over different word senses or topics, which are in turn modeled as a distribution over words. The semantic change is measured by calculating a novelty score for its senses based on their frequency of use. Clustering models assign all uses of a word into sense clusters based on some contextual property; it is also called contextualized embedding. Word sense clustering models are similar to topic models in that they map usage to senses. Accordingly, the semantic change of a word is measured similarly as in topic models. The vector space model works at the word level, whereas the topic distribution models and sense cluster models work at the sense or meaning level. Sense cluster models or contextualized word embeddings are used by the massive corpus-based NLP models like GPT by Open AI. For a detailed overview of current computational methods of meaning detection see Tahmasebi et al. (2021).

There have been efforts to incorporate the innovations in computational linguistics in context of semantic shift in lexicography at least since 2013 (Cook et al. 2013). Even after the innovations in recent years, these methods face several challenges. One challenge arises from the distributional hypothesis itself, as it often conflates lexical meaning with cultural and topical information available in the corpus used as a basis for the model. Moreover, only a limited number of approaches currently propose techniques capable of analyzing semantic change in words with relatively few occurrences, and the existing methods often operate at a relatively coarse time granularity. This is primarily due to the inherent complexity associated with incorporating multiple time bins into the analysis (cf. Hengchen et al. 2021: 349).

Given that contemporary semantic shifts initially manifest with sparse corpus evidence, achieving fully automated detection of new word senses in real time becomes practically unattainable. These endeavors falter at the outset, specifically in identifying potential candidates. In light of this challenge, it seems more logical that lexicographers manually identify words that may have undergone semantic shifts, while making use of available technology to substantiate their intuitions. One recommended strategy involves making use of monitor corpora and paying attention to words that have recently surged in frequency, however this method is not foolproof, as there can be lexemes, which can undergo semantic change without any visible change in frequency.

4. Semantic Shift during Corona Pandemic

During the corona pandemic languages across the world experienced the innovation of speech community and a huge number of formal as well as semantic neologisms have been observed. The German language was no different. Both formal and semantic neologisms in the German language during the corona pandemic have been well documented by IDS and DWDS.[3] As mentioned above, for this paper three words were selected, namely (1) *Dauerwelle*, (2) *Babyelefant* and (3) *Bürgertest*. First of all, we will present a frequency analysis with the help of *DWDS Verlaufskurven*.[4] Next, we present an analysis of context with the help of a corpus created on the Sketch Engine.

4.1 Frequency Analysis

In the case of all these three words, a clear spike could be identified between 2019 and 2020. This makes them good potential candidates for our purpose.

3 DWDS-Themenglossar zur COVID-19-Pandemie of the Digital Dictionary of the German Language by Berlin-Brandenburg Academy of Sciences. https://www.dwds.de/themenglossar/Corona (Retrieved on January 25, 2024)

4 DWDS Verlaufskurven (DWDS Plot) is a tool offered by the Digital Dictionary of the German Language. https://www.dwds.de/r/plot/?q= (Retrieved on January 25, 2024)

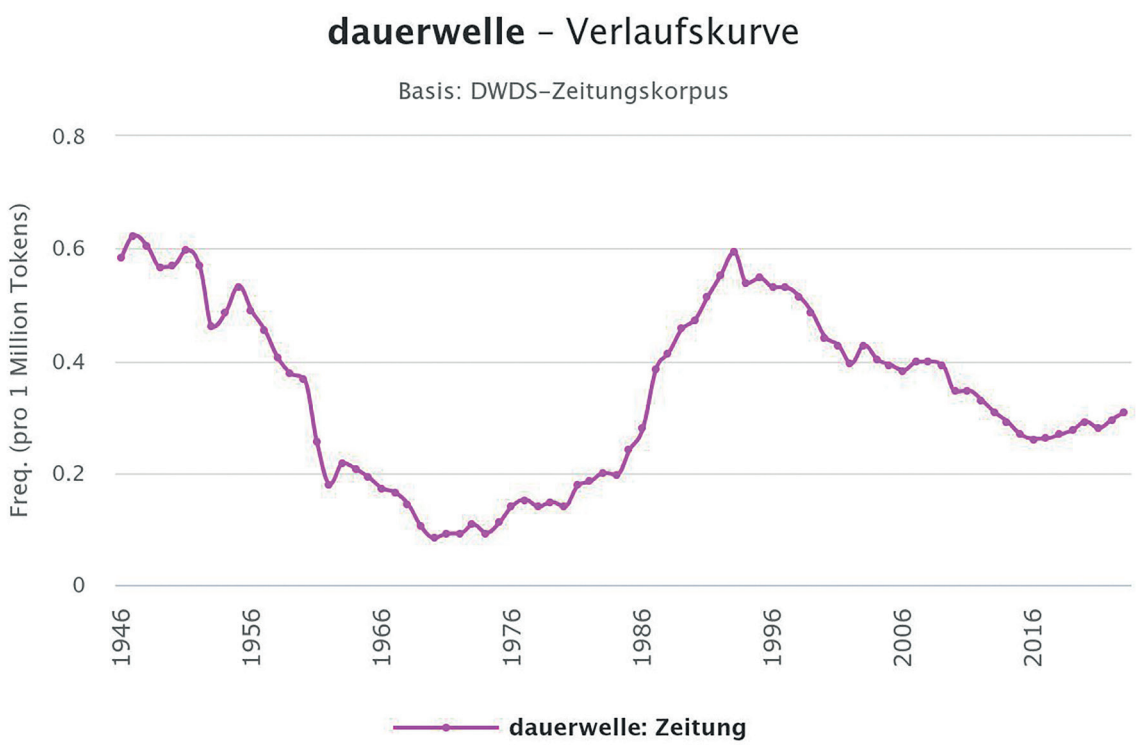

Figure 1: *Frequency Curve for Dauerwelle.*

The word *Dauerwelle* (perm) hit its peak in the 1950s and then in the 1990s (Figure 1), due to the popularity of perms. Since then, the frequency of the word has been on decline. However, if one looks at this graph closely a slight surge could be seen in 2019 and 2020. Since this graph is generated for a large time period this spike is not that pronounced. Similar results can be seen for *Babyelefant* (baby of an elephant) and *Bürgertest* (test by citizen).

4.2 Context Analysis

For this analysis we have created a corpus using the Sketch Engine with texts related to the corona pandemic. This corpus included more than 6 million tokens and over 250,000 unique words. Contrary to our expectations, the collocational profiles generated with our corpus did not always show a clear shift in the context of the words. At first glance, all we could see was the old context, even in a specialized corpus. For example, for *Dauerwelle* the collocational profile does not give any indication towards a semantic shift (Figure 2). The profile is almost identical to a profile generated with any other large corpora for the German language on the Sketch Engine. Two primary factors influence this result: (1) the processing methodology and presentation employed by the Sketch Engine, and (2) the quality of the data within the corpus itself. While the former can be addressed by employing Python or R for enhanced analytical control (cf. Gries 2015), finding a resolution to the latter remains a topic of debate.

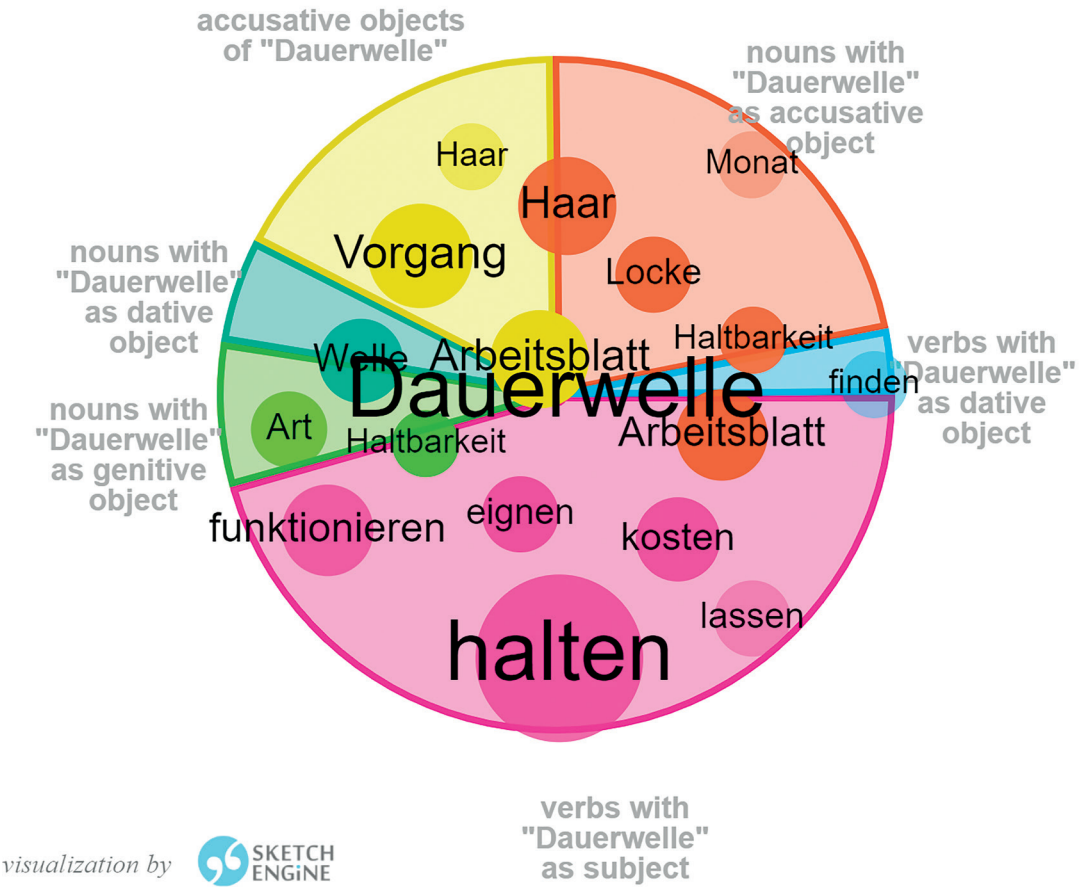

Figure 2: *Context Analysis of Dauerwelle.*

For the other two words, however, the story is different. In the case of *Babyelefant*, the profile includes words like *Abstand* (distance), *Meter* (meter), which show its proximity to preventive measures during the pandemic. It does not mean just 'baby of an elephant' rather this word has now gotten a new semantic feature based on the similarity association and means a distance of one or one and a half meters in the context of corona prevention measurements.

The same holds true for *Bürgertest*. It begins to appear alongside nouns such as *Antigen-Schnelltest* (rapid antigen test), *kostenlos*, and *kostenfrei* (both meaning free of cost), signalling a shift in its meaning. Yet, upon closer scrutiny, it becomes apparent that this term represents a formal neologism. Despite the similarity in form, the new compound

noun *Bürgertest* stems from a word-formation process rather than a process of semantic shift. Interestingly, the DWDS treats the lexeme *Bürgertest* [5] as polysemous.

5. Representation of Contemporary Semantic Shift in Dictionaries

In the current lexicographic practices, there are two main ways to include a new meaning in the dictionaries: (1) in a dictionary of neologisms, (2) in vocabulary lists in online dictionaries, similar to one maintained by IDS for corona vocabulary. We would like to propose a notification feature, which allows a rapid inclusion of new meanings in general purpose online dictionaries.

The implementation of a notification feature in online dictionaries can be very useful to inform users about new meanings. This means that if a user searches for a word that has recently undergone a semantic change, a notification is shown to inform the user about it. Below is a modified the entry page for "*Dauerwelle*" in DWDS (Digital Dictionary of the German Language) [6] to demonstrate how such a notification system could be integrated into online dictionaries (Figure 3).

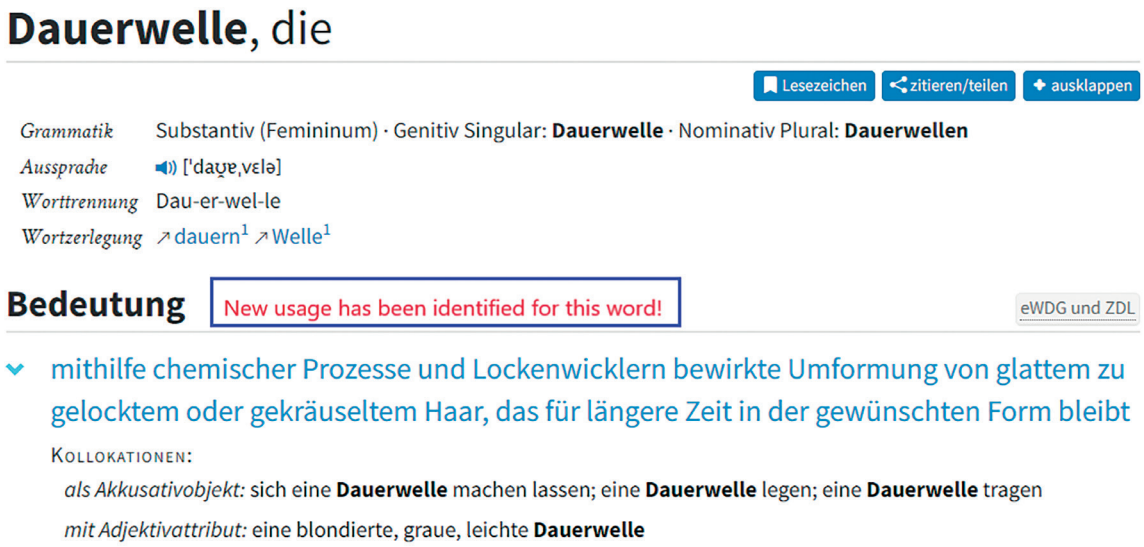

Figure 3: *Modified Entry for "Dauerwelle" in DWDS.*

Once the user sees this notification, they can hover on the notification box or click on this, if other meanings in the dictionary do not satisfy their query or out of curiosity. At this stage the users can be shown the paragraph of meaning and just hovering their pointer on

5 Bürgertest, made available by the Digital Dictionary of the German Language, <https://www.dwds.de/wb/B%C3%BCrgertest>, retrieved on February 9, 2024.

6 DWDS – Digitales Wörterbuch der deutschen Sprache. Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart, edited by Berlin-Brandenburg Academy of Sciences, <https://www.dwds.de/>, retrieved on January 25, 2024.

the notification and clicking it should lead them to vocabulary list or something similar where the users can access example sentences and other data related to that particular sense.

This strategy could also be useful in the cases, where a specific meaning is prevalent only for a short period of time. For instance, numerous word senses emerged during the corona pandemic have since fallen out of use, including the new meaning associated with *Babyelefant*. In these instances, the notification feature can be deactivated if deemed necessary.

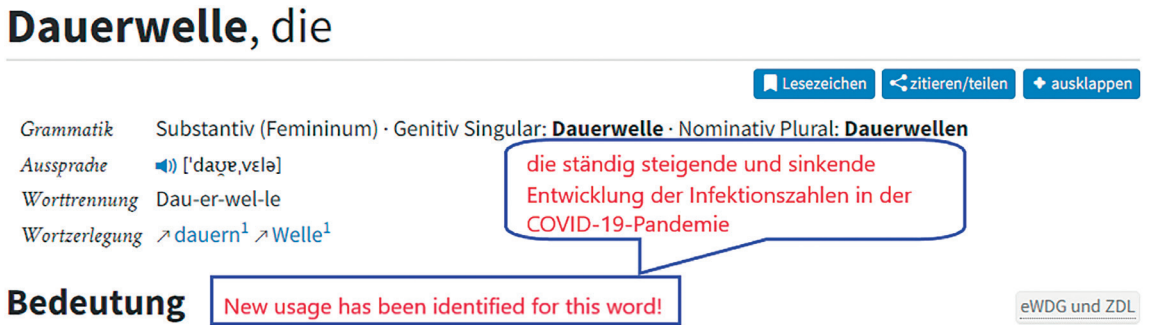

Figure 4: *Paraphrase of New Meaning.*

If a new meaning solidifies its usage, it should be transitioned to the main section of comment on semantics, initially labeled as "new meaning" or something similar. This approach aligns with the practice of labeling archaic or obsolete words. Similarly, formal neologisms can also carry such labels for a certain duration.

6. Conclusion

In this paper we clearly distinguished between the processes and outcomes of the semantic shift and categorized them into clearly identifiable groups. We identified four distinct groups of processes of semantic change. Similarly, we demonstrated the outcomes of the semantic shift, both at the level of language and word. At the word level we established that words can undergo either qualitative or quantitative changes and at the level of language polysemy and grammaticalization can be observed.

The scrutiny of computational methods for meaning detection has revealed recent advancements while recognizing the hurdles encountered in attaining fully automated detection, especially given the sparse evidence inherent in contemporary semantic shifts. This paper endorses a nuanced, human-in-the-loop approach for lexicographers, combining manual discernment with technological tools to proficiently identify and analyze semantic shifts. Furthermore, our manual analysis of the three words underscores that corpora can occasionally yield confusing results both in terms of frequency and context, emphasizing the valuable role of a lexicographer's expertise in such instances.

Based on our observation, we proposed a notification feature for online dictionaries, as a way of enhancing the user-friendliness. The integration of a notification system stands out as a proactive measure to apprise users of recent semantic shifts. This feature has the capacity to transform any general-purpose online dictionary into a dictionary of neologisms, if the users have the option to use filters.

Storrer and Freese (1996: 97-98) observed in their analysis of internet lexicography at the time that online dictionaries failed to fully harness the design potential of the online medium. Remarkably, nearly 30 years later, this observation still holds true. Although the online platform offers a range of dynamic tools for presenting lexicographic data, hardly any dictionary has fully embraced these innovative approaches. Along with quality and authenticity, there is also a need for devising more features, which enable quick retrieval of lexicographic data and provide users more freedom to adapt the presentation of data on screen.

One major complexity with this approach, as this process has to take place in real time, involves the discussion regarding conventionalization, which make a meaning suitable for inclusion in the dictionary. Since there is no universally accepted parameter to measure the degree of conventionalization, each dictionary's editorial board can establish its own criteria for both inclusion and removal of a particular meaning.

References

Bechmann, S. (2013). *Bedeutungswandel deutscher Verben: Eine gebrauchstheoretische Untersuchung.* Narr.

Bechmann, S. (2016). *Sprachwandel – Bedeutungswandel*. utb.

Blank, A. (2005). Etymologie und Wortgeschichte III: Neue Zugänge zu semantischem Wandel. In D. A. Cruse, F. Hundsnurscher, M. Job & P. R. Lutzeier (eds.): *Lexikologie: Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen* (2. Halbbd. (HSK 21.2), pp. 1324-1332). De Gruyter Mouton. https://doi.org/10.1515/9783110171471.2.29.1324

Cook, P., Lau, J.H., Rundell, M., McCarthy, D., & Baldwin, T. (2013). A lexicographic appraisal of an automatic approach for detecting new word-senses. In I. Kosem, J. Kallas, P. Gantar, S. Krek, M. Langemets, M. Tuulik (eds.), *Proceedings of the eLex 2013 Conference* (pp. 49 – 65). Trojina, Institute for Applied Slovene Studies & Eesti Keele Instituut.

Firth, J. R. (1962). A synopsis of linguistic theory, 1930–1955. In *Studies in Linguistic Analysis* (pp. 1–32). Basil Blackwell.

Gries, S. T. (2015). Some Current Quantitative Problems in Corpus Linguistics and a Sketch of Some Solutions. *Language and Linguistics 16*(1), 93–117. https://doi.org/10.1177/1606822X14556606

Grzega, J. (2004). Bezeichnungswandel: Wie, Warum, Wozu? Ein Beitrag zur englischen und allgemeinen Onomasiologie. Universitätsverlag Winter.

Harris, Z. S. (1954). Distributional Structure. *Word 10*(2-3), 146-162. https://doi.org/10.1080/00437956.1954.11659520

Hengchen, S., Tahmasebi, N., Schlechtweg, D., & Dubossarsky, H. (2021). Challenges for Computational Lexical Semantic Change. In N. Tahmasebi, L. Borin, A. Jatowt, Y. Xu & S. Hengchen (eds.), *Computational Approaches to Semantic Change* (pp. 341–372). Language Science Press. https://doi.org/10.5281/zenodo.5040241

Keller, R. (1989). Invisible-Hand Theory and Language Evolution. *Lingua 77*, 113-127. https://doi.org/10.1016/0024-3841(89)90011-9

Keller, R., & Kirschbaum, I. (2003). *Bedeutungswandel: Eine Einführung*. De Gruyter. https://doi.org/10.1515/9783110895315

Kinne, M. (1996). Neologismus und Neologismenlexikographie im Deutschen: Zur Forschungsgeschichte und zur Terminologie, über Vorbilder und Aufgaben. *Deutsche Sprache, 24*(4), 327-358.

Lau, J. H., Cook, P., McCarthy, D., Newman, D., & Baldwin, T. (2012). Word sense induction for novel sense detection. In *Proceedings of 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 591–601). The Association for Computational Linguistics.

Lemnitzer, L. (2010). Neologismenlexikographie und das Internet. *Lexicographica 26*, 65–78. https://doi.org/10.1515/9783110223231.1.65

Lüdtke, H. (1984). Ansätze zu einer Theorie des Sprachwandels auf phonologischer Ebene. In W. Besch, O. Reichmann & S. Sonderegger (eds.), *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Forschung* (1. Halbbd. (HSK 2.1), pp. 731–738). de Gruyter.

Marečková, I. (2011). Neologismus in den Pressetexten: Zum Konzept des Neologismus in diesem Beitrag. *Slowakische Zeitschrift für Germanistik 1*(2). 56–66.

Navigli, R. (2009). Word Sense Disambiguation: A Survey. *ACM Computing Surveys 41*(2), 1–69. https://doi.org/10.1145/1459352.1459355

Renau, I. (2023). A corpus-based study of semantic neology of the Covid-19 pandemic. *Quaderns de Filologia: Estudis Lingüístics 28*, 55–76. https://doi.org/10.7203/qf.0.26663

Schlechtweg, D., McGillivray, B., Hengchen, S., Dubossarsky, H., & Tahmasebi, N. (2022). SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation* (pp. 1–23). International Committee for Computational Linguistics. https://doi.org/10.18653/v1/2020.semeval-1.1

Storrer, A., & Freese, K. (1996). Wörterbücher im Internet. *Deutsche Sprache 24*(2), 97-153.

Tahmasebi, N., Borin, L., & Jatowt, A. (2021). Survey of computational approaches to lexical semantic change detection. In N. Tahmasebi, L. Borin, A. Jatowt, Y. Xu & S. Hengchen (eds.), *Computational Approaches to Semantic Change* (pp. 1–91). Language Science Press. https://doi.org/10.5281/zenodo.5040241

Acknowledgments

This study was supported by the European Master in Lexicography (EMLex).