Title: Determining the Semantic Features of Negative Emotive Words using a Computational Semantic Analysis Approach

Authors and Affiliations:

Martina Katalin Szabó (Corresponding author)
Institute of Informatics, University of Szeged, Szeged, Hungary
Computational Social Science – Research Center for Educational and Network Studies (CSS-RECENS), Centre for Social Sciences, Budapest, Hungary
ORCID: 0000-0002-4192-4352
Email: martina@inf.u-szeged.hu, Szabo.Martina@tk.hu

Péter Sánta
Institute of Informatics, University of Szeged, Szeged, Hungary
Email: speter@inf.u-szeged.hu

Balázs Nagy
Institute of Informatics, University of Szeged, Szeged, Hungary
Email: bnagy@inf.u-szeged.hu

László Vidács
Institute of Informatics, University of Szeged, Szeged, Hungary
Email: lac@inf.u-szeged.hu

Short Title: Semantic Evolution of NEWs

Statements and Declarations:

Financial interests: The authors declare they have no financial interests. Non-financial interests: none. The Author(s) declare(s) that there is no conflict of interest.

Availability of data: The authors confirm that the data supporting the findings of this study are available as supplementary material. Further derived data supporting the findings of this study are available from the corresponding author on request. Corpus sharing is not applicable to this article as no new corpus data were created or analyzed in this study.

Short title:

Analyzing Negative Emotive Word Features: A Computational Semantic Approach

Abstract

This paper employs word embedding to analyze Negative Emotive Words in Hungarian, which initially carry negative meanings but can change contextually (e.g., *terribly nice, the party was terrific!*). These words undergo desemantization, losing negative connotations while acquiring new functions. Our study has two objectives: first, to identify unique semantic traits of Negative Emotive Words using lexicons and word embedding, comparing them to general semantic features; second, to track their semantic shifts over two decades. We examine how suffixes impact these shifts, noting that not all Negative Emotive Words evolve simultaneously. This research, grounded in linguistic theories such as the Law of Parallel Change and the Law of Differentiation, explores whether computational semantic analysis can provide new insights into the linguistic nature and evolution of these words, particularly in low-resource languages like Hungarian.

Keywords: Negative Emotive Words, Semantic-pragmatic development, Word-embedding methods, Intensifiers, Corpus analysis

Acknowledgements:

The research work of Martina Katalin Szabó was funded by the OTKA Postdoctoral Research Grant, the Hungarian Research Fund of the National Research, Development and Innovation Office of Hungary (NKFIH, grant number: PD 132312) and by the International Research Fellowship Program of Japan Society for the Promotion of Science (JSPS, Postdoctoral Fellowships for Research in Japan (Standard)).

This work was supported by the Centre for Social Sciences, the Eötvös Loránd Research Network, the Hungarian Academy of Sciences Centre of Excellence, the University

of Szeged, Department of Informatics and the Institute of Global Studies, Tokyo University of Foreign Studies, Japan.

Determining the Semantic Features of Negative Emotive Words using a Computational Semantic Analysis Approach

1. Introduction

By definition, *Negative Emotive Words* (henceforth NEWs) inherently carry negative semantic content when used in isolation, but they may partially or entirely lose this meaning (as well as the negative connotation) depending on context, as illustrated by examples such as *terribly nice* or *the party was terrific!* (Szabó 2019). The presence of NEWs appears to be a feature not limited to any particular language. Furthermore, the existence of this word group, characterized by its specific semantic properties, can be observed diachronically (Wierzbicka 2002; Jing 2007; Andor 2011; Szabó and Otani 2022).

In the literature, NEWs are primarily discussed within the category of intensifiers, with a focus on their function as intensifiers, as seen in examples such as *terribly nice* or *awfully good* (Chang, 2020). Intensifiers serve to amplify or scale the degree of the quality expressed by the word they modify in a given utterance (Mendez 2008; Chang 2020). However, it's worth noting that intensifiers can fulfill various other roles and convey different meanings as well (Szabó and Otani 2022). For instance, they may indicate a positive evaluation by the speaker, as in the Hungarian phrase *brutális alaplap* ('brutal motherboard' meaning 'high-quality motherboard'). They may also convey the speaker's surprise, as in the example *D-23 for prints? That's crazy!* Additionally, they can function as interjections, exemplified by the Japanese term *yabai*. Despite acquiring these new functions, intensifiers may still retain their original meaning. For example, the Hungarian word *durva* ('coarse') can still be used in accordance with its original lexical meaning, as in *durva szemcse* ('coarse grain'). It also has a figurative sense, indicating 'physically or verbally abusive', as in *durva ember* ('rude person') (Szabó and Otani 2022).

NEWs present a formidable challenge from both theoretical and computational linguistic perspectives due to their complex polysemy. Their semantic and pragmatic content, such as whether they denote a positive or negative feature, hinges on the specific context and domain in which they are employed. Then, NEWs provide especially interesting research material not just due to their complex semantic-pragmatic nature, but also because of their linguistic development process. In particular, they proceed through a desemantization as well as a grammaticalization procedure over time. During the temporal desemantization, they gradually lose their lexical meaning (Lorenz 2002; Der 2013), and in a grammaticalization process, they gain an intensifier role.

To gain a deeper understanding of the semantic features of NEWs and their temporal development, it is essential to leverage a vast amount of language data. The corpus used for investigation should also facilitate temporal analysis. Additionally, employing computational analysis methods capable of surpassing manual analysis of limited language data can yield valuable new insights (Hamilton et al. 2016).

We present a case study applying the word2vec embedding method (Mikolov et al., 2013) to analyze the context for target words based on the cosine similarity metric. Distributional semantics provides the tools to uncover semantic features of NEWs and "highlight patterns in semantic change that would not otherwise be noticeable" (Rodda and Davies 2017). By analyzing the change of context for NEWs over time, we aim to track their semantic change over a two-decade period using a large Hungarian corpus of online news texts. Our work is grounded on the premise that a systematic comparison of dense word vectors across different time periods can capture regular processes of linguistic drift (Traugott and Dasher 2001; Xu and Kemp 2015).

From a purely methodological standpoint and within a broader perspective (beyond the examination of NEWs), this study aims to determine whether meaningful results can be obtained using the applied distributional semantics method based on a corpus representing a relatively short period. To our knowledge, a distributional semantic analysis for detecting language modification over time using a corpus representing only a relatively short time period has not been conducted thus far. Additionally, we seek to assess the applicability of this method in the automatic examination of linguistic development in low-resource languages such as Hungarian.

In our analysis, we compare the vector representations of each NEW in the entire corpus, as well as in each time period, with those of other NEWs. To achieve more interpretable results, we will also utilize lexicons containing standard-register intensifiers (non-negative intensifiers, e.g., *nagyon* 'very') (Kochetova et al. 2023), as well as positive and negative sentiment words (Szabó 2015; Ring et al. 2024). This approach enables us to detect whether some of the examined words are semantically closer to negative words (indicating that the prior negative sentiment content of the given word is still active) or standard register intensifiers (indicating that the given words have already undergone lexicalization at least to a certain extent).

The goal of this paper is twofold. Firstly, utilizing specific lexicons and employing a word embedding method, we aim to uncover semantic peculiarities of NEWs in general, as well as specific words compared to their general features. Secondly, we will partition our primary corpus into four sub-corpora, each spanning a 4-year time period. Subsequently, employing the same methods and tools, we will endeavor to track the diachronic semantic changes of these words. Additionally, we will investigate the role of suffixes in the semantic change process of NEWs, specifically whether their presence helps stimulate the semantic development process.

In evaluating our results, we will focus on specific characteristics of NEWs. Specifically, we will conduct separate detailed examinations of some of the most frequent words. Furthermore, we will select NEWs for the current analysis based on our previous research results; that is, we will choose words that have already been analyzed in recent literature using other linguistic analysis methods. In particular, these words include *borzasztó(-an)*, *borzalmas(-an)*, and *durva(-n)*. This approach enables us to compare and evaluate the outcomes of the current analysis with previous findings (Szabó and Bibok 2023; Szabó and Otani 2022; Szabó 2022). Additionally, some of these words can be traced back to the same stem, providing an opportunity to ascertain whether the morphological form has any effect on the semantic development of NEWs.

During the evaluation and discussion of the results, we will also rely on two linguistic hypotheses, namely the *Law of parallel change* and the *Law of differentiation* (Xu and Kemp 2015). These hypotheses argue that words with related meanings tend to change in similar ways over time (parallel change), while near synonyms tend to diverge (differentiation).

The paper is structured as follows. First, we discuss relevant literature on the examination of NEWs, with a focus on papers related to NEWs in the Hungarian language and the utilization of computational semantic methods. Second, we introduce the corpus used for the analysis and describe the methodology employed for corpus processing. Third, we present the analysis results. Fourth, we provide a comprehensive discussion of the analysis results in the context of prior findings on NEWs. Throughout the text, we illustrate certain observations with corpus examples. Lastly, we summarize the key findings of the research and outline our plans for the future.

2. Related Work

More and more authors have been investigating NEWs over the last few decades (e.g., Dragut 2014; Jing and Zhao 2007; Partington 1993; Paradis and Hudson 2000; Paradis 2001; Wierzbicka 2002). Furthermore, a number of studies can be found that examine them in the Hungarian language (Gábor 1988; Nemesi 1998; Péter 1991; Mária 2007; Kugler 2014; Szabó 2019; Szabó and Guba 2022; Szabó et al. 2022).

As mentioned, in the literature NEWs are primarily discussed within the group of intensifiers, with authors focusing on their function as intensifiers. The number of authors focusing on functions of NEWs other than intensification is quite limited (Andor 2011; Szabó 2018; Szabó 2019; Szabó et al. 2022).

As for the semantic-pragmatic development of NEWs, Partington (1993) and Nemesi (1998) scrutinize the delexicalization process of Negative Emotive Intensifiers. Partington (1993) argues that NEWs tend to proceed through a delexicalization process, during which they lose their negative emotive lexical meaning (on the delexicalization process, see, e.g., Bonelli, 2000). These intensifiers also undergo a grammaticalization procedure, during which they acquire "functional (grammatical, pragmatic) meaning components" (Der 2013). Namely, they become intensifiers. Based on Der (2013), the two events, losing the referential meaning and acquiring a new grammatical function, more or less overlap. Furthermore, this semantic developmental process takes place over a longer period of time, in a step-by-step fashion, and as a consequence, not all NEWs are at the same stage of this semantic development in the same time period (Bolinger 2013). This implies that in practice, some intensifiers are more closely related to their initial lexical meaning, while others are further desemantized (Tagliamonte 2005; Wachter 2012).

Despite the fact that more and more studies have recently been investigating NEWs, systematic research aiming to explore trends in the semantic change of these language elements over time is rather scarce. Regarding the analysis of temporal semantic change via computational methods, to the best of our knowledge, only a single paper (Hamilton 2016) has been published on this topic, examining the semantic shift of a NEW (see below).

The application of word embedding methods in computer-assisted automated semantic analyses is now widespread. These methods are based on a distributional semantic approach (Harris 1954), which is grounded on the assumption that the meaning of a word can be captured by its contextual features in a corpus. This suggests that the shift in meaning over time can be traced by comparing the contextual features of the given word in different time periods (Rodda 2017). For instance, changes in the nearest neighbors of a given word may capture even a drastic shift in its core meaning. Some studies focus on testing the explanatory power of this method based on frequency features or tracking syntactic functionality (Mitra et al. 2014; Kulkarni et al. 2015). Wijaya (2011) proposed to identify clusters of topics surrounding the entity over time using Topics-Over-Time (TOT) and k-means clustering.

Some authors have attempted to utilize distributional models to analyze linguistic questions and competing hypotheses about semantic change over time in general (Xu and Reitter 2015; Hamilton et al. 2016a, 2016b). For instance, Xu and Reitter (2015) employed diachronic word embeddings to corroborate the so-called "Law of parallel change" and the "Law of differentiation" concerning the semantic behavior of synonyms over time. The "Law of differentiation" states that near synonyms tend to diverge, while the "Law of parallel change" asserts that words with related meanings tend to change in similar ways over time (Xu and

Reitter 2015). Xu and Reitter (2015) argue that synonyms tend to semantically evolve in parallel rather than follow different routes. Furthermore, Hamilton et al. (2016a, 2016b) used cross-linguistic data to measure changes in pairwise similarities to examine the effect of frequency, as well as polysemy, on the speed of semantic change.

Overall, the results of available research suggest that the application of word embedding algorithms has proven successful in generating high-quality semantic representations of words in large corpora (Desagulier et al. 2019; Amaral et al. 2022). Furthermore, the approach has been found to be a promising tool for diachronic semantics as well (Gulordava and Baroni 2011; Jatowt and Duh 2014; Kulkarni et al. 2015; Xu and Reitter 2015; Hamilton et al. 2016a, 2016b). Since this method allows us to measure how far a word has moved in the semantic space between two time periods (Hamilton et al. 2016a, 2016b), it has the potential to highlight patterns in semantic change that do not rely solely on the intuition of the researcher (Rodda 2017).

As for the NEWs mentioned above, to the best of our knowledge, only a single paper has been published (Hamilton 2016) on this topic that examines the semantic shift of a NEW, namely the word *terrific* in the English language based on an analysis of a large dataset. In order to demonstrate that *terrific* has become more positive over the last 150 years, the authors computed the sentiment values of this word using the SentProp algorithm on historical data.

Our main goal here is to encourage more research like this and provide some useful supplements. We aim to discover whether computational semantic analysis methods can offer new insights into the linguistic nature of the semantic development of NEWs over time and help us detect some regularities in these linguistic changing processes. Furthermore, we would like to evaluate the applicability of this method in the automatic examination of linguistic development in low-resource languages such as Hungarian.

3. Materials and Methods

3.1 The Corpus of the Current Analysis

The corpus used for this study consists of political news articles published between 2002 and 2018 on one of the largest online news sites in Hungary. The main focus of the analyzed content is to present Hungarian as well as international daily news in an easily comprehensible and often entertaining style. The corpus was provided by the Institute for Political Science of the Centre for Social Sciences. The data collection contains a total of 340,149 text samples consisting of 4,562,670 sentences and 106,213,091 tokens altogether.

The raw text of the corpus was stored in plain text using the UTF-8 character encoding format. While the overall quality of the text was generally acceptable, we encountered various issues such as character representation problems, misspellings, and other forms of noise during our analysis. Therefore, prior to the embedding procedure, we conducted several basic cleaning steps. Initially, we removed numerical characters from strings, converting sequences like '1234text1234' to simply 'text'. Subsequently, we removed punctuation characters, numbers, and other non-alphabetical characters. Additionally, tokens containing the HTML tag 'nbsp' were eliminated. These cleaning steps ensured that the corpus was sufficiently prepared for our proposed analyses. Consequently, some data cleaning steps prior to the automatic analysis were necessary.

In the subsequent text processing steps, including sentence splitting, tokenization, lemmatization, and POS-tagging, we utilized the HuSpaCy toolkit (Orosz, 2022), which is the Hungarian model for the SpaCy language processing Python library [1]. Initially, the corpus was segmented into sentences, followed by tokenization and POS-tagging. Additionally, the tokens underwent lemmatization, where they were converted to their base dictionary form. It's worth noting that lemmatization is particularly crucial for morphologically rich languages like Hungarian. Finally, we removed stop words, mainly particles, pronouns, and connectives, as well as punctuation marks.

The list of NEWs was extracted using a simple dictionary-based method applied to the lemmatized version of the corpus. For this step, we utilized a word collection consisting of 225 wordforms obtained in a previous study (Szabó and Guba 2022).

To gain a better understanding of the semantic changes of the NEWs in question, we opted to compare their vector embeddings with those of other groups of Hungarian words, including non-negative intensifiers (cf. *standard register intensifiers*, Kochetova 2022), and positive and negative sentiment words. Our assumption was that if a given NEW is semantically close to several or an increasing number of standard-register intensifiers over time, it suggests that the NEW has progressed further along the path of grammaticalization. Consequently, we might infer that its intensifier function has become more prominent. Additionally, we hypothesized that the degree of semantic bleaching (i.e., *desemantization*) could be deduced from the ratio of negative, neutral, and positive words obtained through our word-embedding method. For example, if a NEW has a significant number of negative closest relations, it is likely that it has not undergone semantic bleaching, or it has done so to a lesser extent, and still retains a prominent negative semantic component in its meaning.

To extract the standard-register intensifiers from the corpus, we utilized a lexicon consisting of 212 words that can function as intensifiers in the Hungarian language (Szabó 2023). For sentiment analysis, we employed two distinct sentiment dictionaries. One of these dictionaries was a sentiment lexicon compiled for general sentiment analysis purposes (Szabó 2015). The other lexicon was domain-specific, specifically designed for the analysis of political texts by the Institute for Political Science of the Centre for Social Sciences within the framework of the "poltextLAB" project (Ring 2024). It is imperative to use a domain-specific sentiment lexicon alongside a general one because, as emphasized by Hamilton et al. (2016) in their related work, "without domain-specific lexicons, analyses can lead to sentiment assignments that are biased towards domain-general contexts".

It is important to note that during the lexicon-based analyses, each element of the NEW lexicon was assigned one primary role exclusively. This means that if a NEW was also present in any other lexicon, such as the negative sentiment lists, it was automatically extracted as a NEW. This exclusivity rule was crucial for the analysis because most NEWs, due to their prior negative semantic content, also appear in negative sentiment lexicons.

Prior to obtaining the embedding vectors, the corpus was divided into four sub-corpora to ensure comparable time periods for the temporal analysis. Since our data collection consists of political news, each time period was determined based on Hungarian parliamentary cycles, resulting in four 4-year periods in our study.

3.2 On the semantic features of the specifically selected NEWs

In the current analysis, we have selected NEWs that have already been analyzed in recent literature using other linguistic analysis methods (Szabó and Bibok 2023; Szabó and Otani 2022; Szabó 2022). In this section, we focus on the selected words and briefly discuss their linguistic characteristics.

The NEW *borzasztóan* (lit. ‘awfully’, morph: ‘awful-ADV’) shares a common stem with the NEW *borzasztó* (lit. ‘awful’). Similarly, the NEW *borzalmasan* (lit. ‘horribly’, morph: horrible-ADV) is related to *borzalmas* (lit. ‘horrible’). These word pairs have two different relative stems: the former pair is associated with the verb *borzaszt* ‘give someone a shudder’, while the latter is linked to the noun *borzalom* ‘horror’ (for further details, see Wikiszotar, https://wikiszotar.hu/ertelmezo-szotar). Despite not sharing the same relative root, they all trace back to the same etymological root (Czuczor and Fogarasi 1874; Szabó 2022).

The adjective *durva* (lit. 'harsh') encompasses several meanings, which are grouped into five main categories by Géza (1959): 1. Having an uneven surface, not smooth, rough, coarse; 2. Crudely crafted, lacking detail; 3. Poorly executed, imperfect, or involving very demanding, hard work; 4. Aggressive, offensive, or rude (referring to a person); 5. Any action performed by a person described in the previous meaning: impolite, unconventional, or inappropriate (referring to words, deeds, behavior). The latter, more abstract meanings are presumed to have evolved from the former, more literal ones, based on their shared unpleasantness. Records indicate its use as an intensifier as early as the 18th century (Gerstner, 2014). The NEW *durván* (lit. 'harshly', morph: harsh-ADV) shares a common stem with the aforementioned NEW *durva* (lit. 'harsh').

Now, let us summarize the previous findings regarding these NEWs. According to several studies, the word *durva* emerges as a significant "polarity shifter" in Hungarian discourse, as indicated by Szabó (2018) and Szabó and Bibok (2023), where it is frequently used as a positively evaluating word, particularly in Hungarian tweets and speech texts. Additionally, Szabó (2018) suggests that *durván* is progressing towards becoming an intensifier, as evidenced by word association test results. Furthermore, corpus analysis results from Szabó and Otani (2022) reveal further differences in the usage patterns of *durva* and *durván*. Additionally, Szabó et al. (2022) highlight the potential influence of suffixes on the distribution of functions and meanings of the above mentioned NEWs, suggesting that the suffix of a NEW may significantly impact its usage.

Regarding *borzasztó(-an)* and *borzalmas(-an)*, both tend to appear in a negative sentiment context, as observed in the Hungarian corpus of product reviews (Szabó 2018).

However, Szabó (2022) highlights significant differences in their collocation features. Specifically, *borzasztóan* occurs more frequently as an intensifier compared to *borzalmasan*, indicating potential divergence in their language development paths despite sharing the same etymological root. Additionally, Szabó et al. (2023) note that *borzasztó* often collocates with context-dependent and positively polarized words, whereas *borzalmas* tends to collocate with words carrying negative semantic content.

3.3 Building the Distributional Vector Space

Distributional embedding models like the word2vec method (Mikolov et al. 2013) are widely employed techniques for text representation, where a real-valued vector is calculated for each unique word of a corpus to capture its meaning based on its context. In these models, target words are represented as dense vectors in a high-dimensional space, and each dimension records the contextual statistic features of the target words. By virtue of their representation with these vectors, words may be encoded as points in a semantic vector space. Furthermore, semantic similarity of the analyzed words can easily be measured by using vector similarity or distance, where cosine similarity is a widely used approach (Turney and Pantel 2010; Rodda and Pulman 2017). The use of cosine similarity-based measurements for word2vec embedding vectors to explore dynamic changes in the meanings of words was justified in our recent study (Szabó 2021).

The word2vec embedding method (Mikolov et al. 2013) was utilized in this study using the Gensim package (Rehurek and Sojka 2010) in Python. We trained a single embedding model for the cleaned and preprocessed four sub-corpora to ensure direct comparability of the vectors of target words. After the aforementioned text processing steps, we assigned a string tag to each target NEW word to indicate its corresponding time period. This tag was formed by concatenating the word with the string "_n", where "n" denotes the time period. For instance, the target word *borzasztó* would be represented as "borzasztó_0" in the first time period, "borzasztó_1" in the second, and so forth. Subsequently, the four sub-corpora were merged into a single corpus, and a word2vec model was trained. This approach ensures the definition of a unified vector space across all periods, ensuring that target words from different time periods remain distinguishable and each has a unique embedding vector.

During the training of the word2vec model, a context window of size 5 was utilized to calculate the embeddings, encompassing 5 words preceding and 5 words succeeding the target word. The continuous bag-of-words (CBOW) training algorithm was employed, yielding a 100-dimensional embedding space. Additionally, a minimum word count of 5 was imposed on the model, ensuring that words with fewer occurrences were excluded. Notably, we opted to preserve the original morphology of the corpus, foregoing lemmatization during the embedding model training process. As a result, the model encompasses a vocabulary comprising 403,589 tokens, each represented by a 100-dimensional embedding vector.

Our analysis primarily relies on the nearest neighbor analysis of selected target words. This approach entails examining how the neighboring words of a tracked target word change across different time periods, thereby indicating shifts in meaning. In essence, we operate under the assumption that alterations in meaning are reflected by changes in the contextual usage of a target word over fixed periods (Kutuzov et al. 2018).

To measure the distance between vector pairs, we employed the cosine similarity measure. This metric quantifies the similarity between two words, A and B, represented by vectors wAw_AwA and wBw_BwB with elements wA(i)w_A(i)wA(i) and wB(i)w_B(i)wB (i), respectively, through the dot product:

$$S = \frac{w_A \cdot w_B}{w_A w_B}$$

$$= \frac{\sum_i w_A(i) w_B(i)}{\sqrt{\sum_i w_A(i)^2}\sqrt{\sum_i w_B(i)^2}}$$

To track the semantic shifts of the NEWs, we systematically extracted the ten most similar words for each target word in all four time periods. Additionally, we categorized these similar words into NEWs, non-negative intensifiers, positive or negative sentiment words.

4. Results

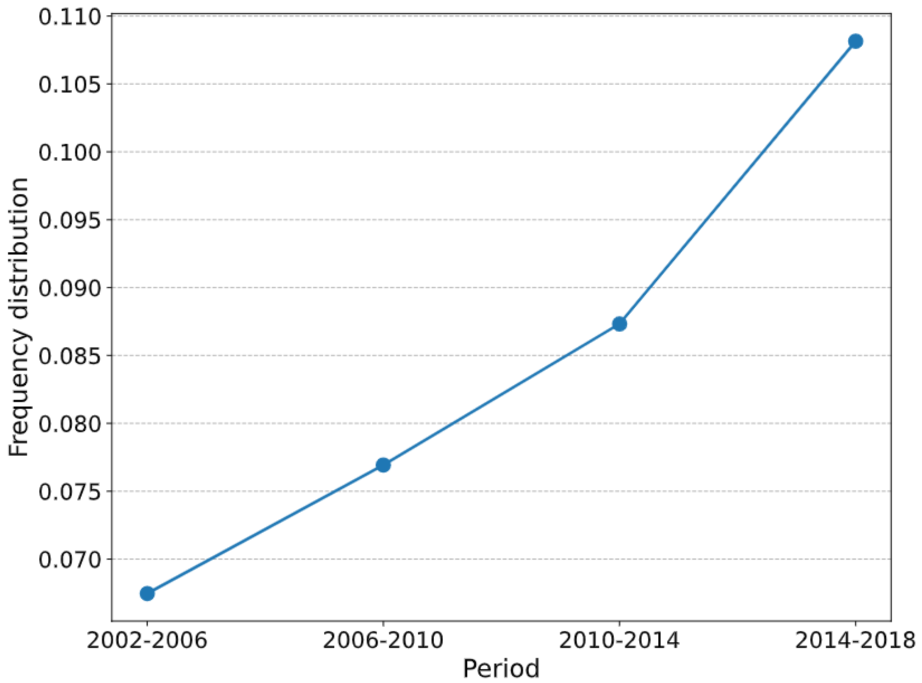
Our corpus comprises a total of 340,149 text samples, containing 106,213,091 tokens in total. Utilizing lexicon-based corpus processing, we identified a total of 51,049 instances of NEWs within the dataset. The basic statistical data of the full corpus and the four sub-corpora are presented in Table 1.

Table 1: Basic statistical data of the corpus divided into four sub-periods.

Time period	Texts	Sentences	Tokens	NEWs
2002-2006	71,426	745,648	17,300,274	6,661
2006-2010	79,778	1,055,192	24,564,024	10,489
2010-2014	98,840	1,384,131	32,201,222	15,373
2014-2018	90,105	1,377,699	32,147,571	18,526
Full corpus	340,149	4,562,670	106,213,091	51,049

As for the overall frequency distribution of NEWs across the four time periods, Figure 1 presents the basic results. In this representation, the number of occurrences is normalized by the total number of tokens in each period.

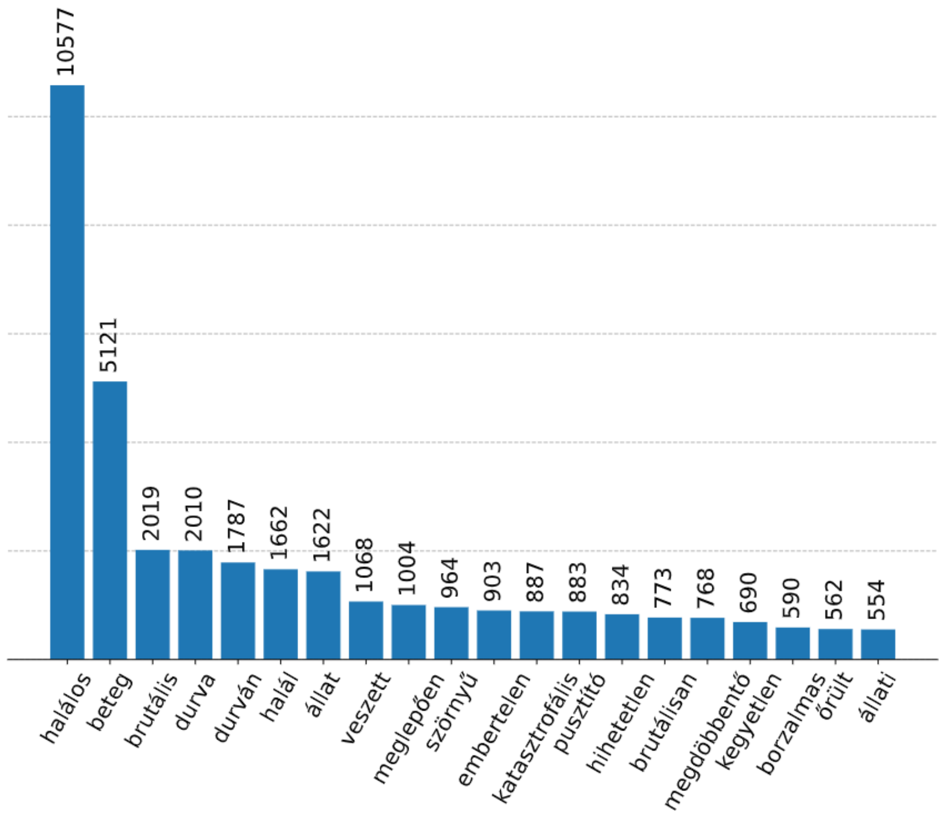
Figure 1: The overall frequency distribution of NEWs over the four time periods.



The frequency of NEWs exhibits a consistent upward trend across the four periods.

Out of the 225 NEWs listed in the NEW lexicon, 196 were identified in our corpus. However, to enhance interpretability, we excluded those NEWs that were entirely absent in any of the time periods. Consequently, a total of 103 NEWs remained for analysis. Figure 2 illustrates the twenty most frequent NEWs in the entire dataset, along with their respective frequencies.

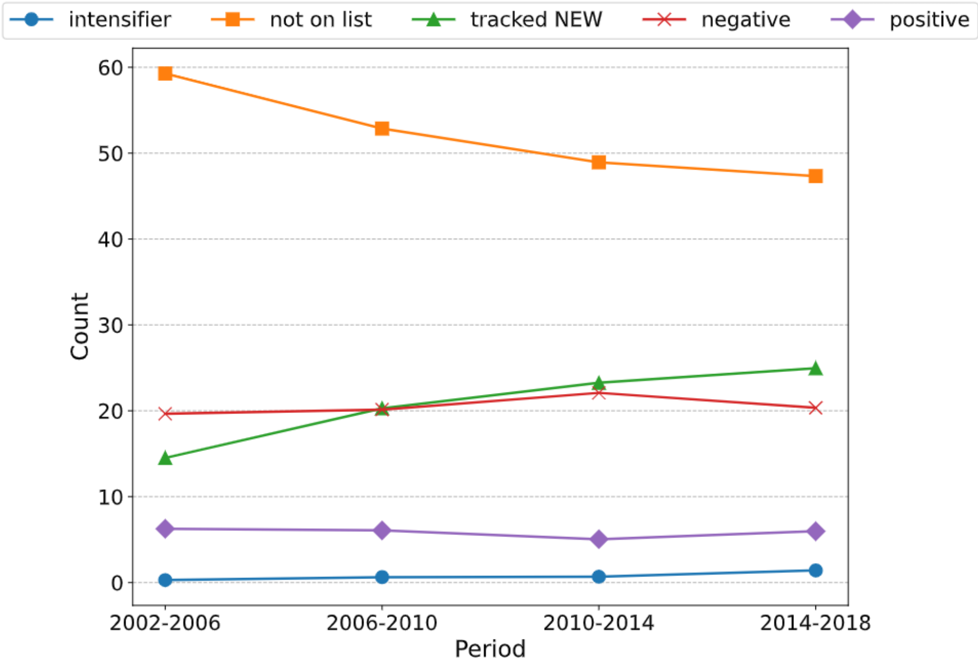
Figure 2: The twenty most frequent NEWs, along with their overall number of occurrence.



As can be seen, there are two words, *halálos* ('deadly') and *beteg* ('sick'), that are much more frequent in the corpus in comparison to the rest of the words, while the frequency of the remaining words is more evenly distributed.

Now, let us examine the semantic characteristics of the NEWs in our corpus by analyzing their closest semantic neighbors. As a first step, let us review the general characteristics of the NEWs in this context. Figure 3 shows the aggregated data regarding the semantically closest words to the NEWs.

Figure 3: The aggregated data of the semantically closest words to the NEWs over the four time periods.



Based on the observations from Figure 3, it is evident that the semantic proximity among the NEWs gradually increases across the examined time periods, with a slight convergence towards standard register intensifiers in the final period. At the same time, the frequency of words that could not be classified into any of the examined categories shows a consistent decrease over time.

Now, let us delve into the data concerning the most frequent NEWs in the corpus. Once again, the word *halálos* ('deadly') stands out as the most common word, with its root *halál* ('death') also ranking prominently as the 6th most frequent word. However, despite their high frequency, these words may not provide significant insights into the linguistic peculiarities of the corpus due to its domain characteristics; primarily, these words appear in their literal sense within the context of the news texts. For example, the closest words to *halál* ('death') include terms such as *akasztás* ('hanging') or *áramütés* ('electric shock'), while those closest to *halálos* ('deadly') encompass *áldozat* ('victim') or *sérült* ('injured'). Similar observations hold for the words *beteg* ('sick') and *brutális* ('brutal'), as well as their derived forms.

Two of the NEWs we selected, namely, *durva* and *durván*, proved to be quite common in our corpus, ranking as the 4th and 5th most frequent NEWs in the dataset, respectively. We will delve into the analysis of these words shortly. However, before doing so, let us examine the frequency of the selected NEWs in each time period, as presented in Table 2 below.

Table 2: The number of occurrences of the NEWs in the four time periods.

	2002-2006	**2006-2010**	**2010-2014**	**2014-2018**	**ALL**
durva	215	391	595	809	2010
durván	127	392	619	1787	2925
borzasztó	44	109	173	216	542
borzasztóan	22	43	66	92	223
borzalmas	69	90	158	273	590
borzalmasan	8	16	21	36	81

The data shows changes in the frequency of individual NEWs over time periods. For instance, the frequency of *durva* and *durván* words increased over time, while other words such as *borzasztóan* and *borzalmasan* are less frequent and distributed more evenly across the periods. The table allows us to compare the frequency of each NEW over time and observe any trends or changes between periods.

Let us commence with an examination of the most prevalent word pair, consisting of the NEWs *durva* and *durván*. The outcomes of this analysis are delineated in Figures 4 and 5 below.

Figure 4: The twenty most semantically similar words to the word durva.

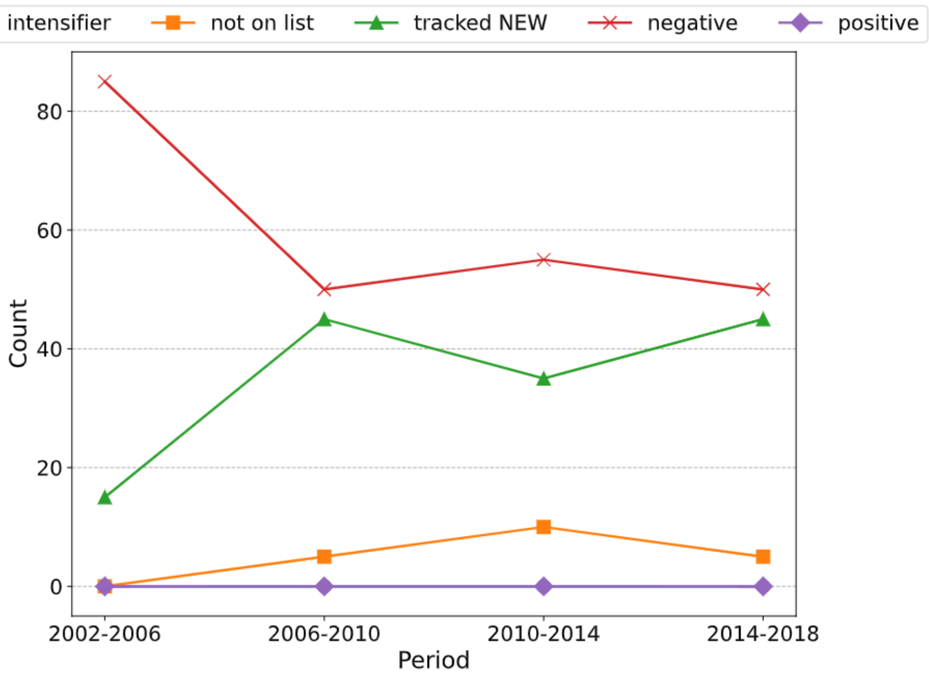
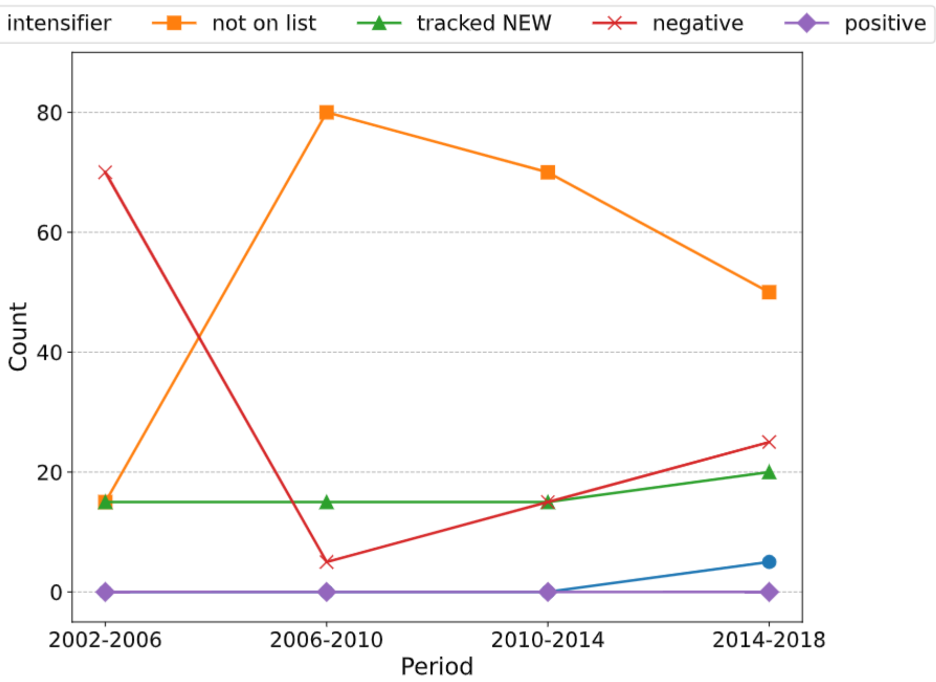


Figure 5: The twenty most semantically similar words to the word durván.



Based on these results, we observe some noteworthy differences. It is evident from the two figures that these two NEWs exhibit distinct semantic development characteristics. The most significant disparities lie in the fact that for *durván*, positive words and standard-register intensifiers emerge in the last period.

To gain a better understanding of the results, we also manually examined the twenty most similar words. They are presented in Tables 3 and 4. In these tables, color codes are utilized to indicate whether a given word belongs to one of the four defined categories (NEWs, positive and negative semantic words, and standard-register intensifiers). Please refer to Table 3 for the color definitions.

Table 3: Color codes of the four defined categories.

Color code	**Categories**
grey1	negative sentiment word
grey2	NEW
grey3	intensifier
grey4	positive sentiment word

Table 4: The evolution of similar words for durván *(lit. 'harshly') across four time periods*

2002-2006	**2006-2010**	**2010-2014**	**2014-2018**
durva_1	durva_0	durva_3	durva_2
otromba	durva_2	durva_1	durva_1
cinikus	durva_3	durva_0	durva_0
durva_2	otromba	cinikus	brutális_2
bántó	brutális_0	otromba	brutális_1
durva_3	brutális_2	brutális_2	brutális_3
kétértelmű	brutális_1	brutális_1	eltúlzott

álságos	kegyetlen_1	csúnya	otromba
hazug	cinikus	sokakban	cinikus
ízléstelen	kegyetlen_2	bántó	kirívó
demagóg	provokatív	kártékony	ijesztő_1
arcátlan	verbális	oktalan	abszurd
szexista	bántó	verbális	megalázó
provokatív	megalázó	ijesztő_3	kegyetlen_3
felháborító	agresszív	öncélú	megdöbbentő_3
álszent	kegyetlen_3	ostoba	bántó
szánalmas	oktalan	ízléstelen	meggondolatlan
megalázó	szexista	primitív	sokakban
megengedhetetlen	meggondolatlan	meggondolatlan	rosszindulatú
öncélú	rosszindulatú	brutális_3	agresszív

Table 5: The evolution of similar words for durva *(lit. 'harsh') across four time periods*

2002-2006	**2006-2010**	**2010-2014**	**2014-2018**
durván_3	durván_2	durván_1	durván_2
durván_2	durván_3	durván_3	durván_1
durván_1	durván_0	nominálisan	durván_0
dobálózó	végösszeg	reálértéken	csúnyán
százalékukat	tíz-	durván_0	valósággal
bántalmaz	rugó	differencia	miközben
fenyegetve	differencia	értékvesztés	eléggé

igazságtalanul	rúgó	árfolyamváltozás	akkorára
lealacsonyító	büdzsén	ft-ra	óriásit
melegeket	forintom	nagyjából	nemhogy
taszigálták	kapásból	összegszerűen	ugyanannyival
buzikat	toldja	elszaladása	veri
tettlegesen	ft-ra	áfateher	ennyivel
bélyegzi	kamatként	százmillióra	nullára
arcul	nominálisan	törlesztőrészletük	leértékelődjön
arrogánsan	áfából	kamatteher	hatalmasat
románokat	tízmilliárdot	reálértelemben	rettenetesen_3
deprivált	öngólt	majdnem	öngólt
szidta	összegszerűen	inflációval	búzaár
homoszexuálisokat	százmillióra	száltra	felelőtlenül

When comparing the data from these two tables, the semantic distinctions between the two words become evident, as well as their divergent evolution over time. In the case of *durva*, the semantically closest words show no significant changes across the examined time periods. However, notable changes are observed for *durván*. For instance, in the first period, *durván* has more negative semantic neighbors, but from the second period onward, the number of such words notably decreases. Additionally, *durván* acquires a sense of 'approximately' starting from the second time period, as evidenced by words like *majdnem* ('almost') and *nagyjából* ('roughly'), along with words denoting numerical values and other terms semantically linked to the domain of economics (Vincent, 1982).

Data from the fourth period also reveals that *durván* strengthens its function as an intensifier, as it consistently includes two standard-register intensifiers among its twenty nearest semantic neighbors (Weinreich et al., 1968).

Moving forward, let us provide an overview of the results pertaining to the members of the "borz" word group. We will commence with the word pair *borzalmas* and *borzalmasan*.

Figure 6: The twenty semantically closest neighbors of the word borzalmas *(lit. 'horrible').*

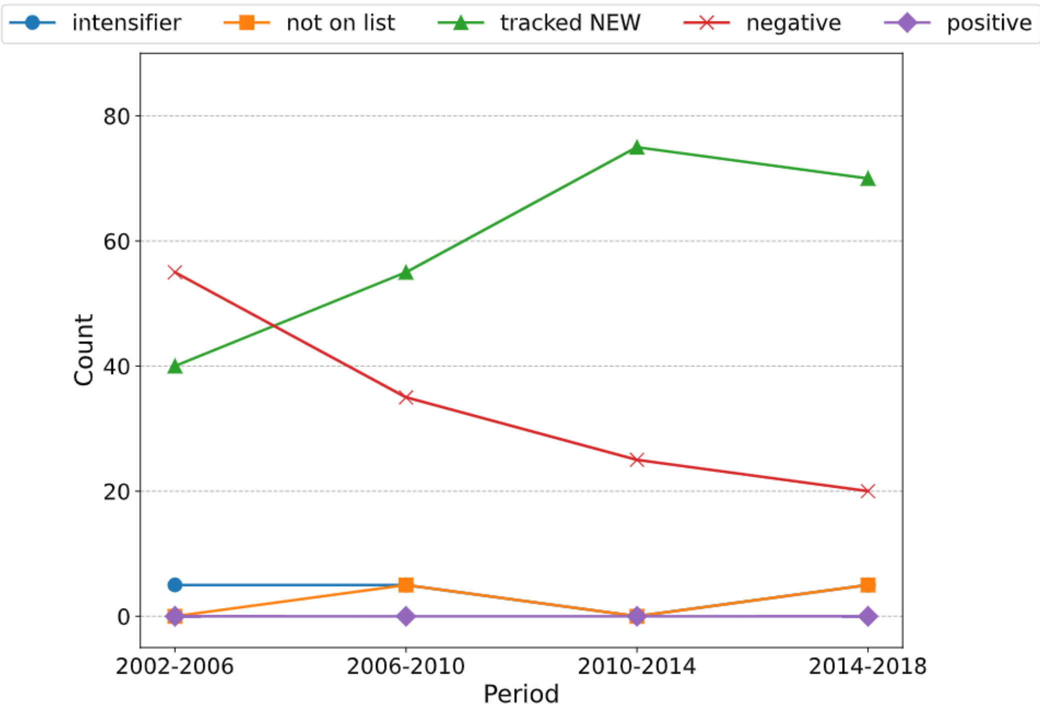
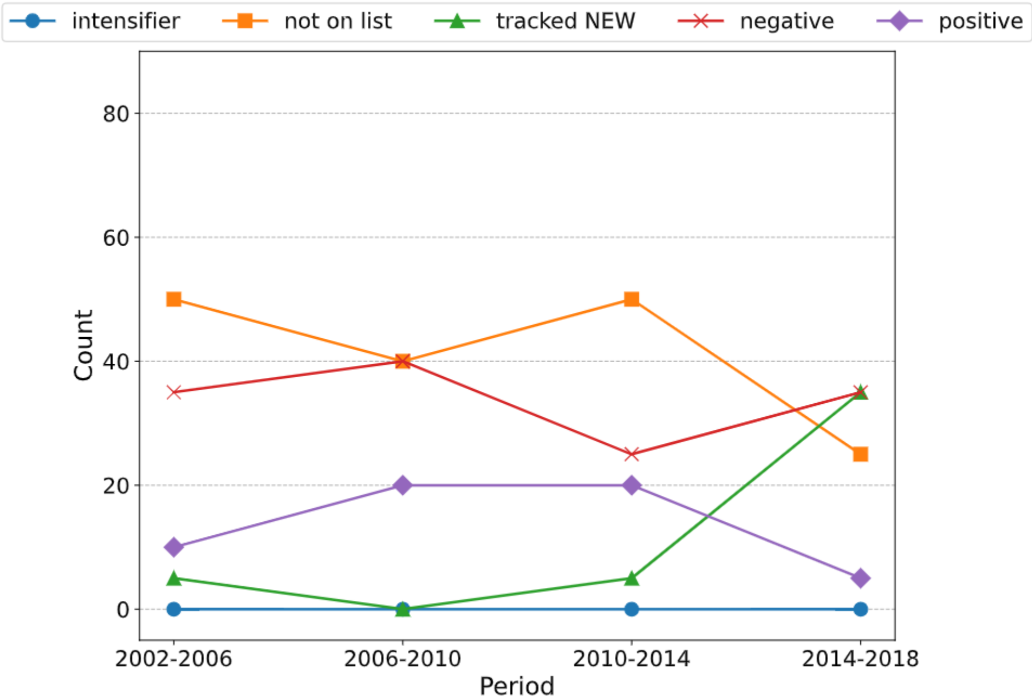


Figure 7: The twenty semantically closest neighbors of the word borzalmasan (lit. 'horribly').



When comparing the numerical data from these figures, a striking difference emerges. In the case of *borzalmasan*, positive words appear relatively frequently among the twenty nearest neighbors, with the most common words falling into the „no list" category. Conversely, for *borzalmas*, no positive words appear in the data, the proportion of „no list" words is much smaller, and NEWs appear much more frequently among the twenty nearest neighbors. Additionally, the proportion of negative words is higher for *borzalmas*. These observations suggest that *borzalmasan* is more commonly used in a non-negative context, while *borzalmas* exhibits stronger semantic connections with other NEWs. However, it is intriguing to note that the proportion of negative words decreases over time for both words, and the prevalence of NEWs notably increases for *borzalmasan* in the last period.

Let us now delve into the detailed results provided in Tables 6 and 7 below.

Table 6: The evolution of similar words for borzalmas *(lit. 'horrible') across four time periods.*

2002-2006	2006-2010	2010-2014	2014-2018
meggyötört	rettenetes_3	borzalmas_3	szörnyű_3
leírhatatlan	borzalmas_2	borzasztó_3	szörnyű_2
elvetemült	borzalmas_3	borzasztó_2	borzalmas_2
szörnyű_3	borzasztó_0	szörnyű_2	szörnyű_1
átéltek	ocsmány	szörnyű_3	rettenetes_3
rémisztő_1	leírhatatlan	rettenetes_3	borzasztó_3
borzalmas_3	rettenetes_2	rettenetes_2	rettenetes_2
szörnyű_1	rémisztő_1	borzalmas_1	szörnyű_0
horrorisztikus	borzasztó_2	borzasztó_0	borzalmas_1
borzasztó_0	borzasztó_3	borzasztó_1	borzasztó_2
riadtan	szuicid	rettenetes_1	horrorisztikus
szörnyű_2	eltorzult	ocsmány	borzalmas_0
rettenetes_2	öntudatlan	szörnyű_1	rettenetes_1
fájdalomtól	gyötri	leírhatatlan	borzasztó_0
mártírként	szörnyű_3	szerencsétlen	tragikus
szörnyű_0	fogai	undorító	leírhatatlan
lövöldözős	legyengült	szörnyű_0	átélt
émelyítő	szörnyű_2	stresszes	gyomorforgató
szuicid	pszichotikus	pokoli_3	{mentálisa
zavarodott	borzalmas_0	szürreális	kegyetlen_3

Table 7. The evolution of similar words for borzalmasan *(lit. 'horribly') across four time periods.*

2002-2006	**2006-2010**	**2010-2014**	**2014-2018**
zavaromban	kinevetni	falumban	borzasztóan_3
ismeretségünk	diogenidész	gyógyulásról	borzasztóan_1
szerethetőbb	túlélésben	irtózatosan_2	higgyétek
válaszol	szolidaritásával	primán	örüljetek
rémesen_0	bármin	nyálú	borzasztóan_2
emlékezet	őszintesége	disznóságot	iszonyatosan_1
csatakiáltás	zavarával	fényképeiket	szörnyen_3
ólommal	lábszagú	édesanyámon	üldözési
métely	nyomorultul	minnesotát	okoskodni
publicisztikákban	hazugsággyár	vidámsággal	irtózatosan_2
megettük	világunkat	mosolyodott	szörnyen_2
eltévedek	felfuvalkodott	többieken	bőrömön
monomániás	rejtező	kályhába	ízük
elfelejtem	vigyázzak	kardélre	idegileg
tálcák	versenyhelyzettel	fültövön	pénzemen
érezted	felkészületlenség	enervált	felpuffadt
gépies	ötlettelen	látványról	ijedve
hiábavalóan	kérdés	kiskamaszként	jézusom
életemről	sztorikra	szegényen	fiamra
kivonható	állóvízben	csetepatéra	bűnösen

From the data, it's evident that in the case of *borzalmasan*, positive words like *őszinteség* ('honesty'), *szolidaritás* ('solidarity'), or *vidámság* ('joy') appear prominently in the results.

Conversely, as previously noted, such positive words are absent among the 20 nearest neighbors for *borzalmas*. Additionally, for *borzalmasan*, if the nearest words are NEWs, they often carry a determinative suffix *-an*, such as *borzasztóan* ('awfully'), *irtózatosan* ('terribly'), and *szörnyen* ('terribly'). This contrast with *borzalmas*, where the nearest NEWs typically lack a suffixed form, appearing instead as non-suffixed adverbs like *rettenetes* ('horrible') and *szörnyű* ('terrible'). Moreover, *borzalmas* exhibits a higher frequency of negative words among its nearest semantic neighbors, which are often semantically linked to the root of the word, *borzalom* ('horror'), including terms like *ocsmány* ('groaty'), *undorító* ('disgusting'), *stresszes* ('stressed'), and *horrorisztikus* ('horrific').

Let us now proceed with our comparison of the word pair *borzasztó* and *borzasztóan*.

Figure 8: The 20 semantically closest neighbors of the word borzasztó *(lit. 'awful').*

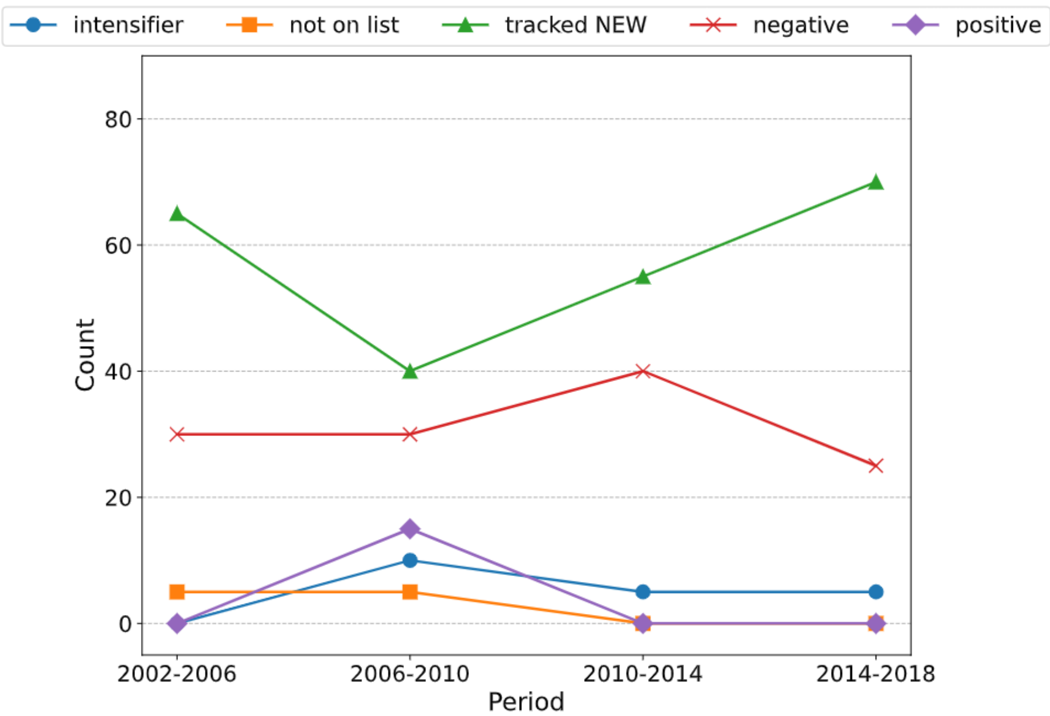
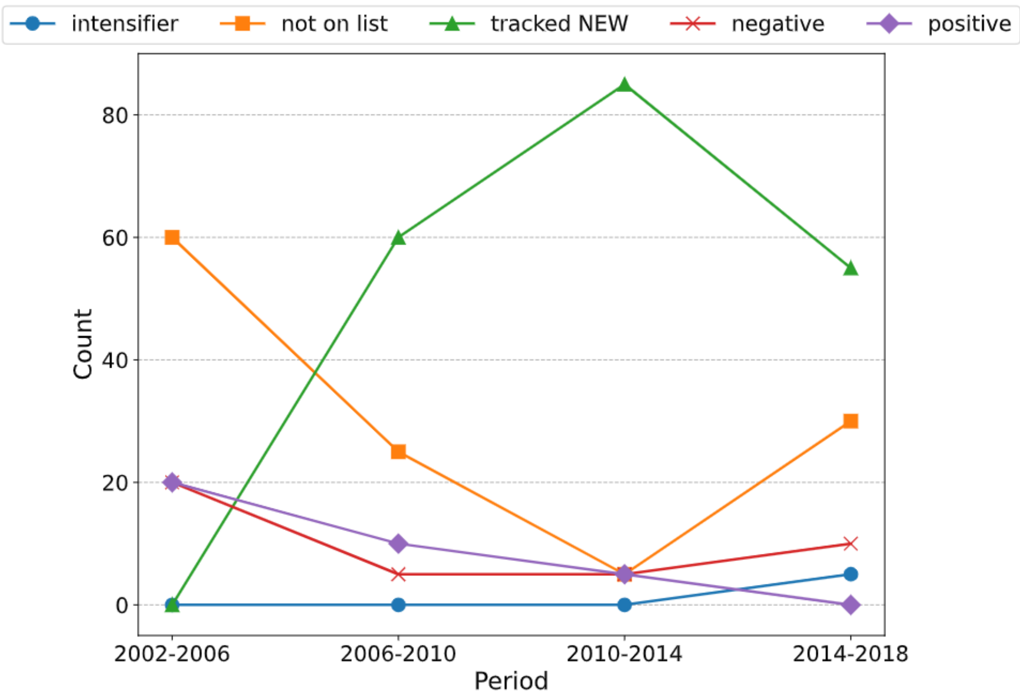


Figure 9: The 20 semantically closest neighbors of the word borzasztóan *(lit. 'awfully').*



Perhaps the most striking difference in the figures is that in the case of *borzasztó*, there are notably more negative words in all periods compared to *borzasztóan*. Also, it is worth noting that both words have positive words among their semantically nearest neighbors, which, as we have seen, was not seen in the case of *borzalmas*.

Now, we present the twenty nearest words in both cases. See tables 8 and 9.

Table 8. The evolution of similar words for borzasztó *(lit. 'awful') across four time periods.*

2002-2006	**2006-2010**	**2010-2014**	**2014-2018**
leírhatatlan	borzasztó_2	borzasztó_3	borzasztó_2
gyomorforgató	borzasztó_3	borzasztó_1	borzalmas_2
borzasztó_2	borzasztó_0	borzalmas_2	borzasztó_1
borzasztó_1	idegesítő	borzasztó_0	borzalmas_3
undorító	borzasztóan_1	csúnya	stresszes
rettenetes_3	borzasztóan_2	rettenetes_3	rettenetes_3

borzalmas_1	szar	szerencsétlen	borzasztó_0
rémisztő_1	rettenetes_1	leírhatatlan	szörnyű_3
félelmetes_3	leírhatatlan	borzalmas_3	pokoli_3
frusztráló	borzalmas_2	pokoli_3	rettenetes_2
borzalmas_2	éreztük	stresszes	szörnyű_2
borzasztó_3	totál	szörnyű_2	szerencsétlen
ocsmány	fárasztó	rettenetes_2	lehangoló
szörnyű_2	frusztráló	gyomorforgató	kényelmetlen
elvetemült	félelmetes_2	undorító	leírhatatlan
idegesítő	csúnya	szürreális	borzalmas_1
félelmetes_2	bosszantó	szörnyű_3	frusztráló
rémisztő_2	bizsergető	hihetetlen_3	ijesztő_2
ijesztő_0	win-win	ocsmány	borzasztóan_1
rettenetes_0	jópofa	frusztráló	ijesztő_3

Table 9. The evolution of similar words for borzasztóan *(lit. 'awfully') across four time periods.*

2002-2006	**2006-2010**	**2010-2014**	**2014-2018**
összezavarodik	borzasztó_1	borzasztó_1	borzasztóan_2
ijedős	borzasztóan_2	borzasztóan_1	borzalmasan_3
ifjonti	rettenetesen_2	borzasztóan_3	rettenetesen_2
elhiggye	irtózatosan_2	találékony	borzasztóan_1
végzed	borzalmasan_3	baromi_3	rettenetesen_3
rágja	szentimentális	iszonyatosan_3	lelkileg

melegszívű	borzasztóan_3	rettenetesen_2	bántóan
elmúltnyolcévben	rettenetesen_0	baromi_2	érzelmileg
felkent	rettenetesen_3	iszonyatosan_2	baromi_3
aha	érzésre	mértem	rettentően_2
munkál	viselkedéséből	rettenetesen_3	rettenetesen_1
csatakiáltás	bőrömön	hihetetlenül_0	rettentő_2
tudatosuljon	borzasztó_3	iszonyú_3	elhiszem
felkiált	bármennyire	szörnyen_3	irtózatosan_2
fintorogva	szörnyen_3	frusztráló	hihetetlenül_3
jó-jó	borzasztó_2	félelmetes_2	lelkű
életszemléletét	bántóan	borzalmasan_3	végtelenül
tetszhet	mily	hihetetlenül_3	idegileg
önelégült	rettenetesen_1	borzasztó_3	frusztráló
letakarni	önzetlennek	borzasztó_2	bármennyire

Upon closer examination of the twenty nearest words individually, we observe that in the case of *borzasztóan*, there are several positive words such as *önzetlen* ('selfless') or *melegszívű* ('warm-hearted') (Vincent 1982). Additionally, the semantic content of the nearest neighbors extends beyond their sentiment value. Many of these neighbors refer to emotions and mental states, as exemplified by *lelkileg* ('mentally'), *idegileg* ('neurally'), and *érzelmileg* ('emotionally') (Vincent 1982).

The semantically closest words categorized as NEWs, similar to observations in the case of *borzalmasan*, predominantly appear as suffixed forms, such as *rettenetesen* ('horribly') or *iszonyatosan* ('terribly') (Weinreich, Labov and Herzog, 1968). However, in terms of suffixes, there are exceptions, as evidenced by words like *félelmetes* ('scary') or *baromi* ('bestial'), which lack suffixes yet remain among the nearest neighbors (Weinreich et al., 1968). In spite of the aforementioned, the nearest words for *borzasztó* are closely related to the negative emotive semantic content of the examined word, often associated with emotions like fear or disgust, such as *frusztráló* ('frustrating'), *gyomorforgató* ('disgusting'), *szar* ('crap'), or *undorító* ('disgusting') (Lépez Morales 1981; Vincent 1981, 1982).

5. Discussion of Results

In the following section, we will discuss our research findings and compare them with the results obtained in previous studies.

Our manual analysis of the semantically closest words to *durva* and *durván* revealed intriguing differences, suggesting distinct semantic development characteristics between these two NEWs. Particularly noteworthy is the emergence of positive words and standard-register intensifiers in the last period of our analysis in the case of *durván*. This observation aligns with previous arguments regarding desemantization and lexicalization discussed in Section 2. As noted in Vincent (1982), *durván* has exhibited signs of intensifier evolution. What is more, the suffix of a given NEW (as in the case of *durván*) may significantly influence the distribution of its various functions and meanings, as emphasized in Weinreich et al. (1968). Our findings corroborate these prior insights.

The manual examination of the 20 most similar words revealed consistent semantic patterns for *durva* across the examined time periods. In contrast, notable changes were observed for *durván*, indicating its evolving semantic characteristics over time. Specifically, the frequency of semantically closest words with negative connotations decreased significantly from the second period onwards, suggesting a shift in usage. Moreover, in the fourth period, *durván* demonstrated a strengthening of its intensifying meaning, as evidenced by the presence of standard-register intensifiers among its 20 nearest semantic neighbors.

We now juxtapose our recent research findings concerning the words of the "borz" group, namely: *borzalmas*, *borzalmasan*, *borzasztó*, and *borzasztóan*, with previous research results.

Prior research, as described in Vincent (1982), suggested that both *borzasztó(-an)* and *borzalmas(-an)* were commonly used in product reviews, with no apparent distinction noted between these two NEWs in this context. However, recent findings presented in Szabó (2022) and Szabó et al. (2023) have brought forth new insights into their semantic evolution. Firstly, in the analysis of the word pair *borzasztóan* and *borzalmasan*, the former appears more frequently as an intensifier compared to the latter. Secondly, in the examination of the word pair *borzasztó* and *borzalmas*, *borzasztó* is found to co-occur with context-dependent and positively polarized words in their analyzed corpus, whereas *borzalmas* tends to associate with words bearing negative semantic connotations.

Upon analyzing the word pair *borzalmas* and *borzalmasan*, it became evident that *borzalmasan* is frequently accompanied by positive words among its 20 nearest neighbors, whereas *borzalmas* lacks any positive associations and exhibits a higher proportion of negative words. This observation underscores the role of affixes in the semantic evolution process. Furthermore, there were notably more negative words among the nearest neighbors of *borzasztó* across all periods compared to *borzasztóan*, suggesting that *borzalmasan* is more commonly utilized in a non-negative context than *borzalmas*. Additionally, it was observed that the proportion of negative words decreases over time for both words, indicating a gradual departure from their initial negative semantic connotations as they evolve. Moreover, the proportion of NEWs increased significantly for *borzalmasan* in the last period.

Regarding the word pair *borzasztó* and *borzasztóan*, notable distinctions are observed. Throughout each period, *borzasztó* exhibits a higher frequency of negative nearest neighbors compared to *borzasztóan*. This disparity underscores the influence of affixes in shaping semantic evolution. Furthermore, both *borzasztó* and *borzasztóan* display positive words among their closest neighbors, a feature absent in the case of *borzalmas*. Recall that the 20 nearest words for *borzalmas* predominantly consist of negative terms associated with emotions such as fear and disgust. This discrepancy aligns with prior research findings, which highlighted that *borzasztó* tends to co-occur with context-dependent and positively polarized words, while *borzalmas* exhibits a propensity for collocating with terms carrying negative semantic connotations (Szabó et al. 2023).

In the case of *borzasztó*, the presence of positive words is notable among its 20 nearest neighbors, alongside frequent occurrences of terms related to emotions and mental states. Conversely, for *borzasztóan*, the presence of NEWs among its nearest neighbors, akin to *borzalmasan*, predominantly comprises suffixed adverbials. However, exceptions exist, as observed with words like *félelmetes* ('scary') and *baromi* ('bestial'), suggesting complex correlations that warrant further analysis. Additionally, the presence of standard intensifiers may indicate an advanced degree of lexicalization. These findings underscore the divergent trajectories in the development of the two examined word pairs.

In summary, our new research findings provide fresh insights into the evolving semantic characteristics of Hungarian NEWs. As for *durva* and *durván*, while previous research highlighted their roles as polarity shifters and intensifiers, our analysis reveals nuanced changes over time, particularly for *durván*. These findings offer valuable insights into the dynamic nature of word meanings and their usage in the Hungarian language. Regarding the "borz" word group, our results align with previous research, confirming that shared etymological roots do not necessarily lead to similar language development paths. Moreover, they underscore the subtle differences between previously considered synonymous words, such as *borzalmasan* and *borzasztóan*, elucidating their unique semantic trajectories.

Our results merit examination from the perspective of two linguistic assertions that offer contrasting predictions about semantic change over time: the *Law of differentiation* and the *Law of parallel change* (Xu 2015). As discussed in Section 2, the *Law of differentiation* posits that near synonyms tend to diverge, while the *Law of parallel change* suggests that words with related meanings evolve in similar ways over time. NEWs exhibit linguistic behavior that aligns with the *Law of parallel change* to some extent. Initially, they carry negative semantic content and undergo a grammaticalization process to acquire an intensifier function (Heine and Kuteva 2013). According to Szabó et al. (2022), the presence of a DEGREE semantic component is pivotal in interpreting NEWs as intensifiers in certain constructions, suggesting that this component exists from the early stages of their semantic development. Consequently, NEWs evolve in parallel over time as they transition into intensifiers, thereby conforming to the *Law of parallel change*.

However, our analysis also reveals instances where different NEWs diverge over time. For example, within the "borz" group, *borzalmas* and *borzalmasan* tend to have more negative nearest semantic neighbors compared to *borzasztó* and *borzasztóan*, which appear to be more neutral. These findings suggest that both the *Law of differentiation* and the *Law of parallel change* play significant roles in the semantic development of NEWs over time. Depending on the context and linguistic environment, one of these laws may prevail during a given period. Thus, the semantic evolution of NEWs serves as a compelling illustration of how two seemingly contradictory linguistic laws can coexist and operate simultaneously during a particular period, depending on the perspective taken.

6. Conclusion and Future Work

This study employed a word embedding methodology to examine the semantic nuances of Negative Emotive Words (NEWs) in the Hungarian language. The objectives of this paper were twofold: to unveil the semantic features of NEWs and trace the temporal shifts in their semantics across different time periods. More precisely, by utilizing specific lexicons and employing word embedding techniques, we sought to unveil the semantic features of NEWs in both a general sense and within specific word categories, juxtaposing them with general language features. Then, we endeavored to trace the temporal shifts in these semantic features of these words across different time periods.

Our investigation into the semantic shifts of Hungarian NEWs, focusing on the word pairs *durva* and *durván*, as well as *borzalmas(-an)* and *borzasztó(an)*, has yielded significant findings. By comparing our results with previous research, we have uncovered noteworthy distinctions and gained valuable insights into the semantic-pragmatic development of these words over time. For instance, we observed a notable transition in the usage of *durván* as it gradually evolved into an intensifier, characterized by the emergence of positive words and standard-register intensifiers in the later period of our analysis. Our study reinforces the concepts of desemantization and lexicalization, providing empirical evidence for the gradual loss of negative sentiment associated with *durván*.

Additionally, the comparison between *borzasztó(-an)* and *borzalmas(-an)* underscored the varying rates of semantic changes in these words, highlighting that etymological similarity does not necessarily guarantee parallel linguistic development paths. This observation further emphasizes the importance of considering each word's unique semantic trajectory when analyzing its evolution. Moreover, it emphasizes the more diverse usage of the NEW *borzasztó*, indicating that it may exhibit a wider range of semantic nuances compared to *borzalmas*. The role of suffixes was also noted in several instances throughout this research project, indicating their influence on the semantic evolution of the studied words. This suggests that morphological peculiarities play a crucial role in shaping the semantic development of NEWs in Hungarian.

The discussion of our results was informed by linguistic hypotheses, specifically the *Law of differentiation* and the *Law of parallel change*. We propose that the semantic evolution of NEWs over time is significantly influenced by the interplay of these linguistic laws. Essentially, one of these laws tends to prevail during a particular phase of linguistic development. Consequently, the temporal trajectory of NEWs' semantics provides a compelling illustration of how these seemingly contradictory laws can coexist within the same timeframe, depending on the perspective adopted.

The insights gleaned from this study pave the way for several promising avenues of future research. Firstly, expanding the temporal coverage could provide a more comprehensive understanding of the semantic evolution of NEWs. A broader dataset may capture subtle changes and linguistic nuances that were not apparent within the scope of this study. Secondly, exploring alternative methods for contextual embeddings, such as leveraging the BERT language model, could enhance our understanding of the semantic evolution of these words by providing more accurate contextual representations. Lastly, the development of specialized lexicons and resources tailored specifically for tracking the semantic evolution of NEWs could facilitate more focused and in-depth research in this area.

References

Aho, A. V., and Ullman, J. D. 1972. *The theory of parsing, translation and compiling* (Vol. 1). Englewood Cliffs, NJ: Prentice-Hall.

American Psychological Association. 1983. *Publications manual*. Washington, DC: American Psychological Association.

Borschinger, B., and Johnson, M. 2011. A particle filter algorithm for Bayesian word segmentation. In *Proceedings of the Australasian Language Technology Association Workshop 2011* (pp. 10-18). Canberra, Australia.

Bolinger, D. 2013. *Degree words*. De Gruyter Mouton.

Bonelli, E. T. 2000. Corpus classroom currency. *Darbai ir dienos, 24*, 205-243.

Chandra, A. K., Kozen, D. C., and Stockmeyer, L. J. 1981. Alternation. *Journal of the Association for Computing Machinery, 28*(1), 114-133. https://doi.org/10.1145/322234.322243

Desagulier, G. 2019. Can word vectors help corpus linguists? *Studia Neophilologica, 91*(2), 219-240. Taylor and Francis.

Dér, C. 2013. Grammaticalization: A specific type of semantic, categorical, and prosodic change. *Berliner Beiträge Zur Hungarologie. Schriftenreihe Des Fachgebiets Für Ungarische Literatur Und Kultur an Der Humboldt-Universität Zu Berlin*, 12, 57-78.

Dragut, E., and Fellbaum, C. 2014. The role of adverbs in sentiment analysis. In *Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore (1929-2014)*. 38-41.

Geeraerts, D., et al. 1997. *Diachronic prototype semantics: A contribution to historical lexicology*. Oxford University Press.

Gulordava, K., and Baroni, M. 2011. A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. In *Proceedings of the GEMS 2011 workshop on geometrical models of natural language semantics*. 67-71.

Gusfield, D. 1997. *Algorithms on strings, trees and sequences*. Cambridge, UK: Cambridge University Press.

Hamilton, W. L., Leskovec, J., and Jurafsky, D. 2016. Cultural shift or linguistic drift? Comparing two computational measures of semantic change. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (Vol. 2016, pp. 2116). NIH Public Access.

Hamilton, W. L., Leskovec, J., and Jurafsky, D. 2016. Diachronic word embeddings reveal statistical laws of semantic change. *arXiv preprint arXiv:1605.09096*.

Haspelmath, M. 1999. Why is grammaticalization irreversible? *Linguistics*, *37*(6), 1043-1068. Walter de Gruyter.

Jatowt, A., and Duh, K. 2014. A framework for analyzing semantic change of words across time. In *IEEE/ACM Joint Conference on Digital Libraries*. 229-238. IEEE.

Jing-Schmidt, Z. 2007. Negativity bias in language: A cognitive-affective model of emotive intensifiers. *Cognitive Linguistics, 18*(3), 417-443.

Kulkarni, V., Al-Rfou, R., Perozzi, B., and Skiena, S. 2015. Statistically significant detection of linguistic change. In *Proceedings of the 24th international conference on world wide web*. 625-635.

Kugler, N. 2014. A nyelvi polaritás kifejezésének egy mintázata, avagy milyen a félelmetesen jó? *Magyar Nyelvőr, 138*(2), 129-139. Magyar Tudományos Akadémia.

Laczkó, M. 2007. Napjaink tizenéveseinek beszéde szóhasználati jellemzők alapján. *Magyar Nyelvőr, 131*(2), 173-184.

Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., and Nowak, M. A. 2007. Quantifying the evolutionary dynamics of language. *Nature, 449*(7163), 713-716. Nature Publishing Group.

Méndez-Naya, B. 2008. Special issue on English intensifiers. *English Language & Linguistics, 12*(2), 213-219. Cambridge University Press.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, 3111-3119. http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf

Pagel, M., Atkinson, Q. D., and Meade, A. 2007. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, *449*(7163), 717-720. Nature Publishing Group.

Paradis, C. 2000. It's well weird: Degree modifiers of adjectives revisited: The nineties. In *Corpora Galore*. 147-160. Brill.

Paradis, C. 2000. Reinforcing adjectives: A cognitive semantic perspective on grammaticalisation. *Topics in English Linguistics, 31*, 233-260. WALTER DE GRUYTER & CO.

Paradis, C. 2001. Adjectives and boundedness. *Cognitive Linguistics*, *12*, 47-65. Walter de Gruyter.

Paradis, C. 2008. Configurations, construals and change: Expressions of DEGREE1. *English Language & Linguistics, 12*(2), 317-343. Cambridge University Press.

Partington, A. 1993. Corpus evidence of language change: The case of the intensifier. In M. Baker, G. Francis, and E. Tognini-Bonelli (Eds.), *Text and Technology. In honour of John Sinclair*. Amsterdam: John Benjamins.

Péter, M. 1991. *A nyelvi érzelemkifejezés eszközei és módjai*. Tankönyvkiadó Vállalat.

Rasooli, M. S., and Tetreault, J. R. 2015. Yara parser: A fast and accurate dependency parser. *Computing Research Repository*, arXiv:1503.06733. http://arxiv.org/abs/1503.06733

Řehůřek, R., and Sojka, P. 2010. Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. 45-50. Valletta, Malta: ELRA.

Sturtevant, E. H. 1917. *Linguistic change: An introduction to the historical study of language* (Vol. 60). University of Chicago Press.

Szabó, M. K., and Bibok, K. 2019. Értékvesztésre és értékváltásra képes lexémák újabb vizsgálata. *Argumentum, 15*, 639-649. University of Debrecen.

Szabó, M. K., and Guba, C. 2022. Analysis of negative emotive intensifiers in Hungarian tweets. Manuscript.

Tolcsvai Nagy, G. 1988. A mai magyar nyelv normarendszerének egy jelentős változásáról az ifjúsági nyelv kapcsán. *Magyar Nyelvőr, 112*(4), 398-405.

Traugott, E. C., and Dasher, R. B. 2001. *Regularity in semantic change* (Vol. 97). Cambridge University Press.

Weinrich, H. 1962. Stephen Ullmann—Semantics. An introduction to the science of meaning. *Romanistisches Jahrbuch, 13*(1), 186-188. De Gruyter.

Xu, Y., and Kemp, C. 2015. A computational evaluation of two laws of semantic change. In *CogSci*.