Automatic Detection of Word Sense Shift from Corpus Data

Ondřej Herman

Lexical Computing, Brno, Czech Republic
Natural Language Processing Centre, Masaryk University, Brno, Czech Republic
E-mail: ondrej.herman@sketchengine.eu

Abstract

Language evolves continuously, rendering static dictionaries quickly outdated. While previous research has addressed the automatic detection of new words, identifying subtler semantic changes in existing words remains a challenge. In this work, we propose a robust, language-independent methodology for the automatic detection of word sense shifts using diachronic corpus data. Our approach builds on the Adaptive Skip-Gram algorithm for word sense induction, enabling us to model polysemy directly from raw text without reliance on external sense inventories.

We calculate the temporal distribution of induced senses and apply trend estimation techniques—specifically linear regression and the Theil–Sen estimator—to detect statistically significant shifts. This two-stage architecture decouples sense induction from trend analysis, increasing overall robustness and interpretability. Unlike traditional methods in lexical semantic change detection, which often target dramatic historical shifts, our method is designed to detect emerging or evolving senses over shorter timescales using large web corpora.

We evaluate our method on Timestamped corpora in English and Czech and present several examples of detected sense shifts. The results demonstrate the feasibility of scalable, automatic sense shift detection and its potential applications in lexicography and linguistic research.

Keywords: word sense induction; neologisms; trends

1. Introduction

1.1 Motivation

Language is an ever-changing entity, which responds to the needs of the speakers and reflects the inherent progress of the external world. Some words gain new meanings, while other meanings become obsolete.

On the other hand, a dictionary reflects a stationary view of a language at a specific point in time, and starts becoming obsolete the moment it is released. Creating a dictionary from scratch represents a large effort – it is very laborious and therefore expensive. Updating a dictionary is still difficult, because much of the work currently needs to be redone, and evidence needs to be revisited. Otherwise, we risk that the dictionary entries cease to be representative of the real state of the language.

To save work, we aim to devise a mechanism, which would let us identify words, which gained new senses, or some of the existing senses changed.

1.2 What is a word sense, anyway?

Traditional approaches assume that words have discrete and enumerable senses, and lexicographical practice reflects this. Polysemous words are divided into individual entries, often conventional in nature.

Following Wittgenstein's principle that *meaning is use* (Wittgenstein, 1953), meaning is fluid. Use of words is incredibly varied and cannot be neatly categorized into separate bins, much less along the dimensions which would provide the conventionally expected list of meanings as present in most dictionaries.

Nevertheless, it is clear that a word and its senses need to be serialized *somehow* to make a useful and descriptive dictionary entry, which would communicate valuable information to the reader of a dictionary. They are reading the dictionary to gain insight into the language and sidestep the need to play the language game to obtain the same knowledge.

Whether two different uses of a word belong to the same word sense is often subjective. The agreement of human annotators is often low when determining what the sense division for a particular word should look like. Human annotators provide sense clusterings of comparable quality as WSI algorithms in many cases (Herman & Jakubíček, 2024), but this does not generalize over all words, which appear in very varied contexts.

Words represent labels of real world entities. On the other hand, WSI approaches lack ontological grounding and operate on surface-level cues, not conceptual distinctions. The clusters obtained from WSI reflect conflated differences in discourse domains (genre, topic, register, etc.) with actual polysemy. The senses produced by WSI should therefore be best interpreted as patterns of contextual variation rather than distinct meanings.

In effect, it seems that there will need to be a human–native speaker present in a dictionary production pipeline for the foreseeable future, as the machine-provided word senses do not carry the same nuance and still describe mostly superficial features in the case of WSI.

1.3 Related Work

Cook et al. (2013) explored a method for identifying novel word senses based on LDA topic models. The difference in normalized frequency of senses in a corpus from 1995 and a corpus from 2008 is used to find new candidate entries suitable for inclusion into an English learner's dictionary. Nimb et al. (2020) use bigram frequency differences between two time periods to identify semantic change and introduce new dictionary entries into a Danish corpus. Compared to this research, our methods aim to identify subtle changes over shorter time periods. To achieve this, we add a statistical trend estimation step to improve noise rejection and improve robustness.

A related area of research outside lexicography is Lexical Semantic Change Detection (LSCD) Schlechtweg (2023); Tahmasebi & Dubossarsky (2023). LSCD focuses on determining whether a word's meaning has changed between two time periods.

The body of research in LSCD aims to track language changes over a long period of time, on the order of 50 to 500 years, and describe the historical behavior of words. Additionally, LSCD methods generally operate on a limited amount of sparse data, and the algorithms reflect these constraints. In principle, the LSCD methods can be used to identify word sense discrepancy across any two text types, not necessarily differing in time.

Approaches to LSCD include co-occurrence based methods (Sagi et al., 2009; Gulordava & Baroni, 2011; Kahmann et al., 2017), topic modeling (Frermann & Lapata, 2016), word embedding based approaches (Mitra et al., 2015; Tahmasebi & Risse, 2017) and various word sense induction based methods (Kulkarni et al., 2015; Fiser & Ljubesic, 2018; Yao et al., 2018).

The words in LSCD test sets describe words, which changed in meaning nearly completely and share very little between the examined periods. Our approach, on the other hand, aims to study shorter-term word sense changes on large, Web-scale corpora with fine-grained timestamps. The availability of precise timestamps allows us to treat time as a continuous variable, so that statistical trend analysis can be applied, leading to better robustness towards noise.

We want to identify subtler changes as they are happening and find new word senses when they are still not widely represented in the corpus. The words as a whole will carry their other meanings with a high frequency at this point, overshadowing these small changes.

Unlike LSCD approaches, which typically operate on carefully curated historical corpora with relatively sparse data, our methodology is designed for large-scale, time-stamped web corpora where the volume of data is not a limitation but a challenge. In this setting, the main problem is not detecting *any* change, but identifying *salient* and meaningful shifts amid a background of various sources of noise and word usage fluctuations. This difference in data availability and signal-to-noise ratio necessitates a different methodological focus, one centered on robustness, scalability, and relevance filtering rather than sensitivity to rare historical events.

On the other hand, LSCD datasets provide a valuable and deeply explored resource on historical semantic change, so being able to apply our methods to LSCD would help in development. At this point, we haven't found a way to bridge the differences and apply our methods in a meaningful way.

2. Methodology

We start from a tokenized corpus text, and, at the end of the process, we obtain a list of words, for which some word sense shift took place. The major steps to get there are the following:

105. Inducing word senses
106. Calculate diachronic word sense frequency distribution for each word
107. Diachronic frequency normalization
108. Estimate trend for each of the word senses
109. Identify statistically interesting word senses

We will now discuss these steps in detail.

2.1 Inducing word senses

As a first step, we apply a word sense induction (WSI) algorithm to the source corpus to identify the range of senses that each word can exhibit. Unlike word sense disambiguation, which assigns word occurrences to predefined senses from an external inventory, WSI derives senses directly from the corpus itself. This distinction is crucial, as emerging or context-specific senses may not be represented in existing lexical resources.

The selected WSI method was required to meet several constraints. It must function efficiently and autonomously on large corpora in any language, including those with limited or no existing linguistic resources. The induced senses should be of high quality and correspond closely to human intuitions about meaning. Furthermore, the method must rely exclusively on the data contained in the corpus, without depending on external sense inventories or auxiliary datasets, which may not be available for certain languages or time periods under investigation.

These criteria rule out the use of large language models (LLMs). Such models are often unavailable or insufficiently accurate for low-resource languages, and even for well-resourced languages, they are limited to knowledge available at the time of training. As a result, LLMs may fail to capture novel or emerging word uses, which are central to our study.

To meet these requirements, we adopt the Adaptive Skip-Gram (Adagram) algorithm (Bartunov et al., 2016), a word sense induction method that extends the Skip-Gram model of word2vec (Mikolov et al., 2013) by allowing multiple vector representations per word. While standard word embeddings conflate multiple senses of a word into a single vector, Adagram assigns distinct vectors to different senses, capturing polysemy more effectively.

Adagram operates by modeling word meaning in context. During training, it predicts a target word from its surrounding context words, but with an additional latent variable representing the sense. The number of senses per word is not fixed in advance but is capped at a user-defined maximum. A key parameter α controls the trade-off between sense granularity and generalization; in our experiments, we use $\alpha = 0.1$. This probabilistic framework allows the model to dynamically allocate senses only when there is sufficient evidence in the data.

One of the main advantages of Adagram is its scalability. The algorithm operates in a similar way as word2vec and can be applied efficiently to large corpora. Although modeling multiple senses introduces a constant-factor overhead relative to single-sense embeddings, this is partially offset by the ability to reduce the dimensionality of the embeddings (we use 64-dimensional vectors without sacrificing quality). Furthermore, Adagram employs hierarchical softmax, which ensures that time complexity scales as $\mathcal{O}(nlogn)$, where n is the corpus size. This makes the approach computationally feasible even for large datasets.

To further improve performance and integration with our infrastructure, we reimplemented the original Adagram algorithm in Rust. This version is optimized for speed and allows direct interaction with Manatee corpus indexes, eliminating the need for preprocessing and enabling more efficient training on large-scale data. This implementation also increases robustness and supports integration with our broader corpus processing pipeline.

In terms of sense quality, we evaluated Adagram using the ShadowSense test set (Herman & Jakubíček, 2024), which is designed to assess the alignment between automatically

induced senses and human intuition. The results confirmed that Adagram strikes a practical balance between quality, efficiency, and independence from external resources. While more recent neural architectures or transformer-based methods may achieve marginally better semantic clustering, their computational cost and dependence on pretraining data make them unsuitable for our goals, especially in a diachronic, multilingual setting.

2.2 Diachronic word sense frequency distribution

At this point, we have constructed the WSI model that captures information about the senses associated with each word in the corpus.

We then use this model to extract the raw frequency distributions of word senses over time. Each occurrence of a given headword is assigned to one of the induced senses and to a specific time period. We count the number of instances of each sense within each period.

A potentially improved trend frequency Adagram performs probabilistic disambiguation: given a word and its surrounding context, the model produces a probability distribution over its possible senses. We currently simplify this output by assigning each word occurrence to the single most probable sense—an approach commonly referred to as hard assignment. This allows us to treat sense labeled tokens unambiguously, which simplifies downstream processing.

Alternatively, the full sense probability distributions can be accumulated across occurrences. This would retain information about less likely senses and result in a smoother frequency distribution. Our observations suggest that this probabilistic aggregation may provide more robust estimates during trend analysis and reduce sensitivity to classification noise.

Nevertheless, we opted for hard assignment in the present work to maintain greater interpretability and transparency in the analysis, as this facilitates debugging and manual inspection of the sense labeled data. In a future work we may explore soft assignment as a way to detect more subtle or gradual semantic shifts that might be obscured by hard clustering.

This process yields a time series reflecting the frequency of individual word senses across the defined temporal intervals.

2.3 Diachronic Frequency Normalization

The raw frequency counts obtained in the previous step are not directly suitable for identifying trends, so we first normalize the raw frequencies obtained in the previous step. In the following, $f_{raw}(s, e)$ represents the raw occurrence count of sense s in the epoch e; S is the set of all senses, while E represents the set of all epochs.

To account for differences in epoch size, we employ the relative frequency, defined as:

$$f_{rel}(s, e) = \frac{f_{raw}(s, e)}{N(e)}$$

where $N(e)$ is the norm for the epoch e, the total amount of tokens in it.

We investigated three normalization approaches described below.

2.3.1 Epoch Normalized Frequency

The first approach is a straightforward extension of the normalization used in Herman & Kovar (2013). It normalizes the frequency of a sense across time while treating each sense independently:

$$f_{en}(s,e) = \frac{|E| \cdot f_{rel}(s,e)}{\sum_{e' \in E} f_{rel}(s,e')}$$

This formulation ensures that the total normalized frequency of a given sense across all epochs is equal to the number of epochs, $|E|$. If a sense is uniformly distributed over time, then $f_{er}(s,e) = 1$ for all e, making this metric easily interpretable.

2.3.2 Global Normalized Frequency

The second approach introduces a global normalization over all senses and epochs:

$$f_{gn}(s,e) = \frac{|S| \cdot |E| \cdot f_{rel}(s,e)}{\sum_{s' in S} \sum_{e' \in E} f_{rel}(s',e')}$$

This version retains the proportionality of raw frequencies and reduces the impact of low-frequency or fringe senses, which may otherwise dominate the normalized distribution in f_{er}. The average value of f_{gn} across all senses and epochs is 1.

2.3.3 Sense Relative Frequency

The third approach contrasts the distribution of senses within each epoch:

$$f_{sr}(s,e) = \frac{f_{raw}(s,e)}{\sum_{s' \in S} f_{raw}(s',e)}$$

This metric reflects the *proportion* of a word's usage accounted for by a given sense in a specific epoch. The sum over all senses for a single epoch is 1, highlighting shifts in the internal distribution of senses over time, independent of overall frequency trends.

2.4 Statistical Trend Estimation

In this step, we apply statistical methods for trend estimation to assess whether the normalized frequency of a word sense exhibits a statistically significant change over time, and to quantify the strength of such change. We build on the methods described in Herman & Kovar (2013).

The underlying assumption of statistical trend estimation is as follows: given a sequence of n temporal intervals, we observe the state of a process at discrete time points x_i and obtain a corresponding value y_i for each $i \in \{1, .., n\}$. Using these observations, we estimate the parameters a, the slope, and b, the intercept, in the linear model:

$$y = ax + b$$

We consider two estimation techniques: ordinary least squares (OLS), a standard approach in regression analysis, and the Theil–Sen estimator, a non-parametric method from robust statistics that is less sensitive to outliers.

The input to the statistical estimator is $x_i = i$ for every epoch ordered chronologically and y_i is set to the corresponding normalized frequency.

The output of this step is the trend slope and the corresponding p-value for every induced sense of a word.

2.5 Identify statistically interesting word senses

In the last step, we choose only those word senses, which have changed in a significant way, that is, those with p-values smaller than a given threshold. The selected words are then sorted by the trend slope in descending order and presented to the user as potentially interesting candidate headwords.

3. Evaluation

We evaluated the methodology on texts from the English and Czech Timestamped corpora provided by Sketch Engine (Kilgarriff et al., 2014). Actual data for this is hard to come by, and various existing LSCD test sets (e.g. Schlechtweg et al. (2020) or Zamora-Reina et al. (2022)) are not amenable to processing using our methods. Therefore, we were only able to estimate the precision of the methods at this point.

3.1 Timestamped Corpora

Source data. The corpora are continuously built from texts obtained from the web through RSS news feeds. The crawling started around 2014 by the Jožef Stefan Institute Trampus & Novak (2013), and since 2021, we have been extending the corpora using a custom-built web feed crawler, as described in Herman et al. (2025). In addition to English and Czech, there are currently 25 other language versions of the corpora built with the same pipeline.

Text processing. From the raw web pages, paragraphs of text are extracted using the JusText (Pomikálek, 2011) tool. Paragraphs, which are too short or which do not contain words from the target language, are discarded. The text is then processed with the Onion (Pomikálek, 2011) tool, removing duplicate and near-duplicate paragraphs. Another step is tokenization using Unitok Michelfeit et al. (2014) and then part-of-speech tagging and lemmatization by TreeTagger Schmid (2013).

Indexing. The corpus text is stored in a binary format and indexed using the Manatee corpus manager Rychly (2007) for fast access. The corpus is also available through the web interface to enable the examination of the results by end users.

Metadata. For every document in the corpus, the following annotations are available: title, URL of the source feed, URL of the document, and the publication timestamp. Where the information can be specified for a whole RSS news feed, there are genre, topic, and location annotations present.

3.2 Test Data

We focused on the period of two years between May 2023 and May 2025. For the English corpus, this represents approximately 12 B tokens, while for the Czech corpus approximately 1 B tokens.

For this experiment, we used the combination of lemmas and part-of-speech tags as the target attribute.[1]

3.3 Test Parameters

To induce the word senses, we ran the adaptive-skip gram for 3 epochs with the maximum number set to 10 and the granularity parameter to the default 0.1. The embedding dimensionality was set to 64. We then calculated the diachronic trend statistics for the top 30,000 headwords ordered by frequency using the three normalization strategies and two statistical trend estimators described above. For every corpus, we obtained 6 lists. From each of these lists, we took the top 100 headwords, which exhibited the highest trend slope, at a statistical significance level of $p < 1e - 4$. We only investigated words starting with a lowercase letter to filter out most proper names, which are only rarely interesting from a linguistic perspective.

3.4 Evaluation Strategy

We presented these results to annotators in random order and asked them to categorize each word into one of these categories:

- **OK** - I understand what is changing and this result looks interesting.
- **Bad** - I do not understand what is changing / the result is not interesting at all.
- **Error** - The lemma is bad or some other processing issue.

3.5 Results

The annotation result for the first 100 words with the most trending senses are shown in Table 1 for the English Trends corpus and in Table 2 for the Czech Trends corpus. The results are quite similar for all of the methods. Note that even though there is a significant overlap between the result sets, but it does not explain the similarity – only around 50 % between every two pairs of result sets are the same overlapping.

The proportion of interesting word senses was on average **50.7 %** for the Czech corpus, while for the English corpus it was **25.7 %**.

At this stage, only a single annotator examined the data, so the inter-annotator agreement is not known known to us.

The lower yield in the English data is likely the result of several contributing factors. These include shifts in corpus composition over time and a greater number of spurious

[1] For highly flexive languages, lemmatization is necessary to properly identify the different contexts which represent to a single headword, but for languages with limited morphology, operating over word forms would provide viable results.

Normalization	Estimator	OK	Bad	Error
Epoch Normalized Frequency	Linear Regression	28	57	15
	Theil-Sen	24	64	12
Global Normalized Frequency	Linear Regression	26	59	15
	Theil-Sen	24	62	14
Sense Relative Frequency	Linear Regression	24	58	18
	Theil-Sen	22	62	16

Table 1: Annotation results for the top 100 results extracted from the English Trends corpus.

Normalization	Estimator	OK	Bad	Error
Epoch Normalized Frequency	Linear Regression	49	47	4
	Theil-Sen	50	49	1
Global Normalized Frequency	Linear Regression	49	47	4
	Theil-Sen	50	48	2
Sense Relative Frequency	Linear Regression	52	46	2
	Theil-Sen	54	43	3

Table 2: Annotation results for the top 100 results extracted from the Czech Trends corpus.

changes that tend to emerge simply due to the larger corpus size. Additionally, English is a frequent target for spam content, which is often difficult to detect during preprocessing and may only become apparent later as more data is collected. While the precise impact of these issues remains uncertain, they are likely to introduce additional noise into the English results.

Despite the reduced proportion, a 25 % yield still represents a substantial and valuable set of candidates for investigating lexical change.

4. Examples

In this section, we present several interesting examples we encountered during our analysis of the data.

Each plot displays the frequency distribution of the induced senses over time. The senses are ordered by the magnitude of their estimated trend, with the most prominent (i.e., steepest) trend shown first. Sense identifiers correspond to those assigned by the WSI algorithm and are represented by their nearest neighbors in the embedding space generated during induction. The scale of the vertical axis is arbitrary, but consistent across all subplots to allow for direct visual comparison.

4.1 English Examples

The noun *cot* as shown in Figure 1 is trending because it is used to refer to the AI technique *chain of thought*. The meaning representing a type of children's bed is represented strongly, with other senses referring to a baseball team, financial analysis, or improper spelling of COTS, an acronym standing for *common-over-the-shelf*–a product readily available to buy. The noun *snail* (Figure 2) is trending strongly in association with beauty products, as

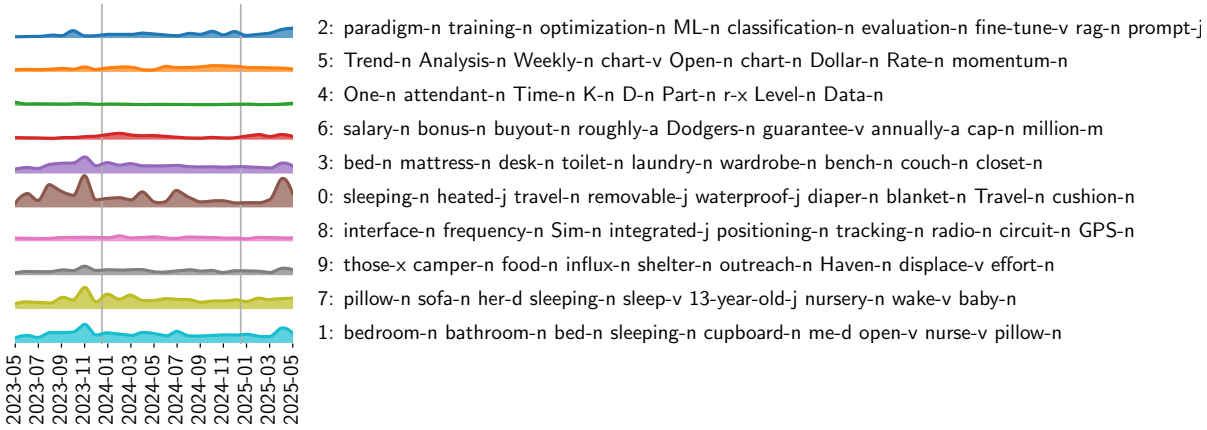

Figure 1: Sense Distribution for the English noun *cot*

creams made from snail mucus are getting popular. Other senses include snails as garden pests, snails as interesting animals, and also snail-mail.

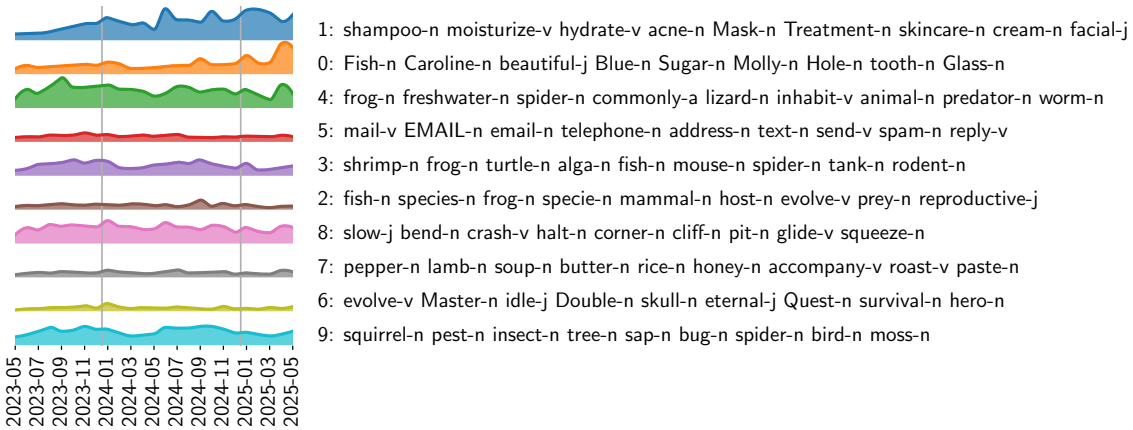

Figure 2: Sense Distribution for the English noun *snail*

Whale (Figure 3) appears to be trending in the area of finance, where it represents a large investor, which can sway the markets easily. Other senses represent the marine mammals in various contexts (scuba diving companion, cute animal, target of fishing), and as a name of a movie.

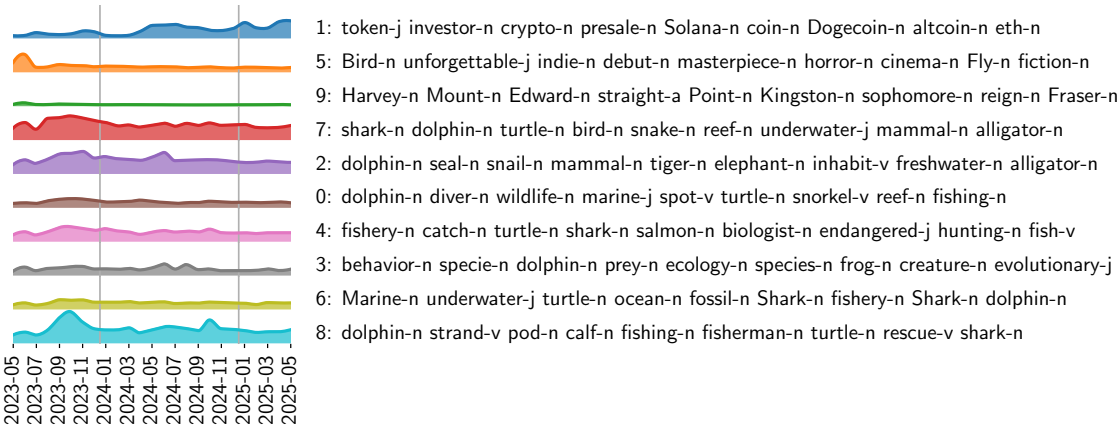

Figure 3: Sense Distribution for the English noun *whale*

4.2 Czech Examples

The adjective *vrstvený* (Figure 4) can be translated as *layered*. We can see it trending strongly in the area of fashion. Other, stationary, senses include the areas of gastronomy, construction or the concept of *layered defense*.

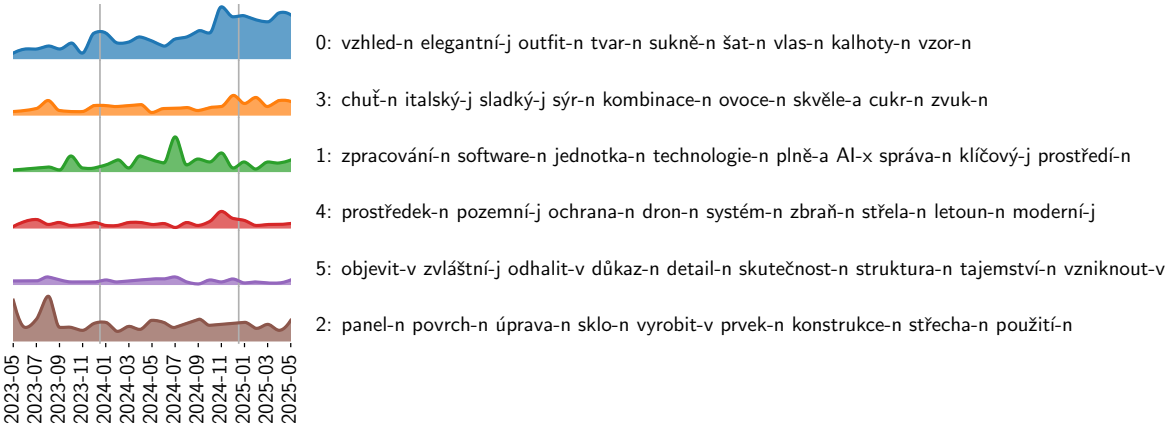

Figure 4: Sense Distribution for the Czech adjective *vrstvený*

The Czech noun *pětistovka* (Figure 5) represents the number 500. Here we see it trending as a specific kind of sports tournament. Other senses are varied, including a specific banknote denomination, or a type of motorcycle.

The Czech noun *hláska* (Figure 6) is strongly trending when used to refer to a specific kind of tower. Other senses refer to *a vowel*, an emergency telephone, or a railroad structure.

5. Future Work

There are opportunities for improvement and further research at nearly every stage of the current pipeline.

The most pressing issues, however, stem from the messiness of the real world and the interaction with a multitude of inherently unstable, diverse, and ever-changing data sources,

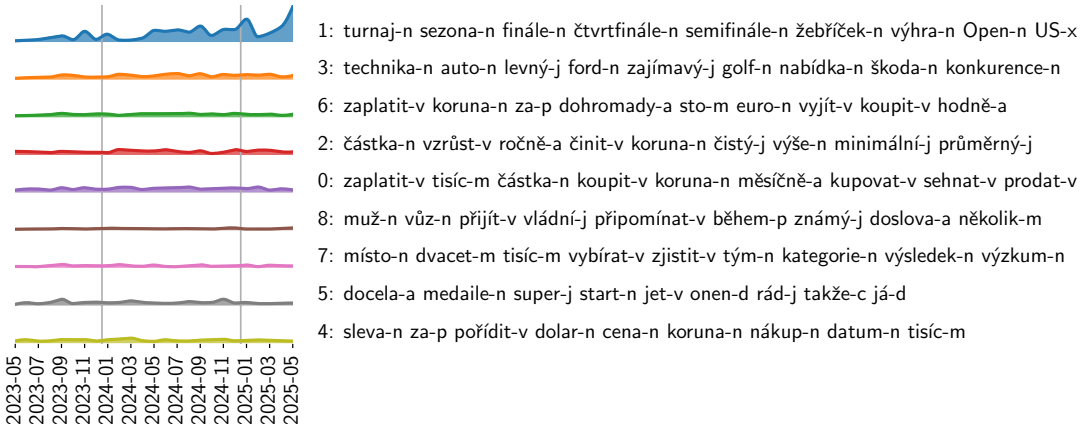

Figure 5: Sense Distribution for the Czech noun *pětistovka*

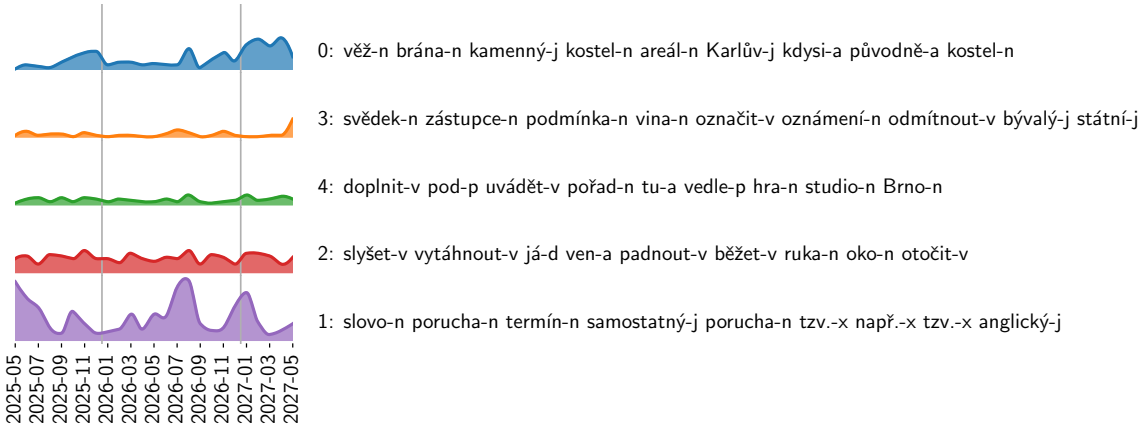

Figure 6: Sense Distribution for the Czech noun *hláska*

with which we need to keep up. A recent and large-scale web corpus will be always biased in some way and suffer from sampling and processing issues. Even though many of these problems can be fixed by the improvement of the tools we use, the reaction time is slow and the result will likely never be perfect, so our main aim is to increase the robustness of the whole system.

5.1 Reducing Influence of Rare Artifacts

Because our goal is to detect subtle shifts in word usage, the method is particularly sensitive to noise introduced during corpus construction and preprocessing. For instance, errors in text extraction from web pages, such as the retention of boilerplate content or encoding anomalies, can introduce systematic artifacts. These artifacts are often associated with specific web feeds or domains rather than being evenly distributed across the corpus. Consequently, a small number of atypical sources may disproportionately influence frequency counts for certain words or senses.

To mitigate this, future work should explore methods to downweight or exclude words and senses that appear predominantly in a narrow range of text types or domains. These types of errors also tend to influence other corpus processing tasks, such as keyword or wordlist extraction, so providing a viable solution to this issue would be useful elsewhere.

5.2 Normalization by Source

A related issue stems from the composition of the input data. Adding new web sources can, somewhat counterintuitively, reduce overall result quality: the added material may introduce new patterns that artificially inflate the frequency of certain senses. To address this, we plan to investigate normalization strategies aimed at reducing the influence of abrupt changes in source composition. Such normalization would stabilize sense frequency estimates and improve the robustness of diachronic comparisons.

5.3 Better Sense Descriptions

Currently, we use the Adagram nearest neighbors in the embedding space to describe the different senses of the word. As it turns out, this type of description was often unintuitive to the annotators. Additionally, for close senses, there tends to be some overlap between the neighbors. We believe that identifying tightly associated collocates of the word for each of its senses might serve as better disambiguators.

5.4 Thorough Evaluation

Currently, we only evaluated a small sample of the result. While the preliminary outputs look promising, we do not understand well the behavior over different time periods, for other corpora, and across different parameters, mainly the p-value cutoff. The evaluation is very laborious, and annotators find it difficult to carry out due to the aforementioned quality limitations of sense descriptors. A following evaluation will also include inter-annotator agreement estimation.

6. Conclusion

In this article, we presented a methodology for identifying words, which changed meaning over time in some way, based on corpus data.

We described the inner workings of the method: word sense induction algorithm is used to categorize word occurrences in the corpus into different word senses. The word sense occurrences are then counted, and statistical trend estimation is applied after applying one of three normalization strategies.

The evaluation describes the performance of the methods on a 2-year period of the English and Czech Timestamped corpora provided by Sketch Engine.

7. Acknowledgements

I would like to thank the anonymous reviewers for their very valuable suggestions.

Software

Bartunov, S., Kondrashkin, D., Osokin, A. & Vetrov, D. (2016). Breaking sticks and ambiguities with adaptive skip-gram. In *Artificial Intelligence and Statistics*. pp. 130–138.

Cook, P., Lau, J.H., Rundell, M., McCarthy, D. & Baldwin, T. (2013). A lexicographic appraisal of an automatic approach for detecting new word senses. *Proceedings of eLex*, pp. 49–65.

Fiser, D. & Ljubesic, N. (2018). Distributional modelling for semantic shift detection. *International Journal of Lexicography*, 32(2), pp. 163–183.

Frermann, L. & Lapata, M. (2016). A bayesian model of diachronic meaning change. *Transactions of the Association for Computational Linguistics*, 4, pp. 31–45.

Gulordava, K. & 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*. pp. 67–71.

Herman, O. & Jakubíček, M. (2024). ShadowSense: a Multi-annotated Dataset for Evaluating Word Sense Induction. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. pp. 14763–14769.

Herman, O., Jakubíček, M., Kraus, J. & Suchomel, V. (2025). From Word of the Year to Word of the Week: Daily-updated Monitor Corpora for 25 Languages. *Electronic lexicography in the 21st century. Proceedings of the eLex 2025 conference*.

Herman, O. & Kovar, V. (2013). Methods for Detection of Word Usage over Time. In *Seventh Workshop on Recent Advances in Slavonic Natural Language Processing, RASLAN 2013*. Brno: Tribun EU, pp. 79–85.

Kahmann, C., Niekler, A. & Heyer, G. (2017). Detecting and assessing contextual change in diachronic text documents using context volatility. *arXiv preprint arXiv:1711.05538*.

Kilgarriff, A., Baisa, V., Busta, J., Jakubicek, M., Kovar, V., Michelfeit, J., Rychly, P. & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1), pp. 7–36.

Kulkarni, V., Al-Rfou, R., Perozzi, B. & Skiena, S. (2015). Statistically significant detection of linguistic change. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, pp. 625–635.

Michelfeit, J., Pomikálek, J. & Suchomel, V. (2014). Text Tokenisation Using unitok. In A. Horák & P. Rychlý (eds.) *RASLAN 2014*. Brno, Czech Republic: Tribun EU, pp. 71–75.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. pp. 3111–3119.

Mitra, S., Mitra, R., Maity, S.K., Riedl, M., Biemann, C., Goyal, P. & Mukherjee, A. (2015). An automatic approach to identify word sense changes in text media across timescales. *Natural Language Engineering*, 21(5), pp. 773–798.

Nimb, S., Sørensen, N.H. & Lorentzen, H. (2020). Updating the dictionary: Semantic change identification based on change in bigrams over time. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave*, 8(2), pp. 112–138.

Pomikálek, J. (2011). *Removing boilerplate and duplicate content from web corpora*. Ph.D. thesis, Masaryk university, Faculty of informatics, Brno, Czech Republic.

Rychly, P. (2007). Manatee/Bonito–A Modular Corpus Manager. *RASLAN 2007 Recent Advances in Slavonic Natural Language Processing*, p. 65.

Sagi, E., Kaufmann, S. & Clark, B. (2009). Semantic density analysis: Comparing word meaning across time and phonetic space. In *Proceedings of the Workshop on Geometrical Models of Natural Language Semantics*. Association for Computational Linguistics, pp. 104–111.

Schlechtweg, D. (2023). *Human and computational measurement of lexical semantic change*. Ph.D. thesis, Universität Stuttgart.

Schlechtweg, D., McGillivray, B., Hengchen, S., Dubossarsky, H. & Tahmasebi, N. (2020). SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection. In A. Herbelot, X. Zhu, A. Palmer, N. Schneider, J. May & E. Shutova (eds.) *Proceedings of the Fourteenth Workshop on Semantic Evaluation*. Barcelona (online): International Committee for Computational Linguistics, pp. 1–23.

Schmid, H. (2013). Probabilistic part-of-speech tagging using decision trees. In *New methods in language processing*. Routledge, pp. 154–164.

Tahmasebi, N. & Dubossarsky, H. (2023). Computational modeling of semantic change. URL https://arxiv.org/abs/2304.06337. 2304.06337.

Tahmasebi, N. & Risse, T. (2017). On the uses of word sense change for research in the digital humanities. In *International Conference on Theory and Practice of Digital Libraries*. Springer, pp. 246–257.

Trampus, M. & Novak, B. (2013). Internals of an aggregated web news feed. In *15th Multiconference on Information Society*. pp. 221–224.

Wittgenstein, L. (1953). *Philosophical Investigations*. New York, NY, USA: Wiley-Blackwell.

Yao, Z., Sun, Y., Ding, W., Rao, N. & Xiong, H. (2018). Dynamic word embeddings for evolving semantic discovery. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM, pp. 673–681.

Zamora-Reina, F.D., Bravo-Marquez, F. & Schlechtweg, D. (2022). LSCDiscovery: A shared task on semantic change discovery and detection in Spanish. In N. Tahmasebi, S. Montariol, A. Kutuzov, S. Hengchen, H. Dubossarsky & L. Borin (eds.) *Proceedings of the 3rd Workshop on Computational Approaches to Historical Language Change*. Dublin, Ireland: Association for Computational Linguistics, pp. 149–164. URL https://aclanthology.org/2022.lchange-1.16/.