National Research University Higher School of Economics Faculty of Computer Science

Programme 'Master of Data Science'

MASTER'S THESIS

Methods and Tools for Lexical Semantic Change Detection

Student: Vahe Yepremyan

Supervisor: Dr. Nikolay Arefyev (к.ф.-м.н. Арефьев Николай Викторович)

Moscow, 2021

Summary

Lexical semantic change detection (LSCD) is the task of identifying words that change their meaning over time. A specific sub-task is determining how given words changed, which senses appeared, and which disappeared. Currently, computational methods for distinguishing senses of ambiguous words significantly lag behind humans, so lexical semantic change detection remains a task for humans. However, tools that facilitate human-machine collaboration can significantly reduce human effort. In particular, with visual, interactive applications that allow users to understand usage patterns for given words, provide a preliminary, “imperfect” grouping of example usages based on the sense, with the ability to make corrections.

First, this work discusses the existing lexical semantic change detection solutions and their shortcomings, focusing on visual, interactive tools and underlying algorithms. Then, it proposes a toolset to facilitate the task and the research in the field. The proposed system is a web-based application that performs word sense induction (WSI), the task of grouping example usages based on their sense, for a given set of target words. In particular, it employs the context clustering approach for WSI, in which each occurrence of a word is represented as a context vector. Context vectors are grouped into clusters using hierarchical clustering. The system provides a user interface to see the clustering results using multiple different visualizations, charts like scatter plots, histograms, clustering dendrogram, the ability to see example usages, tools to inspect intermediate clusters, and making corrections. It emphasizes the interpretability of clusters and the ability to complement computational results with human judgment.

As an experiment, the system was used to process a set of words from the book “Two Centuries in Twenty Words” (Dobrushina et al. 2016) that evolved their meaning over time. Pre-Soviet and post-Soviet subcorpora of the Russian National Corpus[1] was used as a source of example usages. The results were compared to the results obtained by the book’s authors. The results showed that a set of common errors in computational results could be corrected with minimal human effort, significantly improving final results.

[1] https://ruscorpora.ru/

Contents

1. Introduction

1.1 Motivation

Lexical semantic change detection (LSCD) is the task of identifying words that change their meaning over time. A particular subtask of it is the identification of senses and their change for a given word. The latter is the focus of this work. Words are ambiguous; they may take different meanings/senses depending on the context. For example, the English term ‘cell’ may be used in the senses of ‘cell phone,’ ‘prison cell,’ and ‘body cell.’ In addition, languages evolve, words acquire new senses or lose existing ones. For example, the same term ‘cell’ did not have the sense of the ‘phone’ before mobile phones were invented in the early 90th. These changes are often triggered by significant events, like major technological inventions (creation of the internet, the invention of electricity), revolutions (Russian Revolution, formation of Soviet Union), making the understanding of semantic change over time not only interesting from an etymological perspective but also social and cultural. With the development of information technology in recent decades, various tools, datasets, and information retrieval systems were developed that facilitate the study of word senses for lexicographers, etymologists, and professionals in other related disciplines. However, a big part of the lexicographic process is still manual, specifically the identification of senses and frequency of their use over time. While there was significant progress in developing computational methods for distinguishing the senses of ambiguous words, computers are still much worse than humans when determining the sense of an ambiguous word from the context, so lexical semantic change detection remains a task for humans. Nevertheless, tools that facilitate human-machine collaboration can significantly reduce human effort. In particular, with visual, interactive applications that allow users to explore usage patterns for given words, provide a preliminary, “imperfect” grouping of example usages based on the sense, with the ability to make corrections.

1.2 Goals

This work aims to produce a set of tools that will facilitate the detection of lexical semantic change for a given set of target words and help advance the research in computational lexical semantic change detection.

The ’related work’ section covers a vast landscape for lexical semantic change detection methods. There are also several attempts to visualize the results of computational solutions. However, as we will see later, there are some shortcomings

in each of the existing approaches, especially visualizations that accompany those. On a high level, **the problem** can be framed as follows: Words change their meanings over time, existing computer-based solutions that given two corpora of different periods and a target word, allow finding all its senses in both corpora and measuring the degree of the change for each sense, have shortcomings, in particular, they show the results of computation, without allowing to inspect the intermediate states of it, as well as make corrections to it to obtain better results. Two outcomes the solution is aiming to produce are:

1. A computer-aided method of discovering senses for a given word in a given corpus and inspecting example usages.
2. A method for identifying how the set of senses mapped to a given word evolved over time.

1.3 Subtasks

The proposed solution consists of two parts, the computational and the manual. The former corresponds to an automated, algorithmic way of identifying senses; the latter corresponds to interpreting, reviewing, and correcting results.

For automated sense identification, a specific approach to word sense induction (WSI), the task of grouping/clustering samples of word uses together based on their sense, is used. The method is based on an approach initially proposed by Baskaya et al. (2013) and advanced with more research (Zhikov et al., Amrami et al., 2018, 2019, Arefyev et al. 2019) in recent years. The idea is to predict a substitute word for the target word in a context and use the set of predicted substitutes to represent the word in context. It was already shown before (Peters et al. 2018) that besides demonstrating competitive results, this approach allows the possibility to form interpretable signatures for each identified cluster using the most descriptive substitutions within the cluster. For example, *'motorcycle, truck, plane, driver, train, engine, horse, bus, wagon, passenger'* could be a signature for a cluster with samples where the target word is used in the sense of a means of transportation. As mentioned earlier, interpretability and the ability to explain senses of automatically identified clusters is an essential property of a lexical-semantic change detection system, making substitutions-based WSI a good choice for the computational part of the solution.

The manual part of the solution serves two purposes: communicating the computation results to the user and making adjustments to obtain better overall results. For that, a user interface was built, with several considerations.

- Ability to easily see the identified clusters, their signature, and the example usages they are composed of.

- Ability to see the relative distances between representations of example usages, with breakdowns by corpora, predicted labels, and gold labels (in case of labeled data).
- Interactivity and the ability to review and correct the computed results have been essential considerations.

Based on these, the choice of technology stack landed on a web-based application. The WSI part is seamlessly developed in Python on the server-side, and the visualizations in feature-rich web stack for building user interfaces like Javascript and CSS.

2. Related work

In recent years there has been a substantial amount of research in lexical semantic change detection; along with the development of various methods for change detection, there have also been numerous visualization and user interface solutions developed. This overview follows the categorization structure from a recent survey conducted by Jatowt et al. 2021. While most of the approaches and tools reviewed in this work are already covered by Jatowt et al. 2021, the aim here is to detail each, when possible, and elaborate on gaps and shortcomings. The aim is not to produce an exhaustive list of every tool or method developed for lexical semantic change detection but rather a compilation that tries to capture the breadth of existing tools based on various approaches. A high-level description of the underlying approach and the depth of information the visualization method or the tool provides is covered for each case. In addition, some of the shortcomings and improvement opportunities are pointed out for each case. The solution presented in this thesis attempts to address these shortcomings, at least partially.

2.1. LSCD and the need for visualization tools

In their survey, Jatowt et al. 2021 present how computational methods for lexical semantic change detection are beneficial to various professionals: linguists, historians, sociologists, and practitioners in numerous related fields. Interactive, visual tools on top of these systems are helpful not only for professionals but also non-professionals, especially online visual, interactive systems, easily accessible to the broad public are important to help to spread knowledge about etymology. Precise representation of the meanings of terms in the past is beneficial for a range of NLP and information retrieval tasks, like text search in old text. Visualization systems tend to be attractive as they complement automatic analysis or serve as the primary tool for analysis, not only for professionals and scientists.

2.2. Tools based on frequencies of terms and n-grams

Google search (Appendix A, Figures 1 and 2) tool allows searching for definitions. For example, querying for “define [word]” in Google word search will output basic information about the words’ origin, definition, and popularity over time, from the google dictionary service. The definitions are grouped by topics, with an additional breakdown by part of speech within each topic. The origins are visualized as fragments of directed graphs, where vertices are languages, and edges show the

transitions from older language to newer one, from left to right, in chronological order.

While the Google services are widely popular and accessible, and definition search provides basic information about definition and change frequency of use over time, it does not provide information about the change of the meaning.

Sketch Engine[2] (Appendix A, Figure 3) (Kilgarriff et al. 2004) is a popular online corpus manager and text analysis software. Its purpose is to enable people studying language behavior to search extensive text collections. It supports complex and linguistically motivated queries. The system includes 570 corpora in 97 languages. It also allows the user to upload their corpus and scrape websites based on user-defined addresses and scraping configuration. 134 corpora are timestamped. The capabilities Sketch Engine offers include collocations and word combinations search and ranking by frequency, thesaurus, concordance, n-gram search and ranking by frequency, ranking words and lemmas by the frequency with filtering capability by part of speech, prefix, and suffix. It also has terminology extraction and diachronic analysis tools, particularly for detecting neologisms. The diachronic analysis tool uses DIACRAN (Kilgarriff et al. 2015), a framework for finding the “most interesting” words (or terms, etc.) for analysis. To do that, they first divide the corpora into subparts corresponding to periods and compute frequencies per million words (they also propose another option for normalization, to classify a word as present or absent in a document, and to work with counts for each word per thousand documents) within each subcorpora for each word. Next, they plot frequencies over time and find the ‘best fit’ line using linear regression (another option they mention is Theil-Sen gradient estimation). Next, they give each word a score combined with three characteristics, the high gradient of the line, the correlation, and overall frequency. This method powers the diachronic analysis tool in Sketch Engine; the tool’s output is a list of words ranked according to the criteria mentioned above.

While this method is helpful to identify words for analysis and detect neologisms, the user is not provided with additional tools dedicated to semantic change detection. Instead, after receiving the list of candidate words, the user can use the system’s remaining general-purpose tools for semantic change analysis.

[2] https://www.sketchengine.eu/

Google Books Ngram Viewer[3] (Appendix A, Figure 4) is another tool from Google. Michel et al. 2011 introduced the term 'culturomics,' the study of cultural and historical phenomena based on large textual data. They collected and computed frequencies of n-grams from millions of books available in the Google Books dataset and did a quantitative analysis to showcase its use for investigating cultural trends. Additionally, based on the computed frequencies, they have developed an online tool, 'Google Books Ngram Viewer,' that plots a line chart with the year of use on the X-axis and frequency for a queried n-gram. The tool offers rich querying functionality. It allows the comparison of frequency plots of several n-grams. Wildcard queries allow one to use an asterisk in the place of a word to automatically replace it with top 10 substitutions or do a case-insensitive search. It is also possible to issue queries based on the part of speech tag of words, showing results for words as nouns or verbs, or other parts of speech, by appending _VERB or _NOUN to the search word. Inflection search functionality allows searching for all modifications of a word that represent various grammatical categories such as aspect, case, gender, mood, number, person, tense, and voice. For example, the 'create_INF' query will show the results for 'create,' 'creates,' 'creating,' 'created.' Since the tool is based on word use frequencies, it does not provide a signal about what meanings these uses have.

Variability-based neighbor clustering (Appendix A, Figure 5) is a variation of hierarchical clustering introduced by Hilpert and Gries (2008). They use diachronic corpora to cluster adjacent time periods if the frequency of use for the term does not have significant changes. In a case study, they first constructed vectors of collexeme strengths for 1,201 verb lemmas that occurred at least once after the word *shall,* per time unit. Then compute the Pearson product-moment correlation coefficient as a similarity measure between time periods. Gries & Stefanowitsch (2003) introduced a framework for quantifying the repulsion and attraction between lexemes and grammatical constructions. They refer to lexemes that are attracted to a particular construction as their collexemes. Hilpert and Gries (2008) used an extension of distinctive collexeme analysis (Gries & Stefanowitsch 2003) in which they computed the collexeme strength as a base 10 log of the probability to get at most the observed frequency given the expected frequency for a collexeme. The dendrogram

[3] https://books.google.com/ngrams

of the hierarchical clustering gives visual information about the frequency change over time.

2.3 Tools for context-based approaches

Combining several kinds of NLP approaches (Appendix A, Figures 6,7,8) in a framework with various types of visualizations and plots to maximize the signal of semantic change was proposed by Jatowt and Duh (2014). The framework explores semantic change at three levels: lexical, contrastive-pair, and sentiment orientation. They use both the Google Books Ngrams and COHA datasets; however, their approach is different from Michel et al. 2011, the objective they set is to study the evolution of word meaning by investigating its context across time, rather than the cultural and social aspects based on frequencies of word use. Later, they developed an interactive online system for semantic change analysis (Jatowt et al., 2018); however, they did not use neural network-based approaches for word representation for simplicity. Instead, they used bag-of-words representation and distance-aware bag-of-words, where the distance is the relative position of a context word to the target word. They split the corpora into decade-long subcorpora and built vector representations for each word for each subcorpus. They offer to choose from three options as similarity measures: Pearson Correlation Coefficient $sim(X,\ Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$, cosine similarity $sim(X,Y) = \frac{\sum_{i=1}^{n} X_i Y_i}{\sqrt{\sum_{i=1}^{n} X_i^2 \sum_{i=1}^{n} Y_i^2}}$, Jaccard Coefficient $sim(X,Y) = \frac{X \cap Y}{X \cup Y}$, where X and Y are the word vectors. The system provides three types of visualizations. First is the line charts, with a decade on the X-axis, $sim(D_i,\ D_{i+1})$, $sim(D_t,\ D_i)$ and frequency on the Y-axis, where D_i is the i-th decade, D_t is the target decade. The second visualization is the word cloud of frequencies for co-occurring words, with a line chart of frequency over time per co-occurrence. The system's third and last type of visualization is the temporal 'Word Tree.' It is a tree structure, from left to right, where the target word is the vertex on the left; it branches out into the most frequent words used immediately after it. The tree has three levels in total, the number of branches is configurable.

While the system provides a novel and diverse set of visualizations, it does not contain information about particular meanings of the word. Although nevertheless, it

is still helpful to see high-level changes the word of interest went through the time; for deeper analysis and understanding in what senses the word was used, one would need to turn to other tools and sources of information.

Area charts for word sense frequencies over time (Appendix A, Figure 9), allowing to see the shifts in word meaning at a glance, were proposed in work by Rohrdantz et al. (2011). Authors use latent Dirichlet allocation (LDA) (Blei et al., 2003) to identify the senses of words and track their intensity of change over time. They apply LDA to 25 words before and after the target word to produce a context-specific probability distribution of 'topics' and draw a latent topic z from $P(context\ topic = latent\ topic \mid context)$. Authors assign the highest probability topic $argmax_z P(context\ topic = latent\ topic \mid context)$ to the context, omitting topics with less than 40% probability. To describe topics/senses, they take the five highest probability words from the topic distribution $P(word \mid context\ topic = latent\ topic)$. They apply the method to the New York Times Annotated Corpus[4] and plot stacked area charts per identified sense. Each layer on the Y-axis means 10% of contexts with target words are associated with the sense. The X-axis is the year when the text was created. They stack plots for all senses vertically, providing a convenient way to see the high-level shifts in senses over time.

While the overall method gives a good insight into sense change, there is often significant overlap between words describing topics, making it hard to understand and explain the actual senses of topics.

A Scatter plot of contexts with semantic distances preserved (Appendix A, Figure 10) was proposed by Heylen et al. (2012) as a way to see relative positions of example usages of a given word on a single plot. Authors use bag-of-words representation for windows size five and use non-parametric Multidimensional Scaling to draw a scatter plot of contexts in 2D, with distances between them in semantic vector space preserved. To build vector representations of words, they use the PMI-weighted sum of type-level vectors of the context $\vec{o_i^w} = \frac{\sum_{j_i^w}^n pmi_{c_j}^w * \vec{c_j}}{\sum_j^n pmi_{c_j}^w}$, where $\vec{o_i^w}$ is the token-level vector representation for i-th occurrence of word w, $\vec{c_j}$ is the type-level vector for j-th co-occurrence term in the context, $pmi_{c_j}^w$, is the pointwise

[4] https://catalog.ldc.upenn.edu/LDC2008T19

mutual information $pmi_{c_j}^{w} = \log_2 \frac{P(c_j|w)}{P(c_j)}$. The authors present a case study focused on Dutch words extracted from Dutch newspaper articles published between 1999 and 2005, organized in 218 synsets containing 476 nouns in total. First, they plot the synonyms of the target word in the same space in different colors. Then, they show the actual usage in context when placing the mouse pointer on the point on the plot.
While this method is a convenient way of seeing the overall distances and visually recognizing and exploring clusters, it does little to find and surface meanings of words and their change over time.
Similarly, dimensionality reduction techniques like principal components analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE, van der Maaten & Hinton 2008) have been frequently used to plot shifts in word meanings over time in 2D (e.g., Hamilton et al. 2016, Kulkarni et al. 2015).

Term clouds (Appendix A, Figures 11,12) were used by Xu and Crestani (2017) to visualize the most similar terms to the target word from different time periods and their relative frequency as font sizes. They also draw a heatmap with the Y-axis for most similar words to the target term, X-axis as time periods, and values as the measure of similarity. They build vector representation using the Word2Vec model on the New York Times and the National Geographic datasets.
Authors construct word vectors separately for each time period of interest to measure similarities. They follow Kim et al. (2014) proposal and train a sequence of Word2Vec models for each period one by one, and each model's weight is initialized based on the previous one. They define the similarity for words $w_{1,}w_2$ as $sim(w_1, w_2) = [sim(w_1, t_1, w_2, t_1), sim(w_1, t_2, w_2, t_2), \ldots, sim(w_1, T, w_2, T)]$,
where $T$is the number of time periods, and the similarity for $w_{1,}w_2$within a time period is given by $sim(w_1, w_2) = \frac{w_1 * w_2}{|w_1| * |w_2|}$.
This method gives good insights into contexts in which the target word was used and the change over time; however, there is no information about the sense and no ability to see examples of usages in various senses.

2.4. Dictionary-based tools

The Online Etymology Dictionary[5] **(Appendix A, Figure 13)** is an online service that combines explanations and etymologies of words from over 80 various sources. The website implements an internal search that returns explanations and descriptions of words, with dates when the word was used for the first time. It is easy to use; however, the service is only for English. While it provides explanations for various meanings, emphasizing the past meanings and basic information about the time of origin, it does not surface information about the semantic change over time. It also does not include recent neologisms like “cryptoasset” or “Webinar.”

“Diachronlex” diagrams were introduced in work by Theron & Fontanillo (2015). Authors utilize different editions of Spanish language dictionaries over time: 1780, 1817, 1884, 1925, 1992, and 2001 editions provided by the Royal Spanish Academy to develop an interactive visual tool for exploring semantic change over time. Authors draw a matrix where columns are dictionaries, in chronological order, with the most recent one on the left. Rows are senses of the target word defined by dictionaries, in ascending order, top to bottom. They connect related senses with lines across columns and rows. Authors use BLEU and NIST metrics (Zhang et al. 2004) to connect meaning definitions; two senses are connected if they cross a predefined threshold. The tool offers the ability to annotate and ‘correct’ the connections.
The described system gives a good view of the target word’s senses and evolution over time. However, it is limited to available dictionaries, comprehension, and freshness.

2.5. Other systems

Visualization of the evolution of named entities (Appendix A, Figure 14) was done by Mazeika et al. (2011), who used the Yet Another Great Ontology (YAGO)[6] knowledgebase to extract named entities from the New York Times collection of news articles, discover and track their evolution over time. They came up with a collection of more than 1.8 million news articles published from 1987 to 2007, resulting in a database with 50 named entities. They visualize the evolution of named entities as a timeline, using stacked areas. Given a named entity as a query term, they draw other entities co-occurring on the same Web page, document, or tweet as

[5] https://www.etymonline.com/
[6] https://yago-knowledge.org/

stacks. The stack area depicts how prominent a named entity is to the query term, based on the frequency of co-occurrence.

Word search with specified meanings was proposed by Zhang et al. (2019), as a method to search for usages of a given word in old archives. Authors present a tool that allows users to search for words under certain aspects (e.g., word: euro, aspect: currency) to find uses of replacement words from past time periods. The search result is a list of usages, with replacement words highlighted. As an example, the authors present the search results for 'euro' under the aspect of 'currency.' The list of results shows uses of "lira," "franc," "sterling," etc., for the time period 1987, 1991. Authors use the skip-gram model (Mikolov et al. 2013) to build word vectors for each time period in diachronic corpora. They use a 'transformation' model $\psi(\cdot)$, a three-layer neural network, to align vectors between source and target vector spaces. They train the network with mean squared error loss function $L = \frac{1}{N}\sum_{i=1}^{N}(\phi(x_i^b) - x_i^t)^2$, where x_i^b, x_i^t are the word vectors from source and target spaces, respectively. For aspect-based search, they first retrieve the analog of the aspect a in the target space $a^{'} = argmax_e\,(cos(\psi(a),\, e))$, where e is the word vector in the target space. Next, they compute the similarity between tuples $(q,\, a)$ and $(e,\, a^{'})$, where q is the query, e is the word vector in the target space, $r = argmax_e\,(\lambda(\frac{cos(\psi(q),\, e) + cos(\psi(a),\, a^{'})}{2}) + (1 - \lambda)(cos(\psi(q) - \psi(a),\, (e - a^{'}))))$. The query result is the list of the top 100 most similar tuples. The first part of the equation authors refer to as 'semantic similarity' between tuples and the second part as the 'relational similarity.' The intuition behind the relational similarity is to measure that the relation between the aspect and the query is similar across spaces. The tool is implemented mainly as an information retrieval system and not as a solution for detecting changes in senses for the given word.

2.6. Substitutions-based WSI

As mentioned in the introduction, word sense induction is the task of grouping/clustering examples of word usages based on their sense. Baskaya et al. (2013) introduced a WSI method based on substitute vectors, each use of the target word is represented by a distribution over in-context substitutes for the word, and clustering is performed over these distributions. This approach by (Amrami and

Goldberg, 2018, 2019) further improved the method by replacing words with dynamic patterns "(_ and T)," first using ELMo (Peters et al. 2018), and then BERT (Devlin et al. 2018), instead of the 4-gram model initially used by Baskaya et al. Later, Arefyev et al. (2019, 2021) introduced several enhancements to the method, in particular, to optimize it for the Russian language. They have added XLM-R (Conneau et al.) as a base model and employed a combination of bi-directional patterns like "_ and T" and "T and _." Briefly, the outline of the method for a given word is:

- In every sample usage of the target word, replace the target word with a template "_ and T."
- Using the language model, generate the distribution of possible substitutes for the masked word, keep top K most probable substitutes
- Build TF-IDF vectors over the substitutes for all samples
- Cluster resulting vectors using agglomerative clustering with cosine similarity as an affinity function with average linkage.
- Choose the number of clusters maximizing the silhouette score $\frac{1}{n}\sum_{i=1}^{n}\frac{b_i - a_i}{max(a_i, b_i)}$ where a_i is the average distance between i-th sample and all other samples in the same cluster, b_i is the average distance between the i-th sample and all samples in the nearest cluster (different from the cluster of the i-th sample).

2.7. Conclusions on related work

As mentioned before, this is by no means an exhaustive list of all tools and approaches to lexical-semantic change detection. However, this covers most, if not all, major categories of approaches and tooling. While exploring the vast landscape of solutions and interviewing professionals in linguistics and lexicography, several characteristics were identified that describe a good solution. Firstly, there is a lack of a clearly defined set of capabilities that a system for semantic change detection should provide. We define those as "Take two corpora and a set of target terms as an input, give as an output a set of interpretable senses for each target term, and a degree of difference between two corpora for each sense." Note that this does not necessarily mean that the system should produce the output in an entirely automated way, without human input; on the contrary, effortlessly complementing automated results with human input is essential for such a system. Based on this, we define characteristics of a complete system as:

- It should have the capability to group examples of usages of the target words by sense.
- Meanings should be interpretable/explainable by humans.
- It should have the capability to numerically see the degree of difference between the presence of a sense across corpora

3. The proposed toolset

The proposed system (WordSense) is a collection of tools and algorithms for semantic change detection. It attempts to solve shortcomings of existing solutions, specifically focusing on the interpretability of senses and their change over time.

3.1. The underlying algorithms

The system is based on grouping/clustering word uses based on their sense. It performs WSI to identify clusters of example usages for the target word in the given corpora. Additionally, it provides interactive tools to interpret the results, review, and assign senses to clusters, adding human judgment on top of the computational solution. Each sense is represented by a collection/cluster of example usages and a signature that will be discussed in more detail later. Similarly, each collection/cluster has a signature and computed sense, which may differ from the final, reviewed, and manually assigned sense.

In particular, the used method is the one described in section 2.6. The main advantage of this approach that makes it suitable for the system is its interpretability. Word vectors have substitutions as features; this is an important differentiator for substitution-based WSI than other methods. Furthermore, the collection of most probable substitutions gives a clear, interpretable signal into the meaning/sense of the word in the given context. We will extensively use this in the system.

3.1.2. Semantic change detection

For semantic change detection, we compute the pointwise mutual information between each corpus and each of the identified senses and use that to measure how much a particular sense associates with a given corpus. The intuition behind the PMI is the difference between the conditional probability of the corpus and absolute probability; if the old corpus is more probable within a given sense than in the combined dataset, we conclude that the sense is associated with the old corpus; similarly, if the new corpus is more probable inside the sense than in the combined dataset, we conclude that the sense is associated with the new corpus. Thus high PMI between the new corpus and the sense means that the sense is associated (it is more frequent in the new corpus than in the old) with the new corpus; high PMI between the old corpus and the sense indicates that the sense is associated with the old corpus.

$$PMI(Corpus,\ Sense)\ = log_2\ (\frac{P(\ Corpus\ |\ Sense)}{P(Corpus)})$$

When $p(Corpus\ |\ Sense)\ =\ 0$, the PMI is not defined; in that case, the PMI is given the value of -Inf, indicating the absence of examples of the given sense in the corpus. For example, the PMI value of 3 between the new corpus and a given sense means that the probability of a randomly sampled example from the dataset to belong to the new corpus is 8 times lower than the probability of a randomly sampled example from the collection of items in the identified sense. However, PMI may be misleading for the senses with a few examples. Therefore, the number of examples of a certain sense should be considered. Empirically, $abs(PMI(Corpus,\ Sense))\ >\ 2$ and $\frac{size(Sense)}{size(Dataset)} > 0.05$ is a relatively high confidence signal to associate (or negatively associate in case of negative PMI) the sense with the corpus, given that the number of examples in corpora are not significantly different and overall are large. However, this may vary depending on the dataset. Therefore, after associating senses with corpora, a manual review is required for senses that still have a small number of examples from the non-associated corpus. The review's goal is to ensure that these examples were included by clustering mistake; if that is not the case, then we conclude that it is likely that the given sense changed its frequency over time, but it neither got lost nor is entirely new.

3.1.3. A new approach for measuring the distance between clusters

An alternative approach to word representations has been explored as part of this work. Instead of building and using BoW vector representations for words, use probability distributions of substitutions. In the clustering stage, cosine similarity affinity was replaced with Jensen-Shannon distance, a symmetric distance measure between probability distributions. The benefit of this approach is that it allows us to use probabilities assigned to each of the top K substitutes instead of dropping them like we do when building TF-IDF vectors. In addition, we use the geometric average of probability distributions of all words within a cluster as the probability distribution for the whole cluster, amplifying the importance of high probability substitutes within the cluster. A few experiments during the development showed comparable results to the more common, cosine similarity-based approach; however, this method is implemented as an alternative and not the primary approach used in the tool due to lack of stability. Exploration and a better understanding of the described approach are left for future work.

3.2. Technical implementation

3.2.1. Technology stack

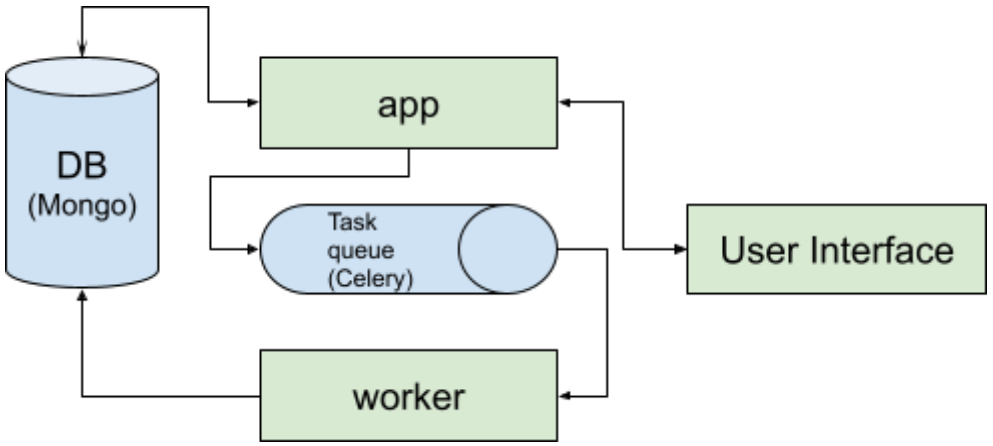
The system is a containerized web application. Docker[7] was used for containerization on the server-side, Celery[8] for asynchronous task queue, Python[9] as the programming language, Flask[10] as a web framework, and MongoDB[11] as a database. Scikit-learn[12], numpy[13], scipy[14], openTSNE[15], pandas[16], seaborn[17], pymorphy2[18], nltk[19], fire[20] open-source libraries were used for machine learning-related tasks in the system. On the front-end, vanilla Javascript was used, along with cytoscape.js[21], d3.js[22], ChartJS[23], Material Design Lite[24] open-source technologies.

3.2.2. System architecture overview

The system consists of two main parts, the backend, and the front end. The backend is executed in two modes: app and worker. Both worker and the app run within Docker containers as services; they communicate using Celery (with Redis as a backend). The front end is served to the client by the API service. Worker service is responsible for computationally heavy tasks; it is where the actual WSI task is executed. App service's primary responsibility is to facilitate the communication between the user interface and the worker. It encapsulates a web server, which takes user input from (dataset, hyperparameters for WSI, and set of target words) the UI, saves it to the database, and sends a message to the worker to process the user input. Worker and app services are implemented as a single monolithic application executed in these two distinct modes; this enables easy sharing of

[7] https://www.docker.com/
[8] https://docs.celeryproject.org/
[9] https://www.python.org/
[10] https://flask.palletsprojects.com/
[11] https://www.mongodb.com/
[12] https://scikit-learn.org/
[13] https://numpy.org/
[14] https://scipy.org/
[15] https://opentsne.readthedocs.io/
[16] https://pandas.pydata.org/
[17] https://seaborn.pydata.org/
[18] https://pymorphy2.readthedocs.io/en/stable/
[19] https://www.nltk.org/
[20] https://google.github.io/python-fire/guide/
[21] https://js.cytoscape.org/
[22] https://d3js.org/
[23] https://www.chartjs.org/
[24] https://getmdl.io/

functionality between two services. We follow the principles of object-oriented programming to organize the code in objects of various classes responsible for particular tasks. We define interfaces between these objects and ensure that dependencies throughout the project are on interfaces and not implementations. It allows having the flexibility to seamlessly change the underlying logic of any object without significant architectural changes. Further in the document, all described dependencies between objects are dependencies on interfaces and not specific implementations. Let us look into each of the three components in more detail.



3.2.2. Worker service

As already mentioned, worker service is responsible for executing computationally intensive tasks. These include the core WSI task and the preprocessing of datasets.

Dataset preprocessing

The tool allows the user to upload, save and reuse datasets. Currently, the tool accepts a custom format, where we provide a CSV file, with the following fields: context (the context in which the word was used), word (target word), positions (start and end index of the target word in the context) and label (gold label in case if the dataset is labeled). Additionally, with each dataset, we provide a file in the .npz format that contains precomputed substitutes for each sample in the dataset. Datasets can consist of one or two corpora of the same format.

WSI task execution

To execute the WSI task, the worker takes as an input an ‘experiment’ object (initially created and saved to the database by the app service based on the user input). The experiment object contains information about the target words, dataset, and hyperparameters. The WSI task execution is done in two stages - vectorization and

clustering. We define vectorizer objects that encapsulate the underlying vectorization method for the vectorization stage. Similarly, we define clusterer objects that encapsulate the underlying clustering method for the clustering stage. The experiment object also contains the choice of vectorizers and clusterers. We allow an array of values for hyperparameters, applying grid search to find the best combination that results in the best silhouette score for the resulting cluster. We use silhouette score as a measure of clustering quality; however other methods also exist; exploration of those is out of the scope of this work. Finally, we implement a 'solver' object to execute the grid search, which takes input vectorizers and clusterers, initialized with all combinations of specified hyperparameters, and the experiment object. It applies vectorization and clustering to all combinations and finds the clustering with the best silhouette score. Next, the solver computes all the necessary metrics and data structures to visualize results on the user interface and save those to the database.

3.2.3. 'App' service

App service's primary responsibility is to connect the user interface with the worker. It encapsulates a web server that serves the HTML and Javascript code to the client web browser. It takes input from the user via the user interface, creates an experiment object based on the input, saves it to the database, and schedules asynchronous tasks to run the experiment for the worker. Another responsibility of the App service is to read the experiment results computed and saved by the worker to the database and serve them to the UI. Another task that the app service handles is dataset uploading. It receives the dataset files from the user interface, saves them in a temporary location, creates a dataset object, saves it to the database, and schedules asynchronous tasks for the worker to read the dataset object and preprocess it.

3.2.4. User interface

The user interface is the central part of our application. It implements visualizations and tools for the user to access the results of the WSI task and additional metrics helpful for the lexical-semantic change detection task. It takes the results of the computations by worker service via the app service and renders visualizations in the user's web browser. The interface is organized into three navigational pages: main screen, experiment screen, dataset upload, and four visualization/tooling pages: scatter plot, dendrogram, inspect senses, clustering summary. Visualization and

tooling will be covered in separate sections in more detail since those are the core of the system.

Main page (Appendix B, Figure 1)

The main page is where the user lands when opening the tool. It is split into two parts; on the left is a form with input fields for clusterer, vectorizer, and target words. Depending on the choice of clusterer, additional inputs are required, such as affinity, linkage, number of clusters (if the choice of clustering algorithm requires the number of clusters as a parameter). Similarly, inputs such as maximum document frequency, minimum document frequency, analyzer (word, character), and top K substitutes are requested based on the choice of vectorizer. Finally, there is a table with all experiments ever executed in the system on the right. Each row in the table represents an experiment; the columns are dataset, start time, end time, status (not started, in progress, finished), and a button to load experiment results.

Experiment page (Appendix B, Figure 2)

Users can open the experiment page by clicking the 'load' button for the experiment of their choice on the main page. Similar to the main page, the experiment page is also split into two parts. On the left of the page is the information about the experiment, the same fields from the table on the main page. Additionally, clusterer and vectorizer with their initial hyperparameters. Average ARI (adjusted rand index) and silhouette scores across all target words are also included on the left side. We use ARI for labeled datasets to measure the clustering quality. There is a table of results for each target word on the right side. Table columns are: 'word,' 'ARI (for max silhouette score),' 'maximum ARI,' maximum silhouette score, and a set of links to visualization tools.

Dataset upload page (Appendix B, Figure 3)

The dataset upload page contains a form for uploading a dataset. The user needs to give the dataset a name, choose a zip archive with all dataset files (currently, the system supports a custom format), give a name to the corpus, and upload the second corpus with the same set of fields.

3.3. Visualizations for WSI and LSCD

The system offers four distinct methods of visualizing and working with the results of WSI tasks for semantic change detection. First, a scatter plot of contexts projects word representations into 2D, preserving the neighborhood of words from the original space as much as possible. Its purpose is to give the user signal on relative distances between usages and help identify clusters (from here and on, the word 'cluster' is used not only for the final identified clusters by the clustering algorithm but also for intermediate ones) structures at a glance. Second, the clustering dendrogram draws a dendrogram of hierarchical clustering. Its purpose is to help the user understand the intermediate state of the clustering process, explore intermediate clusters that may have distinct senses on their own, and dive deep into why specific clusters formed and some others did not. Third, the clustering overview page gives a high-level view of the numerical results of the clustering process and gives the ability to scan examples of usages based on the identified cluster per corpora. The last, inspect senses page, is the primary working tool; it gives the user ability to traverse the clustering tree up and down from identified clusters, quantifies and displays important numerical properties of each intermediate cluster, and gives the ability to manually mark clusters with senses, that way complementing the computational results with the human judgment. Now let us explore each of these tools individually in more depth.

3.3.1. Scatter plot of contexts

As mentioned above, the purpose of this page is to plot words in context in 2D. Such a plot gives the ability to see cluster formations at a glance. Hovering the mouse pointer over a point shows the context with the target word highlighted. It is possible to select a single point or multiple points. When selecting a single point (Appendix B, Figure 4), edges are drawn to the closest points in the original space, labeled with the distance value. A table with the context of the point and information about the most significant substitutes is drawn on the left panel. We show the list of substitutes sorted by the probability generated by the language model. We also show the substitutes of the 10 closest points sorted by their frequency. When multiple points are selected (Appendix B, Figure 5), we show the list of substitutes from the selected group sorted by conditional probability $P(Sub \mid Cluster) = \frac{C_{sub}}{C_{cluster}}$ where the cluster is the set of selected points, C_{sub} is the number of points containing the substitute, and $C_{cluster}$ is the total number of points in the selection. It explains the sense of the

selected collection of points using the highest probability substitutes. However, substitutes that are common across all points in the dataset, regardless of the meaning of the point, are ranked high. To solve that, we follow the approach (Zhikov et al.) and first sort by $P(Sub \mid Cluster)$ then by $PMI(Sub,\ Cluster)$, where the latter is the pointwise mutual information between the substitute and the cluster, given by $PMI(Sub \mid Cluster) = log_2(\frac{P(Sub \mid Cluster)}{P(Sub)})$. Sorting by PMI allows to rank higher the frequent substitutes in the selection and rare in the dataset as a whole; that way, the highest-ranking substitutes become more descriptive. We use the same table of substitutes in multiple other pages in the system; further, we refer to it as the 'substitutes table.' The tool provides controls to set the coloring; they can choose from coloring by the corpus (relevant for two corpora cases), gold label (relevant for labeled datasets), and predicted clusters. It also provides a choice of dimensionality reduction methods. The system supports PCA and t-SNE (van der Maaten & Hinton 2008). When t-SNE is chosen, additional parameters can be set: perplexity, number of iterations, exaggeration, number of iteration during the early exaggeration phase, and the early exaggeration coefficient. Kobak et al. 2019 demonstrated that the mentioned parameters heavily influence the results of dimensionality reduction with t-SNE. We have used the openTSNE library developed as part of their work to give the possibility to the user to fine-tune the reduction technique. Figures 1 and 2 show the scatter plot for the word 'лира,' we see clear separation into two clusters, which correspond to two senses in which the word is used - a musical instrument and a currency.

3.3.1. Clustering summary page

The system draws three plots on the top of the clustering summary page (Appendix B, Figure 6). The first is a line chart with the silhouette score on the Y-axis and the number of clusters provided as hyperparameters on the X-axis. The second is a histogram with identified clusters on the X-axis and the number of samples on the Y-axis. We draw the histogram for both corpora on the same plot in different colors. When clicking the bins on the histogram, the list of all samples within the bin and the substitutes table is loaded under the histogram. Additionally, the tool provides control to change the number of substitutes to be shown along with samples. The measure of pointwise mutual information between the corpus and the cluster, described in section 3.1.2, is also shown. The third and the last plot is a histogram of pairwise distances of word vectors within and between two corpora. It gives a convenient visual intuition into the differences in density of corpora.

3.3.2. Clustering dendrogram

Agglomerative clustering is used to group example usages of the target word. The clustering process is iterative; each sample is a single-element cluster on the first iteration. On each iteration, the two closest clusters are merged. We collect the information about each iteration, the clusters that were merged to build a tree data structure, which we visualize as a right to left dendrogram. Leaf nodes of the dendrogram are the usage examples. On the left of the page, the tool provides control to select the number of clusters. It is equivalent to selecting the number of clusters since the number of clusters equals the difference between the total number of samples and the number of iterations. The clusters are represented as vertices in the tree; we highlight clusters that form after the selected number of iterations with bigger circles. This simple capability gives a way to ‘replay’ the clustering process and explore the intermediate states. It is possible to choose the coloring based on gold, predicted labels, and corpora. Dendrogram’s view helps to see intermediate cluster formations that often correspond to senses that get lost because they are merged into other clusters at later iterations. Three selection modes are possible: cluster node selection, sample node selection, pair selection.

Cluster node selection (Appendix B, Figure 7)

Each node/vertex is either an intermediate or final cluster in the dendrogram tree. By clicking on a vertex, one can select all its leaf nodes, which are representations of example usages of the target word. The substitution table is displayed on the left panel when a node is selected, similar to the scatter plot page. Additionally, a table of nearest clusters is shown. We determine nearest clusters by using the same method that was used during the clustering process. Each node in the table is represented by the top 5 substitutes of its leaves. To determine the top substitutes for a cluster, we use the same method as the substitutions table, based on pointwise mutual information and conditional probability.

Sample node selection (Appendix B, Figure 8)

Usage node selection is similar to point selection in the scatter plot view. The same context block is shown, and additionally, the nearest nodes as in cluster node selection, as displayed.

Pair selection (Appendix B, Figure 9)

When a usage node or a cluster node is selected, the distance from the selection to every other node in the dendrogram is added to each node, as a number next to it. Clicking on a number will select the pair of these nodes. For a selected pair, we draw two tables. The first contains the distance between two selected nodes, the Jensen-Shannon distance between the probability distributions of the substitutes, and the joint ranking of substitutes. The joint ranking of substitutes shows the most descriptive substitutes for the two selections. We compute the joint rank by taking the arithmetic average of substitute ranks defined by conditional probability and PMI, similar to how we do it in the substitutions table. We show next to each substitute the joint rank and the individual ranks; this helps understand why specific nodes are close or far from each other. The second table we draw has been proposed by Arefyev et al. 2019. They refer to it as discriminative substitutes. We show two sets of 10 substitutes; the first set is the list of substitutes with high probabilities in the first cluster and low in the second; the second set is the opposite. Discriminative substitutes give a good sense of differentiating substitutes between two selections.

The dendrogram view is specifically valuable for researching the clustering process; it helps to identify merges of clusters that are suspicious, it uncovers information about too early or too late merges, and it is a powerful general-purpose tool for exploring intermediate states of hierarchical clustering. However, when the sample size is large, the tree view cannot fit and be explored on any reasonable size screen. Furthermore, only seeing small fragments of the tree does not give the desired signal, so the primary use case of this tool is research and troubleshooting on small samples (up to 200).

3.3.2. Inspecting senses

The tools we have covered so far are designed to give insights into the results of the vectorization and clustering processes on varying levels. The scatter plot helps understand how meaningful are the vector representations and the method of measuring the distance between them. The clustering summary gives a high-level view of the numerical results of the clustering, differences between corpora and provides a window into exploring the samples conveniently. Dendrogram helps understand the intermediate states of the clustering process and gives hints about possible lost clusters. As we can see, all of these tools take the computational

results and give visual methods and metrics to understand them better. However, as mentioned earlier, the aim is to combine computational results with human judgment for semantic change detection. The goal of the sense inspection tool (Appendix B, Figure 10) is precisely that. The sense inspection tool is the central instrument for semantic change detection tasks. The tool is built on top of the clustering tree, similar to the dendrogram, but instead of drawing the complete tree, it shows only three levels. In the center is the selected cluster, on the right is the parent cluster. The selection is merged into the parent later in the clustering process. The child clusters that merged into the selected cluster in the earlier stages are on the left. Each node is represented by a 'cluster block,' which contains information about the node.

Cluster blocks are representations of clusters identified by the clustering algorithm. They contain the following set of data points:

- **Signature:** top substitutes determined by conditional probability and PMI (similar to substitutes table in scatter plot view)
- **ID:** a unique identifier for the cluster
- **Total samples:** the number of samples within the cluster (same as the number of leaf nodes in the dendrogram)
- **Iteration:** the iteration when the cluster formed
- **Corpora:** number of samples in each corpus (relevant for the two corpora cases)
- **PMI:** pointwise mutual information between the corpora and the cluster gives a signal about how much more the samples in the cluster associate with each corpus; a detailed description is covered in section 3.1.2 (relevant for the two corpora cases)
- **Class counts:** number of samples of each class within the cluster (relevant for labeled case)
- **Entropy:** Shannon entropy, given by $H(Cluster) = -\sum_{i=1}^{n} P(x_i) log_2(P(x_i))$ where n is the number of classes in the cluster, x_i is the i-th class. It signals the homogeneity of the cluster concerning gold classes of contained samples. (relevant for labeled case)

Clicking on a signature within the cluster block selects the cluster, redrawing the parent and child nodes. That way, the user can navigate the whole clustering tree.

The system makes a distinction between clusters and **senses**. While the computation comes up with clusters that presumably represent various senses, the final call of what cluster represents that sense is left to the user. The tool offers the possibility to assign senses to cluster blocks manually. When assigning a sense to a cluster, it propagates down the clustering tree, assigning the same value to all child clusters and leaves. A summary of all assigned senses and the total number of samples without assigned senses is shown on the left of the page. In addition, the tool provides filtering functionality by senses.

Sense blocks are representations of senses identified by the user. They contain the following set of data points:

- **Sense name:** a free form name for the sense, defined by the user
- **Signature:** top substitutes determined by conditional probability and PMI (similar to substitutes table in scatter plot view) for all samples that have the given sense assigned to them
- **Total Samples:** total number of samples that have the given sense assigned to them
- **Corpora:** number of samples that have the given sense assigned to them, breakdown by corpus (relevant for the two corpora cases)
- **PMI:** pointwise mutual information between the each corpus and the sense given by $PMI(Corpus, Sense) = log_2(\frac{P(Corpus \mid Sense)}{P(Corpus)})$, similar to PMI between the corpora and clusters, gives a signal about how much more the samples that have the given sense assigned to them associate with each corpus, detailed description is covered in the section 3.1.2 (relevant for the two corpora case)
- **Class counts:** number of samples of each class with given sense assigned to them (relevant for labeled case)
- **Entropy:** Shannon entropy concerning classes of samples that have the given sense assigned to them. (relevant for labeled case)

Additionally, each sense cluster contains a visualization indicating the ‘degree of belonging’ to one of the other corpora. It is represented as a horizontal bar, split into left and right parts. The left part, in red, represents the old corpora, the right part, in blue, represents the new corpora. The width of each part is determined by the $PMI(Corpus \mid Sense)$, proportional to $k^{PMI(Corpus \mid Sense)}$ for the left part and

$k^{-PMI(Corpus \mid Sense)}$ for the right part, where k is an arbitrary constant (set to 1.5 to achieve good visual presentation).

For labeled datasets, two types of **'interesting' merges** are defined. First are the ones which increased the average Shannon entropy. We measure the entropy for both child clusters, take the arithmetic average and compare it to the selected cluster's entropy. If the selected node's entropy is bigger than the average entropy of its children, we add an 'increased entropy' tag. The intuition is that the merge resulted in a more noisy cluster than we have already had; it is a good reason to think that the merge should not have happened and is worth investigating. The second type of interesting merges we define as 'imbalanced merge.' We check the number of samples within each child node, and if the child with a more significant number of samples has more than five (the choice is arbitrary, the number can be easily configured) times more samples than the second child, we add the 'imbalanced merge' tag to the selected cluster. Again, the intuition is that if one of the child clusters has grown independently from the other one and the other one did not merge into any other cluster till the late stages of the clustering process, it is likely that those represent distinct meanings, worth investigating.

As already discussed (3.1.2), the pointwise mutual information between clusters and corpora measures the association between the same. High-PMI clusters are interesting for investigation because they contain a skewed distribution of samples per corpora. The system tags clusters with the absolute value of PMI between the cluster and any of the corpora higher than 2 (the value is arbitrary, it proved to be reasonable on the corpora the system was tested on) with 'High PMI' tag. It is possible to filter clusters using that tag as well.

Another type of interesting merge is the 'late merges.' Those are the cluster merges that happened at the late stages of the clustering process. While imbalanced merges are interesting at any stage, merges during the last iterations of the clustering may be 'forced.' The agglomerative clustering algorithm keeps merging clusters until it reaches the predefined number of clusters. The tool provides filtering functionality by tags as well.

3.3. The procedure of sense inspection

The user can review and correct two types of clustering errors using the sense inspection tool (3.3.2). The first is 'incorrect merges' when two clusters with different senses are merged. The second is 'absence of merge' when two clusters of the

same sense are not merged before the clustering process terminates. The following procedure can be used to address both types of errors:

- Starting from the automatically identified clusters, the user traverses (depth-first) the clustering tree down, prioritizing clusters with signatures that do not make immediate logical sense. For example, cluster signature *'motorcycle, truck, plane, driver, train, engine, horse, bus, wagon, passenger'* can be recognized by the user as a means of transportation, while *'machine, box, oven, tower, pipe, installation, board, horse, camera, engine'* is harder to interpret. The core idea is that with each step down the clustering tree, the ambiguity of cluster senses reduces. For example, a cluster with an ambiguous two-substitute-signature for the word 'plane,' 'surface, fly' may have child clusters with signatures 'airport, travel' and 'perpendicular, dimension,' corresponding to non-ambiguous senses 'aircraft' and 'geometric concept.'
- The user checks a random sample of example usages for all clusters with meaningful signatures to confirm that the cluster signature is descriptive of the cluster content. More samples the user chooses to review, the higher the confidence that the cluster represents a valid sense. In case of inconsistencies, the user keeps traversing the cluster down by clicking the signatures on the child nodes.
- When the user is confident about the sense of a cluster, they assign a sense to it using the text field on the cluster block and the 'set sense' button (if the sense was already assigned to another cluster, the user could choose it from a dropdown list in the cluster block).
- When reaching the leaf nodes, they assign senses to them; if they are ambiguous, the user can add the best describing sense with a mention that it is still ambiguous, or assign a sense '[ambiguous]' or '[unclear].'
- Next, the user can either return to the root of the clustering tree and repeat the process or navigate up the tree until they reach the first node without an assigned sense and from there repeat the process.

As a result of this procedure, a set of sense blocks will form and will be displayed on the right panel in the *Sense Inspection* tool with relevant metrics to make conclusions whether the semantic change has happened or not. Empirically, based on the experimentation with the pre and post-soviet subcorpora of the Russian National Corpus, the absolute value of PMI between a sense and a given corpus greater than 3 is a reasonably reliable signal to associate (or negatively associate in

case of negative PMI value) the sense with the corpus, given the number of samples within a sense is greater than 30.

The effort required for this procedure varies based on the quality of clustering that the algorithm performed and the size of the dataset. Nevertheless, even if partially done, the sense inspection procedure still improves the quality of identified senses. Thus the user is free to make a judgment call on how much effort they want to invest in completing the sense inspection and correction.

Mentioned clustering errors result in ‘scattered’ senses when the examples of certain senses are not grouped together in a single cluster but spread across many clusters. Without reviewing all examples, there is no way to identify all senses with absolute confidence. Even when doing so, it is not guaranteed that the user will spot all senses, simply because some examples of uses are ambiguous for humans, just like for machines. However, if there is a cluster with a clear sense and a significant number of examples, then the likelihood of examples in the dataset with the same sense, not included in the cluster, is low. Often these are examples that are ambiguous to humans as well.

4. Experiments

4.1. Experimental setup and the process

As for experimentation, the subcorpora of the Russian National Corpus was used. The dataset consists of pre-soviet (approx. 5.2M sentences), soviet (approx. 8M sentences), and post (approx. 6.5M sentences) parts; as targets for experimentation, a set of words from the book ‘Two centuries in twenty words’ (Dobrushina et al. 2016) were used. For each target word, two tasks were performed. First, identify senses and link them to senses identified in the book. Second, conclude whether the target word has changed its sense between two corpora. For the first task, recall is used as a metric to avoid penalizing the identification of senses that are not covered in the reference book. For the second task, the proportion of words with matching conclusions on semantic change in the book.

The process used for identifying senses is the following:

- Take a random sample of up to 1000 usages from each corpus for a given target word.
- Execute the WSI task for using the system.
- Review the results using the visualization tools.
- Using the sense inspection tool and procedure, identify senses.

- Make conclusions about the semantic change according to 3.1.2.

Not all clusters were reviewed; in the results, one may notice that only a portion of all examples is covered. The reason for this is that the amount of human effort should be limited, so only clusters with a signature that contained some understandable theme with more than 30 examples were reviewed, as well as clusters that were formed at late stages of the clustering process, meaning at 3-4 levels down the dendrogram root.

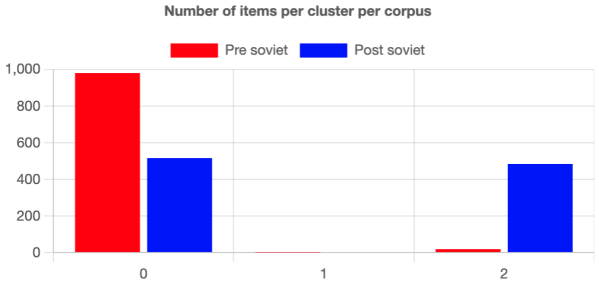
The time periods that the reference book (Dobrushina et al. 2016) used to describe the usage of senses were split into pre-Soviet time, before 1922 and after (as close as possible), mapped to pre and post-soviet corpora. Thus words that were identified as new, the claim is not that they necessarily acquired the new sense in post-Soviet time, but rather in Soviet or post-Soviet time.

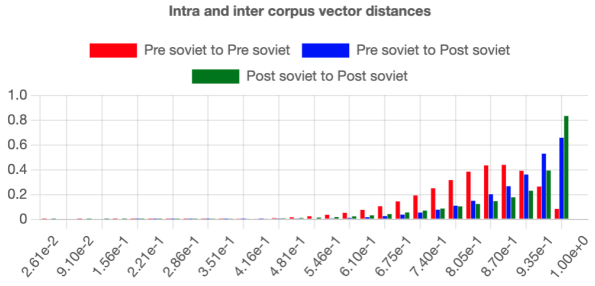
4.2. Results

A detailed analysis of results for Russian words ‘пакет’, ‘свалка’ and ‘машина’ are shared in this section, the remaining results can be found in Appendix C. The WSI task was executed for all words, and the sense inspection procedure described in 3.3.2 was applied. The criteria for sense change were used according to the description in 3.1.2. The manual sense inspection was operated by a person with limited knowledge of the Russian language.

Results for ‘пакет’

The total number of examples used was 2000, a thousand from each corpus. The WSI algorithm (3.1) identified three clusters.





Signature	бумага, коробка, мешок, папка, карточка, конверт, сумка, бутылка, ключ, портфель
Samples	Total: 1496 Pre-Soviet: 980 Post-Soviet: 516
Iteration	1998

PMI	Pre-Soviet: 0.3897434791149338 Post soviet: -0.5356672044643827

Signature	моралес, моралеса, рудольф, робин, брю, шоумен, бланко, сидоров, сидор, домина
Samples	Total: 1 Pre-Soviet: 1 Post-Soviet: 0
Iteration	0
PMI	Pre-Soviet: 0.3897434791149338 Post soviet: -0.5356672044643827

Signature	блок, портфель, остаток, раздел, доля, актив, голос, частный, лицензия, половина
Samples	Total: 503 Pre-Soviet: 19 Post-Soviet: 484
Iteration	1997
PMI	Pre-Soviet: -3.7264870763619413 Post soviet: 0.9444486474690673

The second identified cluster had only one example, and the word ‘пакет’ was used as a personal name, a rare use case. Nevertheless, the signature of the cluster, consisting of names, did make logical sense for that example.

The first cluster had a somewhat interpretable signature, but not quite clearly, it seemed related to the paper container, but the word ‘ключ’ did not seem to be aligned with that pattern.

The third cluster did not yield an interpretable signature, but it had a high negative PMI for pre-Soviet corpus, which may have meant a specific pattern related to when the word was used.

The manual procedure was done according to 3.3 for each of two clusters with 1496 and 503 examples.

Final results for ‘пакет’

Sense	***Source***	***Description***	***Semantic change***
Sense 1	Pred	“По каждому виду облигаций, на основе **пакета** Excel, было построено по пять кривых роста.” (post-Soviet) “Например, в России "благодаря" им **пакет** Microsoft Office практически безальтернативен.” (post-Soviet)	

		“Компьютерную обработку данных проводили на компьютере "Macintosh" с помощью **пакета** программ "DNAstar" ("Lasergen").” (post-Soviet)	
	Gold	No link	
	Pred	**Signature:** модуль, компонент, приложение, сервис, библиотека, инструмент, конфигурация, набор, интерфейс, драйвер **Description:** Software package/library **Total samples:** 36 (1.8% of all examples) **Corpora:** Post-Soviet: 36	Gold: New sense
	Gold	No link	No link
Sense 2	Pred	“При появлении очередного **пакета** с данными до момента срабатывания таймера пакет сохраняется в буферном накопителе и ожидает времени срабатывания таймера.” (post-Soviet) “Главными приоритетами этого проекта является решение проблем взаимодействия между сетями с маршрутизацией **пакетов** и сетями с коммутацией каналов.” (post-Soviet) “**Пакет** уничтожается при переполнении буферного накопителя, когда скорость поступления пакетов из локальной вычислительной сети объекта превышает скорость, с которой данные передаются в канал связи.” (post-Soviet)	
	Gold	No link	
	Pred	**Signature:** соединение, блок, модуль, компонент, код, файл, протокол, образец, диск, сигнал **Description:** Network packet, in telecommunications **Total samples: 1**6 (0.8% of all examples) **Corpora:** Post-Soviet: 16 **Num clusters:** 1	Gold: New sense
	Gold	No link	No link
Sense 3	Pred	“Почтальон при вручении повестки получает на чай гривенник и по доставке **пакета** на дом -- двугривенный.” (pre-Soviet) “На карту наезжала кипа каких-то бумаг, среди которых виднелись несколько телеграмм и два или три **пакета**, запечатанных красным сургучом.” (post-Soviet) “-- Отослан, Игнатий Федорович: ведь все казенные **пакеты** надписываю и отправляю на почту я, -- это по моей части.” (pre-Soviet)	

	Gold	“Наконец, Бельтов снял **пакет** и стал читать письмо; с каждой строчкой его лицо делалось бледнее, и слезы навернулись на глазах его. [А.И. Герцен. Кто виноват? (1841–1846)]” “Графиня с беспокойством развернула **пакет**, прочла несколько строк, — руки ее затряслись, она побледнела. [В.Ф. Одоевский. Кос- морама (1837)]” “Прилагаю **пакет** на имя вашего сиятельства, переданный мне консулом в Занте. [А.Я. Италинский. Документы (1815)]”	
	Pred	**Signature:** билет, бумага, подпись, конверт, карточка, папка, объявление, приглашение, послание, печать **Description:** Sealed paper parcel **Total samples:** 61 (3% of all examples) **Corpora:** Pre soviet: 56, Post soviet: 5 **Num clusters:** 1	- Neither lost nor new - Frequency decreased over time
	Gold	Письмо, конверт, почтовое отправление	- Neither lost nor new - Frequency decreased over time
Sense 4	Pred	“Их внешний вид, форма и объем **пакета** говорят о солидности структуры.” (post-Soviet) “На столе моем стоял сахар в бумажном **пакете** со штемпелем купца Синебрюхова, и сквозь монотонное чтение я услышал шорох в пакете.” (pre-Soviet) “Дверь приоткрылась, и в палату вошла Вероника, держа наполненный пластиковый **пакет**.” (post-Soviet)	
	Gold	“ Вспомните, как удобно хранить продукты в полиэтиленовых **пакетах**. [Я. Муравин. Тысяча первая профессия полимеров // Химия и жизнь (1965. No 7–8)]” “Лицо и голову Прутика закрывал прозрачный **пакет**…” [Ю.М. Дружков. Приключения Карандаша и Самоделкина (1964)] “...банки со сгущенным молоком, хлеб, яблоки, дыни — это на обед, русалочий хвост и целлофановый **пакет** с парфюмерией. [Н. Серпкова, В. Су- етин. Трое в одной лодке, не считая русалки (записки кинолюбите-лей) // Спортсмен-подводник (1965. Вып. 11)]”	
	Pred	**Signature:** коробка, мешок, бумага, бутылка, сумка, ручка, папка, портфель, карточка, замок **Description:** Container either a box or a bag, often made out of paper or plastic **Total samples:** 1111 (55% of all examples)	- Neither lost nor new - Stable over time

		Corpora: Pre soviet: 639, Post soviet: 472 **Num clusters:** 1	
	Gold	1. Мешок, емкость, в том числе, целлофановый, пластиковый пакет 2. Картонные пакеты для жидких продуктов *Not an exact match, prediction has broader sense*	- New sense
Sense 5	Pred	“Итого, во владение иностранцев переходит контрольный **пакет** "Юкоса".” (post-Soviet) “Исходя из этого цена реализованного **пакета** составляет 125 млн долл.” (post-Soviet) “На покупку было потрачено 295 млн долларов, а выручка от продажи части этого **пакета** составила около 555 млн долларов..” (post-Soviet)	
	Gold	“Крупные промышленники и банки, стиснув зубы, держались за **пакеты** акций. [А.Н. Толстой. Гиперболоид инженера Гарина (1925–1927)]” “...концентрирующих в своих руках в целях спекуляции контрольные **пакеты** акций самых разнообразных предприятий. [П. Лапинский. Герои гнили в борьбе с страной строящегося социализма (1930) // Известия (1930. 4 декабря)]”	
	Pred	**Signature:** портфель, доля, голос, остаток, актив, участок, выкуп, процент, залог, половина **Description:** Portfolio of valuable assets **Total samples:** 215 (10% of all examples) **Corpora:** Pre soviet: 1 (unrelated), Post soviet: 214 **Num clusters:** 1	- New sense
	Gold	Пакет акций	- New sense
Sense 6	Pred	“Оклад от 600 у.е. + проценты + социальный **пакет**.” (post-Soviet) “**Пакет** льгот и скидок действует только сегодня.” (post-Soviet) “При аренде автомобиля класса "эконом" на двух человек весь **пакет** услуг стоит $329, при заказе "компакта" -- $359.” (post-Soviet)	
	Gold	“**Пакет** льгот и скидок действует только сегодня. [Столица (1997. 2 сентября. No 16)]” “Итак, у нас получился довольно увесистый **«пакет»** идей. [Работница (1988. No 10)]” “Речь шла не о программе, а о **пакете** наших предложений. **Пакет** предложений был выработан и размножен в тех масштабах, которые были доступны, это порядка 600–700 экземпляров. [Горизонт (1989. No	

		8)]”	
	Pred	**Signature:** бонус, тариф, скидка, карточка, билет, сервис, премия, сертификат, лицензия, купон **Description:** Collection/package of services or plans, e.g. compensation package, internet services package. **Total samples:** 30 (1.5% of all examples) **Corpora:** Pre soviet: 3 (unrelated), Post soviet: 27 **Num clusters:** 1	- New sense
	Gold	‘Наборы’ (‘нематериальная совокупность’) *Not an exact match, prediction has a narrower sense*	- New sense
Sense 7	Pred	Not identified	
	Gold	“Как только начнется лекция, он вытащит из-за стола этот **пакет** с угощением и пустит его по рукам товарищей, но так, чтобы пакет передавался от одного к другому на виду у всех, высоко над столом. [Ф.И. Буслаев. Мои воспоминания (1897)]” “Стол покрылся свежею салфеткою, и на нем появился кипящий самовар, серебряные ножи и вилки и несколько аккуратно свернутых **пакетов**, в которых оказались: индейка, язык, ватрушки и пакетик с солью. [А.А. Фет. Вне моды (1889)]”	
	Pred	Not identified	N/A
	Gold	Упаковка, сверток	- Stable sense
Sense 8	Pred	Not identified	
	Gold	“...сливавшиеся в мутную полосу чугунные валы глотать добела раскаленные **пакеты** сварочного железа и выплевывать их обратно гнувшимися под собственною тяжестью ярко-красными железными полосами. [Д.Н. Мамин- Сибиряк. Три конца (1890)]” (pre-Soviet)	
	Pred	Not identified	N/A
	Gold	Пакет железа *Rare sense*	- Neither lost nor new - Frequency decreased over time
Sense 9	Pred	Not identified	

	Gold	— Что это у вас на шее? — показал атаман рукой на продолговатый **пакет** из лосиной кожи, висевший на шее Рудольфа Карловича. — Это... Это ладанка... [В.В. Курицын. Томские трущобы (1906)]	
	Pred	Not identified	N/A
	Gold	Ёмкость, тара *Rare sense*	- The sense appeared around the time when Soviet Union was founded, over time it evolved into “bag” - Lost sense
Scores		Sense identification recall: 0.571 % of correctly identified lexical semantic change patterns: 75%	

After manual inspection of clusters, two senses were discovered that were not covered in the reference book. It is unclear why these senses were not present in the book; one possible reason is the explicit emphasis on texts of the 19th century that the authors mention in the introduction. Nevertheless, the system generated an intermediate cluster with signature ‘модуль, компонент, файл, сервис, код, блок, приложение, протокол, сервер, конфигурация’, with two child clusters with signatures ‘модуль, компонент, приложение, сервис, библиотека, инструмент, конфигурация, набор, интерфейс, драйвер’ and ‘соединение, блок, модуль, компонент, код, файл, протокол, образец, диск, сигнал’, that unambiguously corresponded to ‘computer software library’ and ‘computer network’ senses. The number of examples for both is relatively small, but given prior knowledge that computer and telecommunications technologies came to popularity late-Soviet and post-Soviet times, one can conclude that these are new senses.

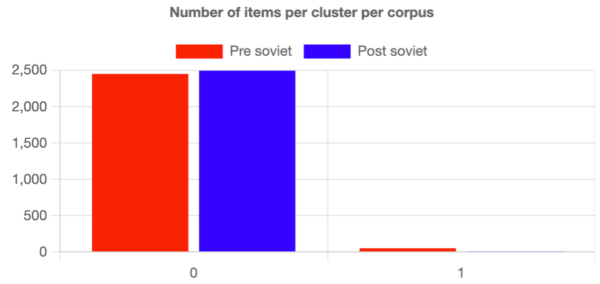
Three senses covered in the reference book were not discovered. ‘Упаковка, сверток,’ ‘Ёмкость, тара’ and ‘Пакет железа’, according to the authors, the latter two are rare senses; however ‘Упаковка, сверток’ had 14% of all usages between 1990 and 1999 years, according to the authors. The system most likely was not able to distinguish it from the rest of the example in the sense with signature ‘коробка, мешок, бумага, бутылка, сумка, ручка, папка, портфель, карточка, замок’, just like it could not distinguish between ‘Мешок, емкость, в том числе, целлофановый, пластиковый пакет’ and ‘Картонные пакеты для жидких продуктов’. The difference is that of the latter; it was possible to see some examples

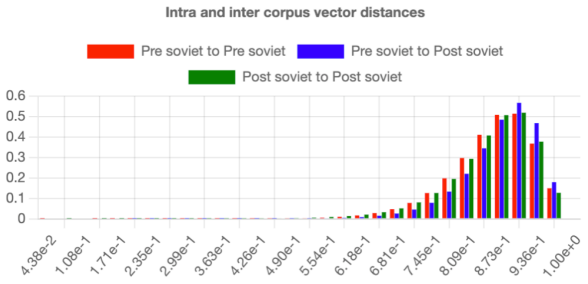
of both gold senses with the large cluster by randomly sampling 30 items, which was not the case for the former.

The largest identified sense with signature ‘коробка, мешок, бумага, бутылка, сумка, ручка, папка, портфель, карточка, замок’ did not show the exact change in meaning over time as the reference. The reason for that is two-fold. First, many of the examples of the sense ‘Письмо, конверт, почтовое отправление’ were included in the same cluster, because of clustering error. These instances were not spotted during the manual inspection due to limited knowledge of the language and brevity of example uses.

Results for ‘машина’

The total number of examples used was 5000. 2500 from each corpus. The WSI algorithm (3.1) identified two clusters.





Signature	мотоцикл, двигатель, машинка, техника, лошадь, самолёт, водитель, собака, поезд, колесо
Samples	Total: 4943 Pre soviet: 2449 Post soviet: 2494
Iteration	4999
PMI	Pre-Soviet: -0.013194132575794715 Post soviet: 0.013074558689897153

Signature	белый, старый, мужской, любить, жена, малый, быть, русский, мой, женский
Samples	Total: 57 Pre soviet: 51 Post soviet: 6
Iteration	4998
PMI	Pre-Soviet: 0.8395353278067539 Post-Soviet: -2.2479275134435857

The first cluster had a theme that immediately stood out - transportation. However, with significantly fewer examples, the second cluster was harder to interpret. The

cluster inspection showed that all samples (except a few with the sense ‘mechanical apparatus’) had the Russian female name ‘Маша’ instead of the target word ‘Машина.’ While these are not the same lemma, they share certain word forms. One of the examples in the cluster was ‘Я педагог, и здесь в доме свой человек, Машин муж…’ which contains the word ‘Машин’ in the sense of the female name. The same word in a different context, ‘...на улице много машин…’ has the sense of automobile.

The second cluster contained the majority of examples; it had a clear, understandable signature but a broad sense.

Final results for ‘машина’

Sense 1	Pred	“Самостоятельно производить ремонт и устранять неисправности в пишущей **машине** запрещается.” (post-Soviet) “Теплота, развиваемая сгорающим топливом, превращается чрез посредство пара в механическую работу **машины**.” (pre-Soviet) “Припуск на шов определяем для себя по ширине лапки нашей швейной **машины**.” (post-Soviet) “О молотильной машине В Земледельческом журнале 1830 года, 29 помещена была статья г-на Флата о молотильных **машинах**.” (pre-Soviet)	
	Gold	“Полюби какого-нибудь человека с состоянием, он тебе купит швейную **машину**. [А.Ф. Писемский. Просвещенное время (1875)]” “Вот, Танечка, — говорила она, указывая на свою стиральную **машину**, — какое подспорье эта машина. [Т.Л. Сухотина-Толстая. Друзья Ясной Поляны (1908–1917)]”	
	Pred	**Signature:** инструмент, двигатель, лошадь, механик, машинка, механизм, электроника, техника, прибор, печь **Description:** Mechanical and/or electrical apparatus built for certain tasks, printing machine, coffee machine, etc. **Total samples:** 352 (7% of all examples) **Corpora:** Pre soviet: 318, Post soviet: 34 Num clusters: 5	- Neither lost nor a new sense - The frequency declined over time
	Gold	Бытовой прибор *Narrower than predicted sense*	- Neither lost nor a new sense - The frequency declined over time

Sense 2	Pred	“Во время пробега очень немногие **машины** имели поломки, которые тут же в пути исправлялись.” (pre-Soviet) “**Машины** затормозили, развернулись резко и умчались.” (post-Soviet) “- У нас хоть и нет **машин**, а тоже гонят изо всех мест” (pre-Soviet) “Пассажиры садились вокруг **машины**, иные на лавках, иные на дровах.” (pre-Soviet)	
	Gold	“В 10 час. поехал в Петербург и посетил автомобильную выставку в Михайлов [ском] манеже. Более 140 различных фирм прислали свои **машины**. [Нико- лай II. Дневники (1913–1916)]” “Ведь у вас все есть: дом, дача, **машина**. [А.В. Вампилов. Прощание в июне (1964)]”	
	Pred	**Signature:** водитель, автобус, багажник, мотоцикл, грузовик, пассажир, палатка, вагон, поезд, кресло **Description:** Automobile, car **Total samples:** 691 (13.8% of all examples) **Corpora:** Pre soviet: 78 (including many examples with sense ‘train’), Post soviet: 613 Num clusters: 5	- Neither lost nor a new sense - The frequency increased over time
	Gold	Автомобиль	- Neither lost nor a new sense - The frequency increased over time
Sense 3	Pred	“Планомерному развитию механизации мешало отсутствие системы **машин** и продуманной агротехники и технологии производства.” (post-Soviet) “Дай-чен-мо командирован в Россию для ознакомления со способами добывания и очистки нефти и металлов и для закупки необходимых для этой цели машин.” (pre-Soviet) “Инженерно-технический работник по надзору за безопасной эксплуатацией грузоподъемных **машин** должен работать по плану, утвержденному должностным лицом, которому он подчинен.” (post-Soviet)	
	Gold	Not linked	
	Pred	**Signature:** агрегат, прибор, инструмент, механизм, аппарат, станок, вагон, установка, трактор, двигатель **Description:** Large, tractor-like machine for	- Neither lost nor a new sense - The frequency increased over time

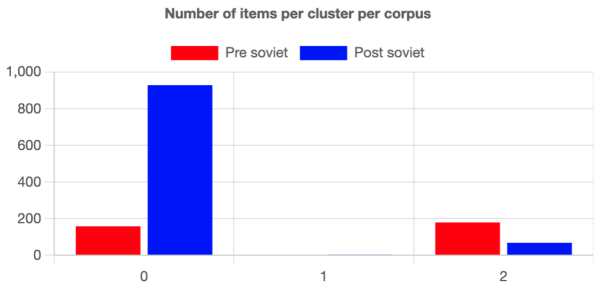
		construction and/or agriculture and similar activities, e.g. crain, cultivator, etc. **Total samples:** 57 (1.1% of all examples) **Corpora:** Pre soviet: 7, Post soviet: 50 Num clusters: 1	
	Gold	Not linked	- Neither lost nor a new sense - The frequency increased over time
Sense 4	Pred	Not identified	
	Gold	“На железную дорогу, а поспеешь к **машине**, так еще сторублевую! [Ф.М. Достоевский. Идиот (1869)]” “Сей грозный исполин, пыша пла- менем, дымом и кипящими брызгами, двинулся вперед... Стоявшие по сторонам дороги зрители изумлялись, видя величественное, ровное, легкое, притом скорое движение **машины**. [Северная пчела (1836. 6 ноября)]”	
	Pred	Not identified	- Lost sense
	Gold	Поезд, паровоз	N/A
Sense 5	Pred	Not identified	
	Gold	“Я, ужаснувшись, дал себе зарок и продал эту **машину**, фривольно называемую пи-си, за полцены. [Н.Ю. Климонтович. Последняя газета (1997–1999)]” “В отличие от компьютеров Commodore у **машин** Atari имеется встроенная программа тестирования, позволяющая проверить работоспособность компьютера. [А. Щедрин. Commodore, Atari и дру- гие... // Техника — молодежи (1991. No 3)]”	
	Pred	Not identified	- New sense
	Gold	Компьютер	N/A
Scores		Sense identification recall: 0.4 (sense 1 is much broader than the linked sense, counting as linked) % of correctly identified lexical semantic change patterns: 100%	

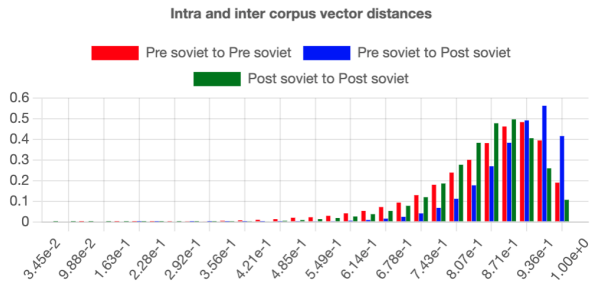
Results for the word 'машина' were significantly harder to interpret. Almost all clusters with over 100 examples had signatures with at least two meanings - some apparatus and transport. It signaled that the language model could not distinguish

from the context in what sense the word was used. Cluster inspection showed that the word 'машина' has a vast range of various uses. For example - ‘Чудовищный зверь-машина.’, ‘Везде, где могут действовать машины, англичане превосходны.’, ‘Вы смотрите на людей, как на машину.’, ‘Сама машина есть явление духа, момент в его пути.’. Many of the examples were highly ambiguous, hard to interpret (‘В России машина может сыграть совсем иную роль, может стать орудием русского духа.’). After filtering for clusters with over 50 examples and high PMI value (meaning the majority of examples from one corpus), some patterns became visible. For example, several clusters with the sense' automobile', some with the sense of 'mechanism/apparatus for a certain job.' The sense of 'automobile' is spread across almost all clusters with sizes larger than 30 examples. Authors in 'Two Centuries in Twenty Words' (Dobrushina et al. 2016) discuss the wide variety of meanings and ways the word is used in their work; in that sense, our findings align with the findings in the reference book.

Results for ‘свалка’

The total number of examples used was 1337. One thousand were from post-Soviet and 337 from pre-Soviet. The reason is that the corpora for experimentation had only 337 examples in the pre-Soviet part. The WSI algorithm (3.1) identified three clusters.





Signature	площадка, забор, шахта, могила, баня, деревня, пустыня, дача, склад, ферма
Samples	Total: 1086 Pre soviet: 158 Post soviet: 928
Iteration	1335
PMI	Pre-Soviet: -0.7928486707157645 Post-Soviet: 0.1921720727544764

Signature	драка, пожар, взрыв, охота, битва, авария, бойня, атака, буря, паника
Samples	Total: 247

	Pre soviet: 179 Post soviet: 68
Iteration	1334
PMI	Pre-Soviet: 1.5236275145910727 Post-Soviet: -1.4419049249030724

Signature	др., прочий, мир, место, бар, город, клуб, соседний, ресторан, дом
Samples	Total: 4 Pre soviet: 0 Post soviet: 4
Iteration	1303
PMI	Pre-Soviet: -Infty, Post-Soviet: 0.41899946543126576

The first and the most significant cluster had an ambiguous theme. Nevertheless, one can interpret that it refers to a specific place. The second cluster, with 247 examples, has a less ambiguous signature; it can be interpreted as 'fight,' 'battle.' Finally, in the third identified cluster with only 4 examples, three were used as named entities, in particular, two as to names for 'nightclubs,' which explains part of the signature. The remaining example described a place in or near the city.

Final results for ‘свалка’

Sense 1	Pred	“Он пришёл к месту городских **свалок**, к оврагу, где рылся с дедушкой Еремеем..” (pre-Soviet) “В Измайлове отличные развалы, рядом полно **свалок** с не нужным обывателям барахлом..” (post-Soviet) “И не нужно будет больше мучиться с вонючими **свалками** и мусоросжигательными заводами.” (post-Soviet)	
	Gold	“Дряхлый капиталистический мир запутался в собственных противоречиях. Обреченный сойти на **свалку** истории, он ищет выхода в новых безумных авантюрах. [Советское искусство (1950. No 25 (1217))]” “...сам напился до такой степени, что более ничего не помнит, и очнулся только вчера за городскими **свалками**, и в виде совершенно голом. [В.М. Дорошевич. Дело о людоедстве (1900)]”	
	Pred	**Signature:** забор, площадка, склад, шахта, полигон, баня, дача, могила, завод, ферма **Description:** Dump, landfill, most often for rubbish **Total samples:** 536 (40% of all examples)	- Neither new nor lost - Frequency increased over time

		Corpora: Pre soviet: 20, Post soviet: 516 Num clusters: 6	
	Gold	Место для сбора мусора, помойка	- Neither new nor lost (appeared around 1900s right before Soviet Union was formed) - Frequency increased over time
Sense 2	Pred	“Длительная, раскаляющаяся ссора перейдет в **свалку**.” (post-Soviet) “...Вот тут-то и началась безумная **свалка** фотографов и кинооператоров.” (post-Soviet) “Описывать это сражение здесь неуместно, да и не умел бы этого сделать, ибо не видал подробностей кровавой **свалки**.” (pre-Soviet) “Над пособием, выданным от правительства, кипела жестокая **свалка**.” (pre-Soviet)	
	Gold	“Началась жестокая **свалка**. Как ни был мал горбун, однакож с ним не так легко было справиться: за слабостью рук он лихо защищался зубами;... [Д.В. Григорович. Переселенцы (1855–1856)]” “Отчаянные татары, сломленные, низверженные сверху стен и башен, стояли твердым оплотом в улицах, секлись са- блями, схватывались за руки с россиянами, резались ножами в ужасной **свалке**. [Н.М. Карамзин. История государства Российского. Т. VIII (1815–1820)]” “— Утром была **свалка** на базаре между народом и сарбазами... [П.И. Огородников. На пути в Персию и прикаспийские провинции ее (1873)]”	
	Pred	**Signature:** драка, пожар, взрыв, битва, охота, авария, бойня, атака, буря, паника **Description:** Fight, battle, mainly chaotic, disorganized, without rules **Total samples:** 239 (17.8% of all examples) **Corpora:** Pre soviet: 178, Post soviet: 61 Num clusters: 1	- Neither lost nor new sense - Frequency declined over time
	Gold	1. Битва 2. Драка	- Neither lost nor new sense - Frequency declined over time
Sense 3	Pred	Not identified	

	Gold	“Из переполненного толпой коридора я попал в буфет. Там была давка и **свалка** у прилавка. В укромном уголке я натолкнулся на Каменева, впопыхах глотающего чай: — Ну что же, стало быть, вы одни собираетесь нами править? [Н.Н. Суханов. Записки о революции. Кн. VII (1918–1921)]” “Оркестр наполнен был при сочинителе музыки из охотников, то есть из людей благородных, которые скрыпок настроить не умеют. **Свалка** была большая. Полиция всемерно трудилась ввести четыреста человек в такие стены, в которых может войти только триста, и, казалось бы, ведь нельзя? [И.М. Долго- руков. Повесть о рождении моем... Ч. IV (1791–1798)]”	
	Pred	Not identified	N/A
	Gold	Толпа, давка	- Neither lost nor new sense - Frequency declined over time
Sense 4	Pred	Not identified	
	Gold	“**Свалки**, перемывки, выработки, ширфы, канавы, кучи песку и галек — все это напоминало издали работу сумасшедшего, который не стеснялся осуществлением своих диких фантазий и то, что вы- рывал в одном месте, сваливал в другом. [Д.Н. Мамин-Сибиряк. Золотуха (1883)]” “— Сказывают, старую **свалку** стали промывать, так, слышь, со ста пудов песку по золотнику падает.”	
	Pred	Not identified	N/A
	Gold	Отработанная руда, груда отработанных отходов	- Neither lost nor new sense - Frequency declined over time
Sense 5	Pred	Not identified	
	Gold	“Он и на фабрику ходит: сядет на **свалку** дров и глядит на меня, как я дрова ношу. [Д.Н. Мамин-Сибиряк. Три конца (1890)]” “...Пролетел кусок левого берега — пристанями, пароходными тру- бами и нечистою **свалкою** пенькой набитых мешков; полетели — пустыри, баржи, заборы, брезенты и многие домики. [Андрей Белый. Петербург (1913–1914)]”	
	Pred	Not identified	N/A

	Gold	Груда	- Neither lost nor new sense - Frequency declined over time
Scores		Sense identification recall: 0.4 (identified sense 2 linked to two distinct senses in the ref book (Dobrushina et al 2016)) % of correctly identified lexical semantic change patterns: 100%	

5. Conclusions

In this thesis, current methods and tools for lexical semantic change detection were reviewed, and the opportunities and gaps in the pool of existing solutions were discussed. A web application was built that combines the substitution-based WSI approach to lexical semantic change detection with visualizations and tools to inspect, interpret and correct the computational results. Experiments were conducted to assess if the system would match the findings of semantic change in a book (Dobrushina et al. 2016) written by professionals in linguistics about the evolution of words in the Russian language.

The experiments show that human-machine collaboration can significantly improve the computational-only results for the clustering-based lexical semantic change detection; although, the user's language knowledge plays a vital role in inspecting and improving the results. The experiments were done by a user with limited Russian language knowledge, mapping senses to English. Often multiple senses were covered by a single English word, for example the Russian word ‘привет’ with senses ‘Формула конца письма с выражением внимания к третьему лицу’, ‘Формула конца письма с выражением внимания к адресату письма’ and ‘Здравствуйте’, were identified simply as ‘greetings’ in English.

In this work, the problem has been cast to the problem of precisely finding all senses for the target word for each time period, using a clustering-based approach, so the quality of clustering has been essential for solving the task.

Substitution-based WSI's interpretability played a central role in building a visual, interactive user interface. However, in the future, there is an opportunity to explore clustering methods that yield better clustering results and tackle the problem of interpretability for each identified cluster separately.

Two types of clustering errors - 'incorrect merges' and the 'absence of merge' discussed in 3.3 are effortlessly addressable if they happen at the late stages of the

clustering process. The primary value the sense inspection process adds is finding significant clusters of different senses that were merged and 'splitting' those into distinct clusters, and finding significant clusters that were not merged to other clusters with the same sense. Both tasks are relatively effortlessly possible to accomplish with the sense inspection procedure; however, when incorrect merges happen at earlier stages of the clustering process, the effort of manually identifying and correcting them becomes significantly higher. The main focus for future work should be on improving the quality of agglomerative clustering at earlier stages.

This work has been done as part of a research project in lexical semantic change detection led by Dr. Nikolay Arefyev. All of the code is committed to the repository of the project.

Lexical semantic change detection is challenging yet exciting and important for computers and humans. Hopefully, the proposed system will serve as an example and a stepping stone to make progress in the field and help researchers interested in clustering-based approaches to lexical semantic change detection improve their methods by better understanding intermediate states of clustering.

7. References

[1] N. V. Arefyev and D. A. Bykov. 2021. Lomonosov Moscow State University, Samsung R&D Institute Russia, HSE University. An Interpretable Approach to Lexical Semantic Change Detection with Lexical Substitution, 31–46,. DOI:10.28995/2075-7182-2021-20-31-46

[2] Adam Jatowt, Nina Tahmasebi and Lars Borin. 2021. Computational approaches to lexical semantic change: Visualization systems and novel applications. In Nina Tahmasebi, Lars Borin, Adam Jatowt, Yang Xu & Simon Hengchen (eds.), Computational approaches to semantic change, 311–339. Berlin: Language Science Press. DOI:10.5281/zenodo.5040320.

[3] Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak & Erez Lieberman Aiden. 2011. Quantitative analysis of culture using millions of digitized books. Science 331. 176–182.

[4] A. Stefanowitsch and Gries, S. 2003. Collostructions: investigating the interaction of words and constructions. Int. J. Corpus Linguist. 8(2), 209–243.

[5] A. Kilgarriff, Rychly, P., Smrz, P., and Tugwell, D. 2004. The Sketch Engine. In Proceedings of the 11th EURALEX International Congress (Lorient, France). 105--116
[6] A. Kilgarriff, O. Herman, J. Busta, V. Kovar, and M. Jakubıcek. 2015. DIACRAN: A framework for diachronic analysis. In Abstract Book. Corpus Linguistics 2015
[7] Adam Jatowt, Ricardo Campos, Sourav S. Bhowmick, Nina Tahmasebi, and Antoine Doucet. 2018. Every Word Has Its History: Interactive Exploration and Visualization of Word Sense Evolution. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 1899–1902. Torino Italy: ACM. https://doi.org/10.1145/3269206.3269218.
[8] Christian Rohrdantz, Annette Hautli, Thomas Mayer, Miriam Butt, Daniel A. Keim, and Frans Plank. 2011. Towards Tracking Semantic Change by Visual Analytics. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 305–10. Portland, Oregon, USA: Association for Computational Linguistics. https://aclanthology.org/P11-2053.
[9] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. Journal of Machine Learning Research, 3:993–1022.
[10] Kris Heylen, Dirk Speelman and Dirk Geeraerts. 2012. Looking at word meaning: An interactive visualization of semantic vector spaces for Dutch synsets. In Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH, 16–24. ACL.
[11] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. Journal of Machine Learning Research 9. 2579–2605.
[12] William L. Hamilton, Jure Leskovec and Dan Jurafsky. 2016. Cultural shift or linguistic drift? Comparing two computational measures of semantic change. In Proceedings of EMNLP 2016, 2116–2121. Austin: ACL. DOI: 10 . 18653 /v1 /D16 -1229.
[13] Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi & Steven Skiena. 2015. Statistically significant detection of linguistic change. In Proceedings of the 24th international conference on the World Wide Web, 625–635. Florence: ACM. DOI: 10.1145/2736277.2741627.
[14] Zaikun Xu and Fabio Crestani. 2017. Temporal semantic analysis and visualization of words. In IIR, 52–62.
[15] Yoon Kim, Yi-I. Chiu, Kentaro Hanaki, Darshan Hegde, and Slav Petrov. 2014. Temporal Analysis of Language through Neural Language Models. ArXiv:1405.3515 [Cs], May. http://arxiv.org/abs/1405.3515.

[16] Roberto Therón and Laura Fontanillo. 2013. Diachronic-Information Visualization in Historical Dictionaries. Information Visualization. https://doi.org/10.1177/1473871613495844.
[17] Ying Zhang, Stephan Vogel & Alex Waibel. 2004. Interpreting bleu/nist scores: How much improvement do we need to have a better system? In Proceedings of LREC 2004. ELRA.
[18] Arturas Mazeika, Tomasz Tylenda and Gerhard Weikum. 2011. Entity timelines: Visual analytics and named entity evolution. In Proceedings of the 20th ACM international conference on information and knowledge management, 2585–2588. ACM.
[19] Yating Zhang, Adam Jatowt, S. Sourav Bhowmick & Yuji Matsumoto. 2019. ATAR: Aspect-based temporal analog retrieval system for document archives. In WSDM2019. Melbourne: ACM
[20] T. Mikolov, K. Chen, G. Corrado, and J. Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In Proc. of ICLR Workshop.
[21] Osman Başkaya, Enis Sert, Volkan Cirik, and Deniz Yuret. 2013. AI-KU: Using Substitute Vectors and Co-Occurrence Modeling For Word Sense Induction and Disambiguation. In Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), 300–306. Atlanta, Georgia, USA: Association for Computational Linguistics. https://aclanthology.org/S13-2050.
[22] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised Cross-Lingual Representation Learning at Scale. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 8440–51. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.747.
[23] Stefan Th. Gries and Martin Hilpert. 2008. The Identification of Stages in Diachronic Data: Variability-based Neighbour Clustering. In: Corpora 3.1, pp. 59–81.
[24] Dmitry Kobak and Philipp Berens. 2019. The Art of Using T-SNE for Single-Cell Transcriptomics. Nature Communications 10 (1): 5416. https://doi.org/10.1038/s41467-019-13056-x.
[25] Dobrushina Nina R., Daniel Mikhail A., Danova Margarita K., Opachanova Anastasiia S., Pechurina Varvara S., Skorinkin Daniil A., Sheshenina Aleksandra V. 2016. Two centuries in twenty words [Dva veka v dvadcati slovah]. The National Research University Higher School of Economics,.

[26] Zhikov V. B. Arefyev N. V. Word Sense Inspector – induced word senses exploration toolkit. https://www.dialog-21.ru/media/4907/zhikovvb-arefyev-nv.pdf
[27] Asaf Amrami and Yoav Goldberg. 2018. Word Sense Induction with Neural BiLM and Symmetric Patterns. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 4860–67. Brussels, Belgium: Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1523.
[28] Asaf Amrami and Yoav Goldberg. 2019. Towards Better Substitution-Based Word Sense Induction. ArXiv:1905.12598 [Cs]. http://arxiv.org/abs/1905.12598.
[29] Gries, St.Th. and A. Stefanowitsch. 2004. Extending collostructional analysis: a corpus-based perspectives on “alternations”, International Journal of Corpus Linguistics 9 (1), pp. 97–129.
[30] Nikolay Arefyev, Boris Sheludko, and Tatiana Aleksashina. 2019. Combining Neural Language Models for WordSense Induction. ArXiv:2006.13200 [Cs] 11832: 105–21. https://doi.org/10.1007/978-3-030-37334-4_10.
[31] Matthew E Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. arXiv preprint arXiv:1802.05365.
[32] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
[33] Adam Jatowt and Kevin Duh. 2014. A Framework for Analyzing Semantic Change of Words across Time. In IEEE/ACM Joint Conference on Digital Libraries, 229–38. London, United Kingdom: IEEE. https://doi.org/10.1109/JCDL.2014.6970173.

Appendix A

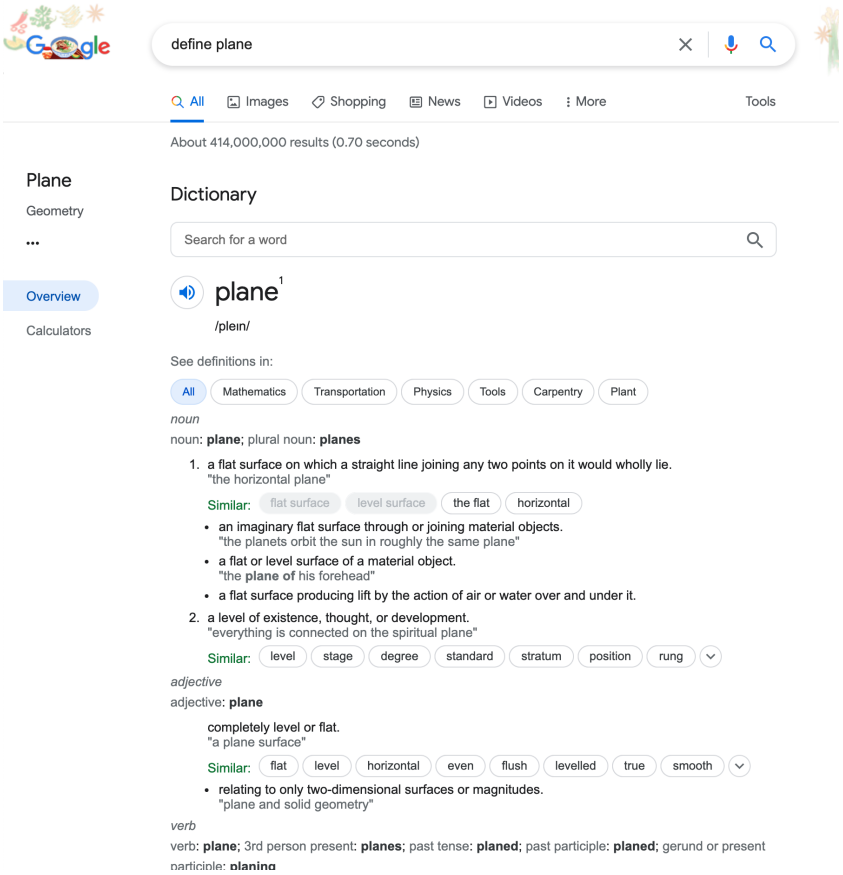


Figure 1. Google definition search, results for ‘define plane’, definitions list, screenshot taken in Dec. 2021

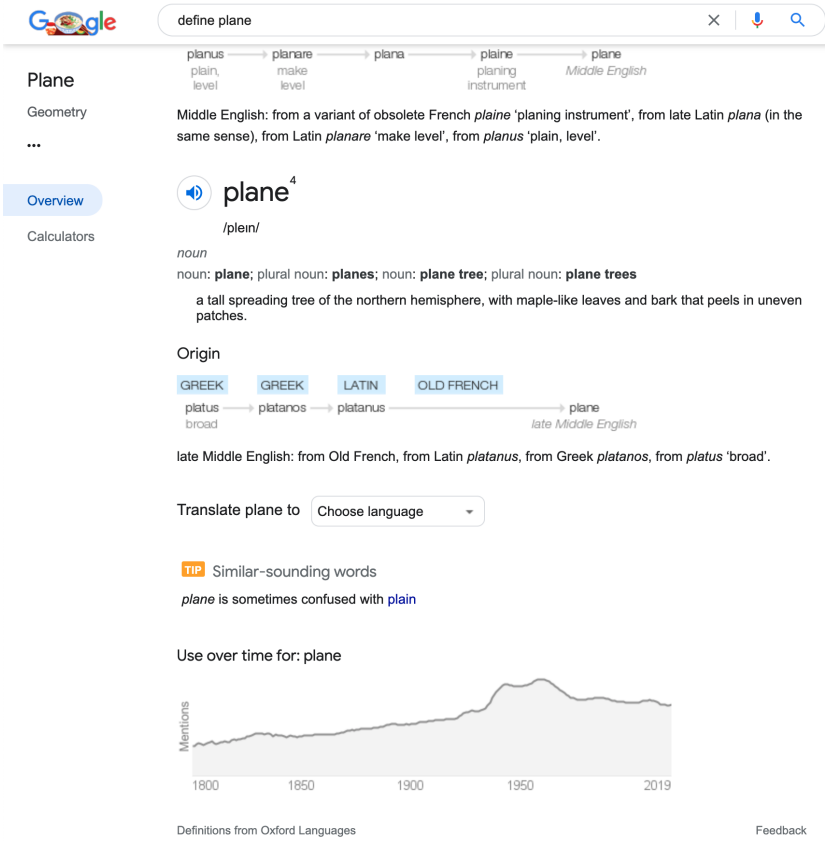


Figure 2. Google definition search, results for ‘define plane’, use over time chart, screenshot taken in Dec. 2021

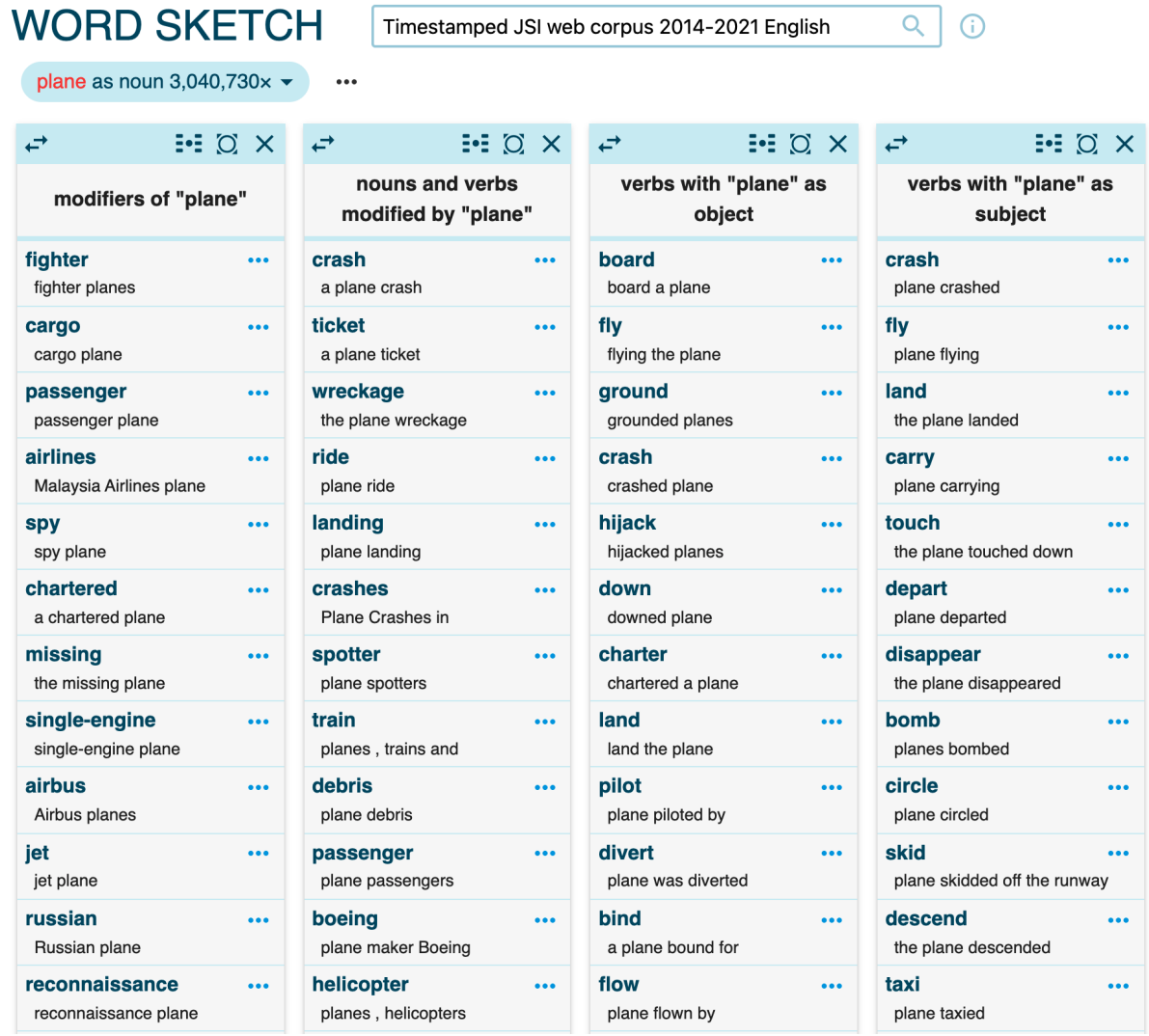


Figure 3. Word Sketch tool in Sketch Engine, results for lemma ‘plane’, screenshot taken in Dec. 2021

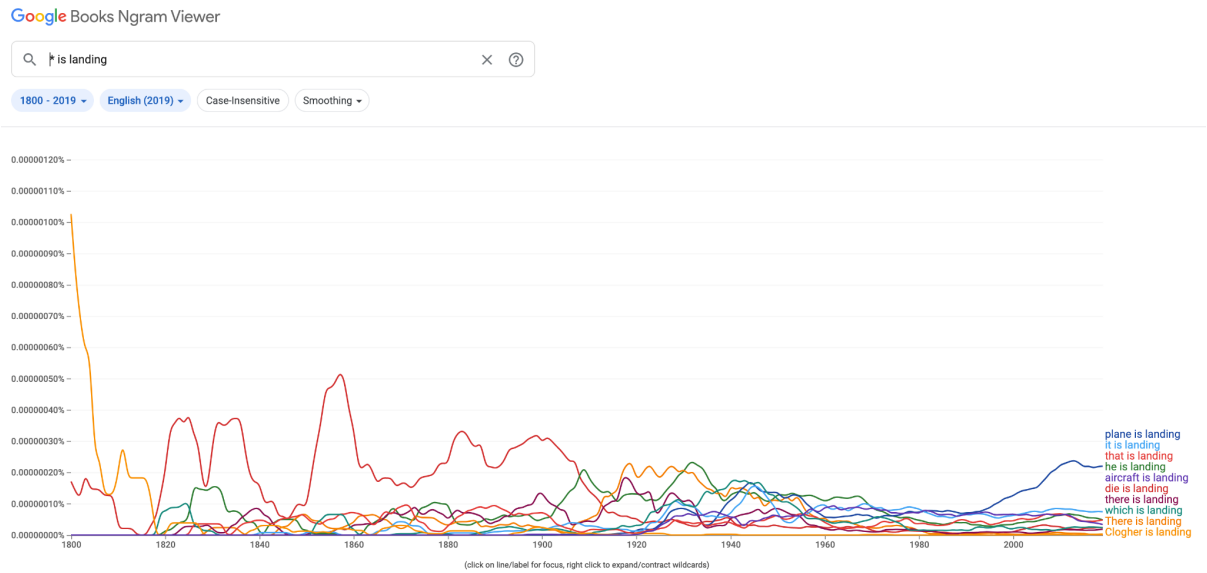


Figure 4. Google Books Ngram Viewer, results for ‘* is landing’, screenshot taken Dec. 2021

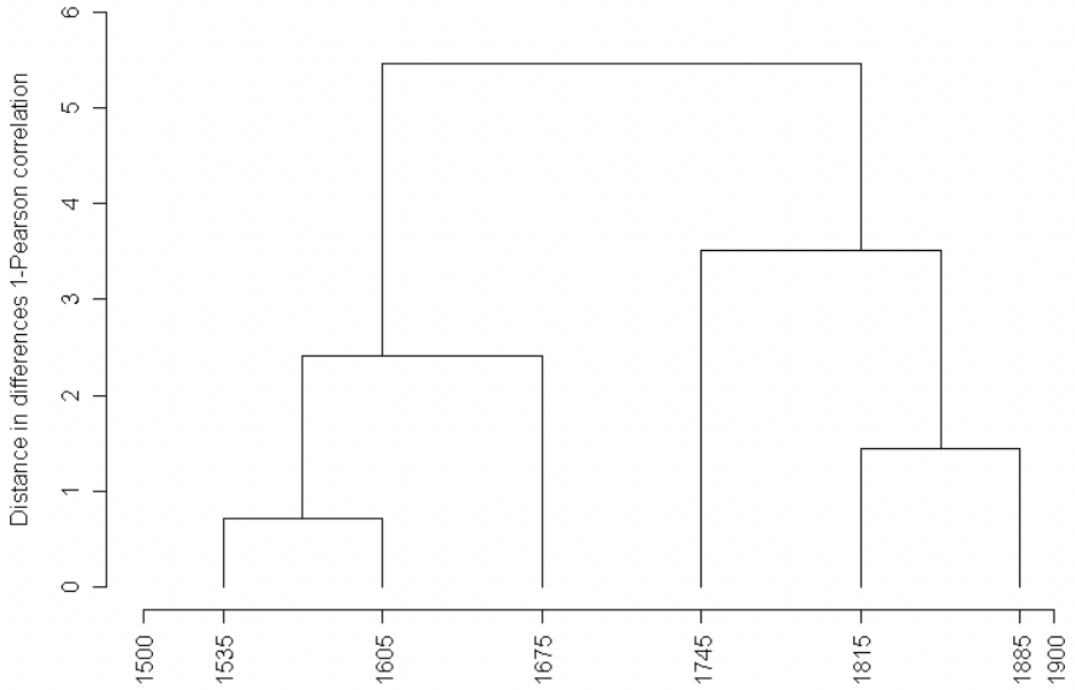


Figure 5. Variability-based neighbor clustering dendrogram, resulting for ‘shall’ + verb (infinitive). Screenshot taken in Dec. 2021 from the original paper. Hilpert and Gries. 2008. The Identification of Stages in Diachronic Data: Variability-based Neighbour Clustering.

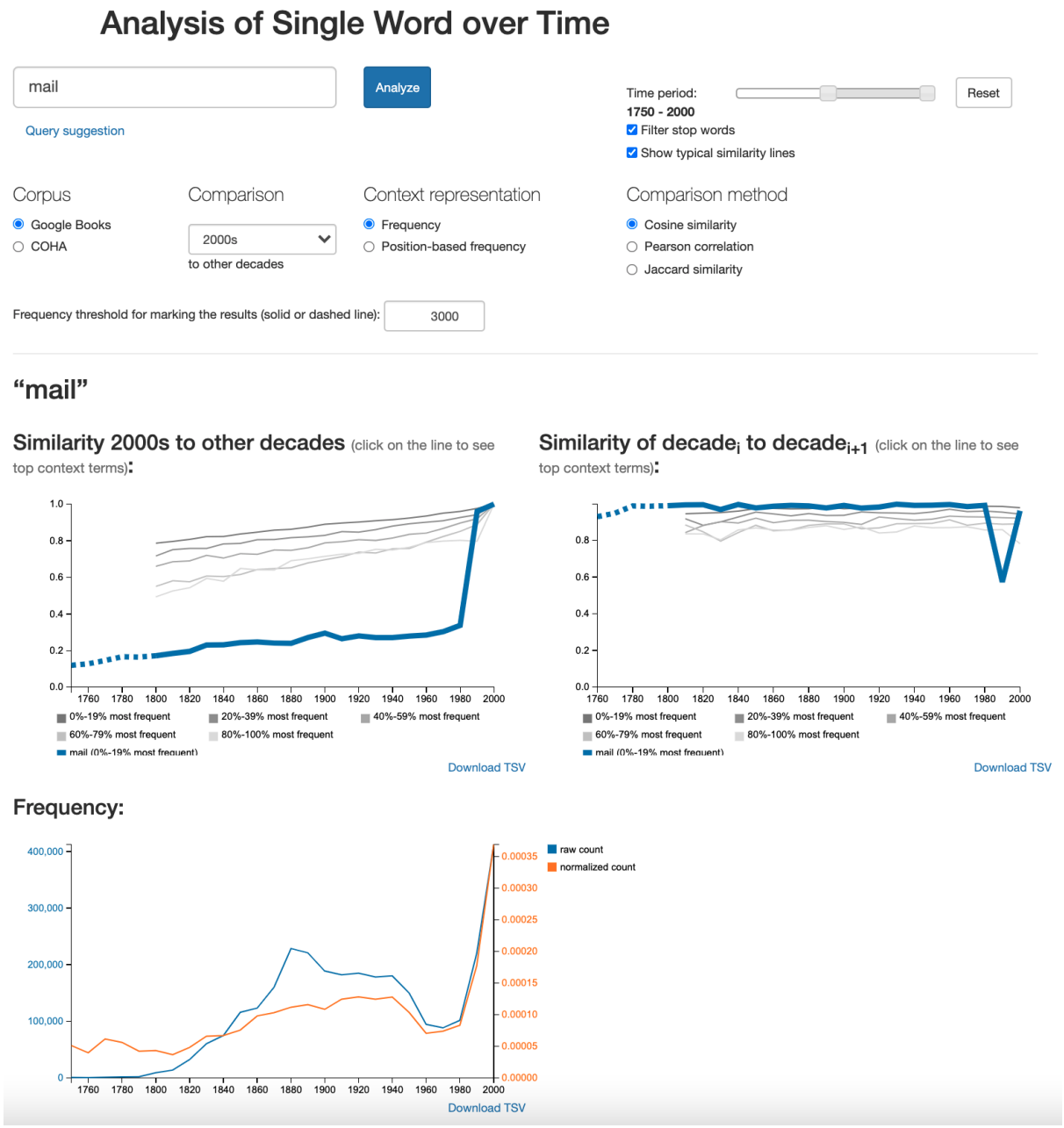


Figure 6. Word Evolution tool, Analysis of Single Word over Time page fragment, decades similarity and term frequency charts, results for 'mail'. Screenshot taken in Dec. 2021. https://www.okayama.silk.jp/WordEvolution/wordUsageOverTime Jatowt et al. 2018. Every Word Has Its History: Interactive Exploration and Visualization of Word Sense Evolution.

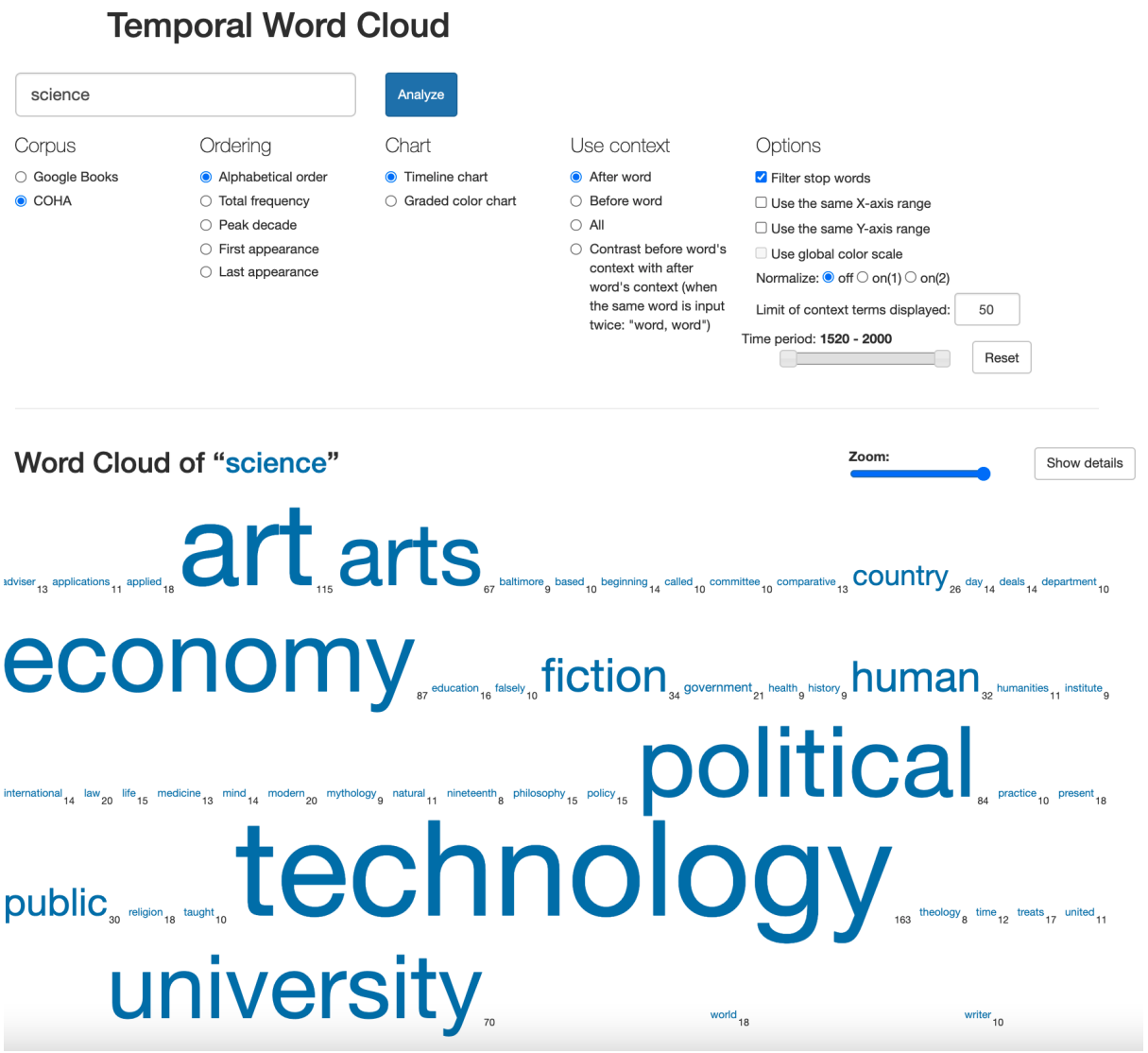


Figure 7. Word Evolution tool, Temporal word cloud, results for 'science'. Screenshot taken in Dec. 2021. https://www.okayama.silk.jp/WordEvolution/wordCloud Jatowt et al. 2018. Every Word Has Its History: Interactive Exploration and Visualization of Word Sense Evolution.

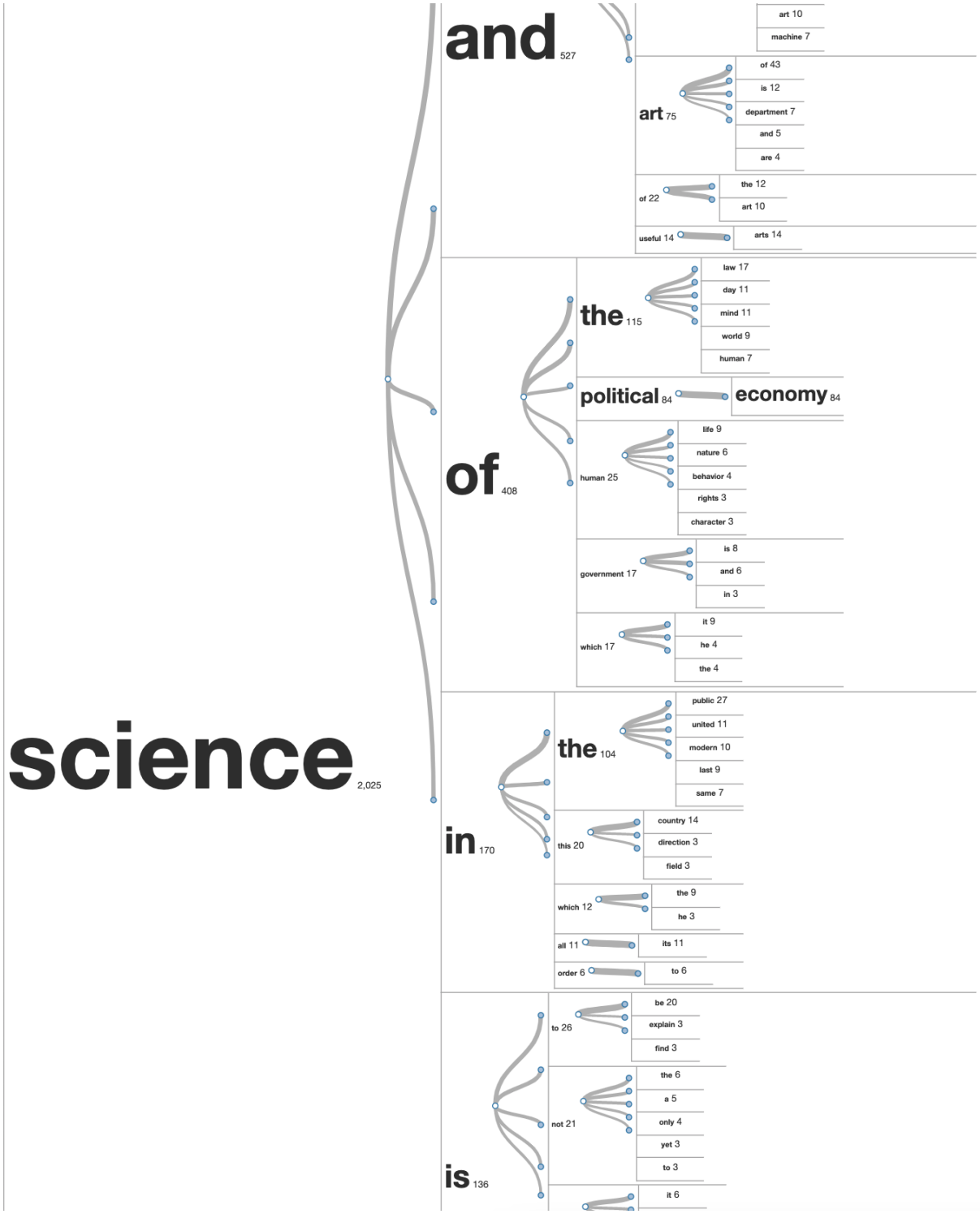


Figure 8. Word Evolution tool, Temporal word tree, results for ‘science’. Screenshot taken in Dec. 2021. https://www.okayama.silk.jp/WordEvolution/wordTreeResult Jatowt et al. 2018. Every Word Has Its History: Interactive Exploration and Visualization of Word Sense Evolution.

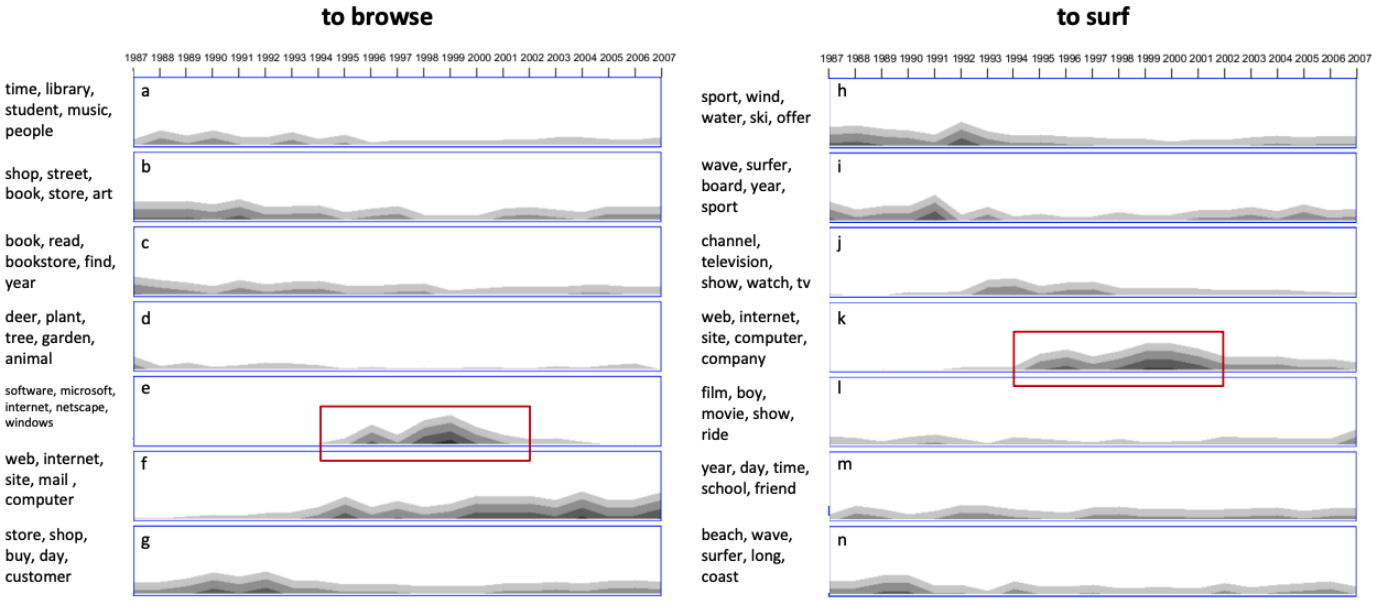


Figure 9. Temporal development of senses for ‘to browse’ and ‘to surf’. Screenshot taken from the original paper in Dec. 2021. Rohrdantz et al. 2011. Towards Tracking Semantic Change by Visual Analytics.

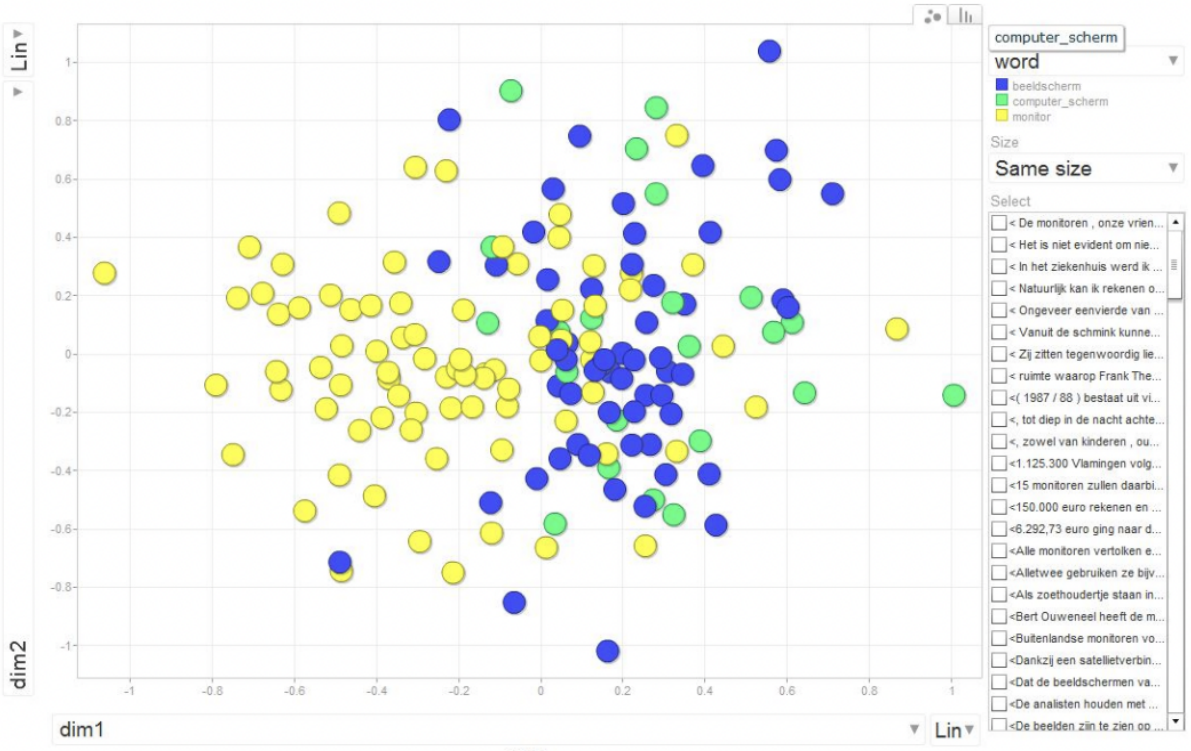


Figure 10. Motion Chart for COMPUTER SCREEN, screenshot taken Dec. in 2021, Heylen et al. 2012. Looking at word meaning: An interactive visualization of semantic vector spaces for Dutch synsets.

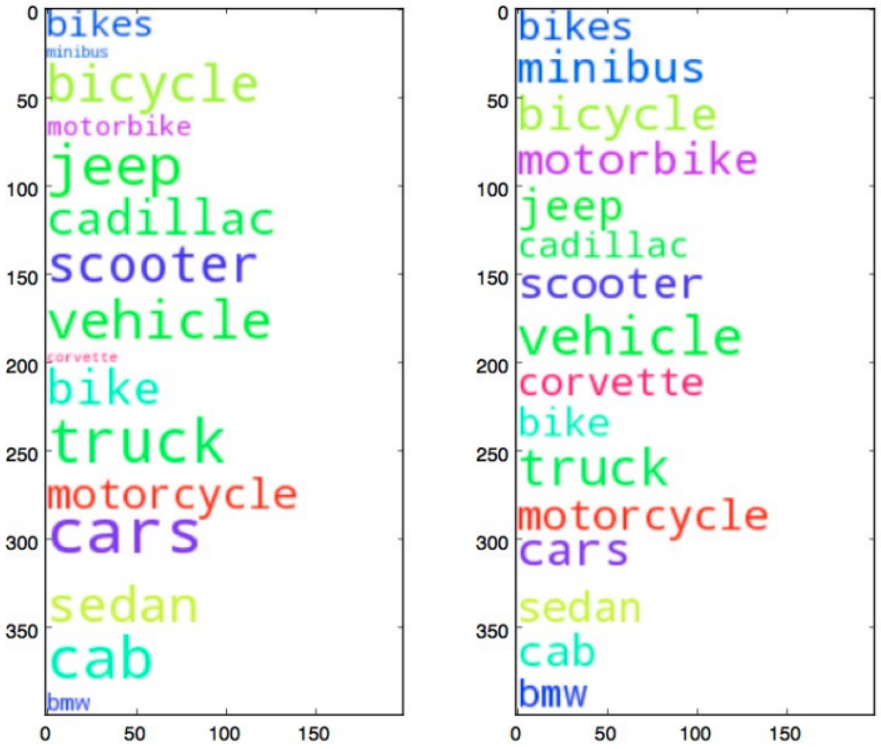


Figure 11. Semantic Similarity Word Cloud, for 1970 on the left and 2016 on the right, words most similar to ’car’ according to the New York Times. Screenshot taken from the original paper in Dec. 2021. Zaikun Xu and Fabio Crestani. 2017. Temporal semantic analysis and visualization of words

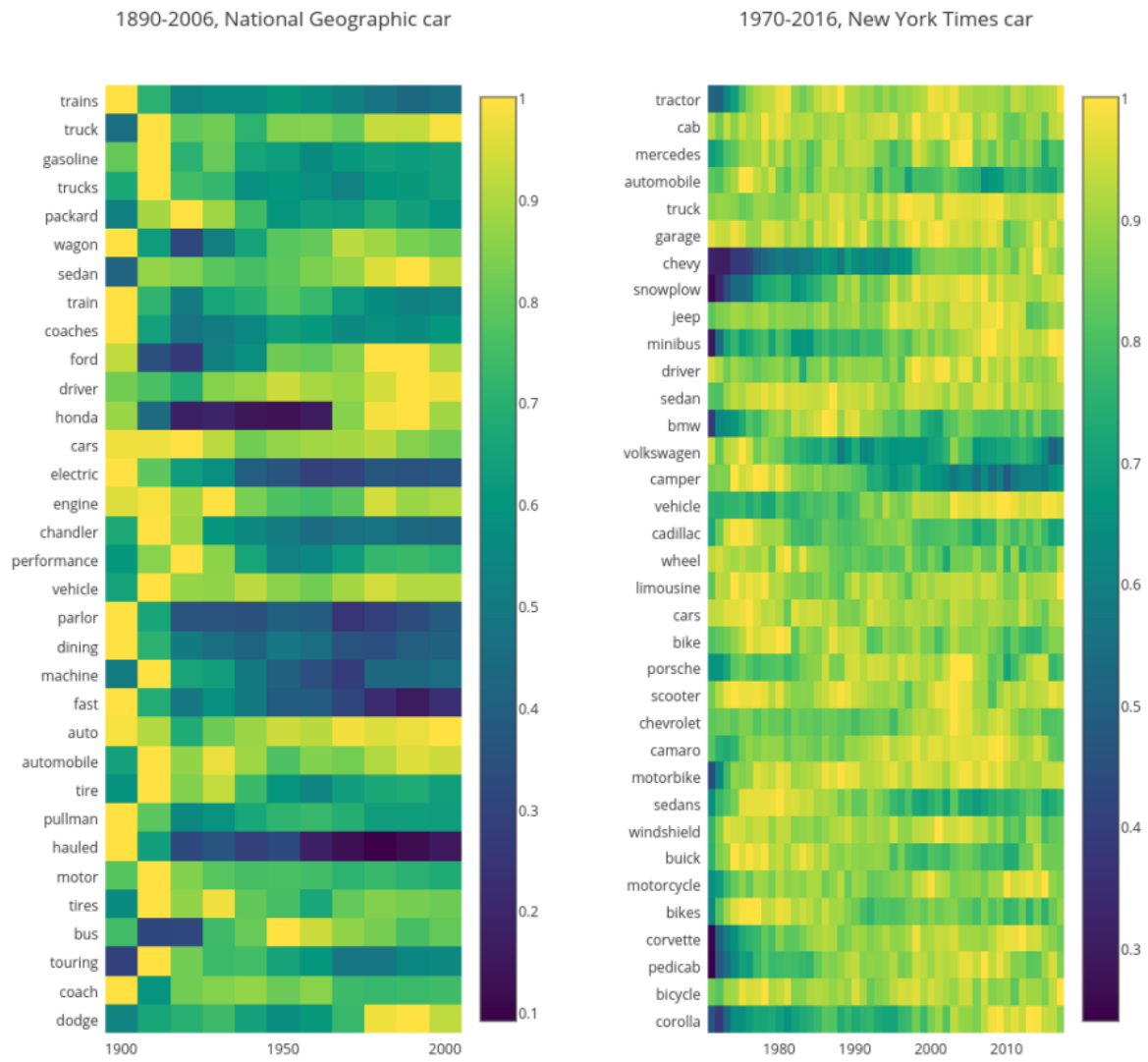


Figure 12. Heatmap visualisation of words most similar to the word ’car’. Screenshot taken from the original paper in Dec. 2021. Zaikun Xu and Fabio Crestani. 2017. Temporal semantic analysis and visualization of words

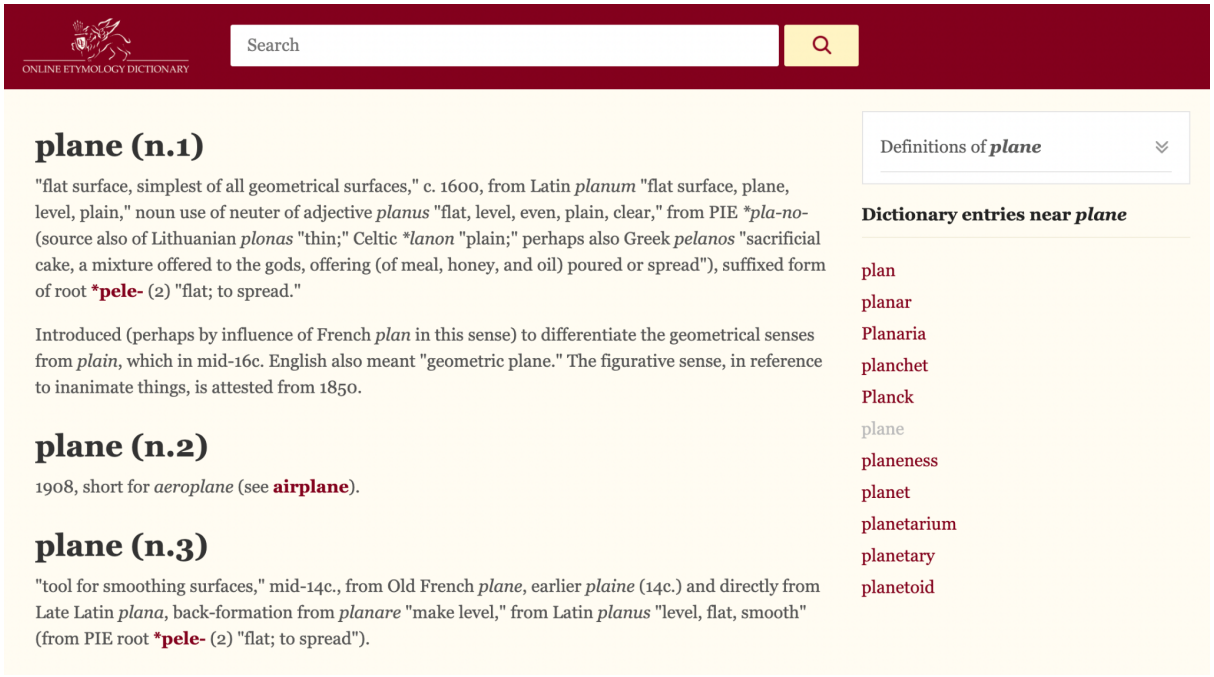


Figure 13. Online Etymology Dictionary, fragment of results for 'plane', screenshot taken Dec. in 2021

https://www.etymonline.com/word/plane

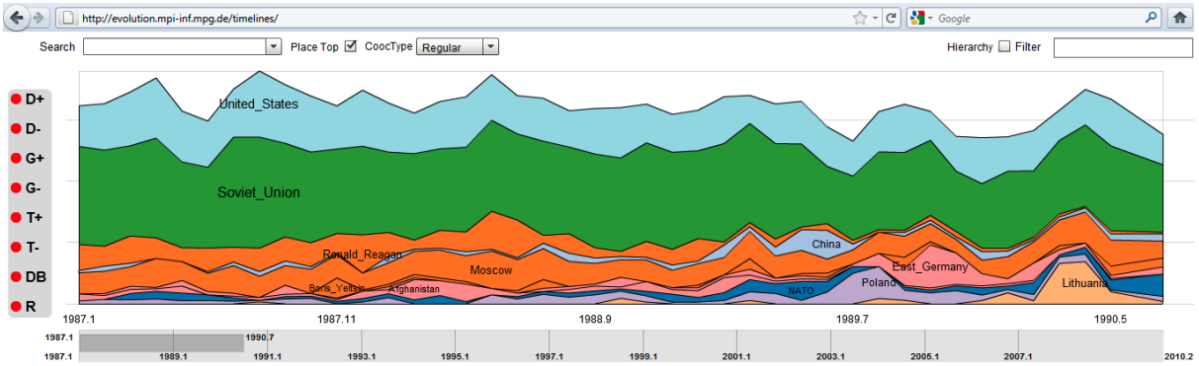


Figure 14. Entity timeline for Mikhail_Gorbachev. Screenshot taken from the original paper in Dec. 2021. Arturas Mazeika, Tomasz Tylenda and Gerhard Weikum. 2011. Entity timelines: Visual analytics and named entity evolution

Appendix B

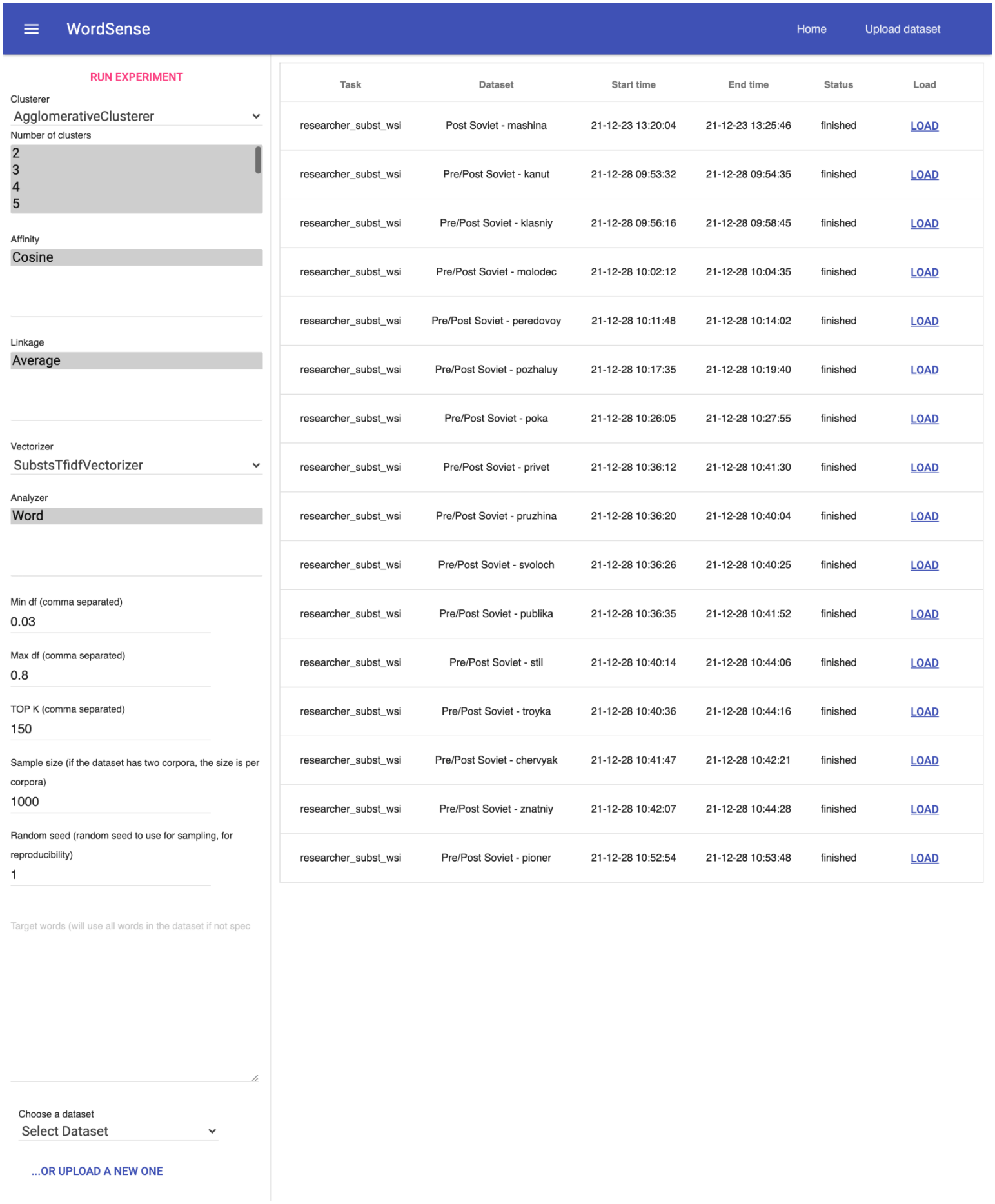


Figure 1. Main page, WordSense tool, screenshot taken in Dec. 2021

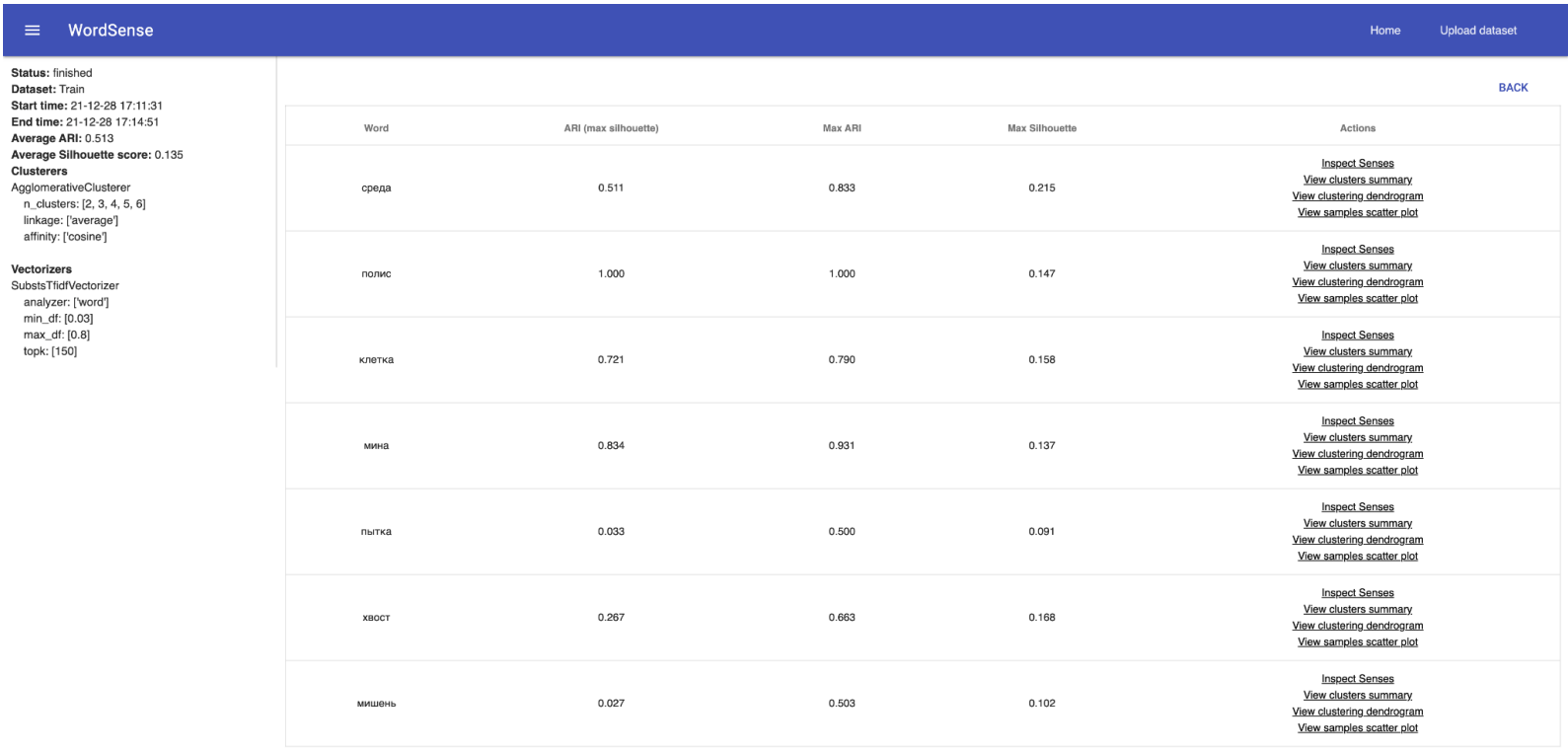


Figure 2. Experiment page, WordSense tool, screenshot taken in Dec. 2021

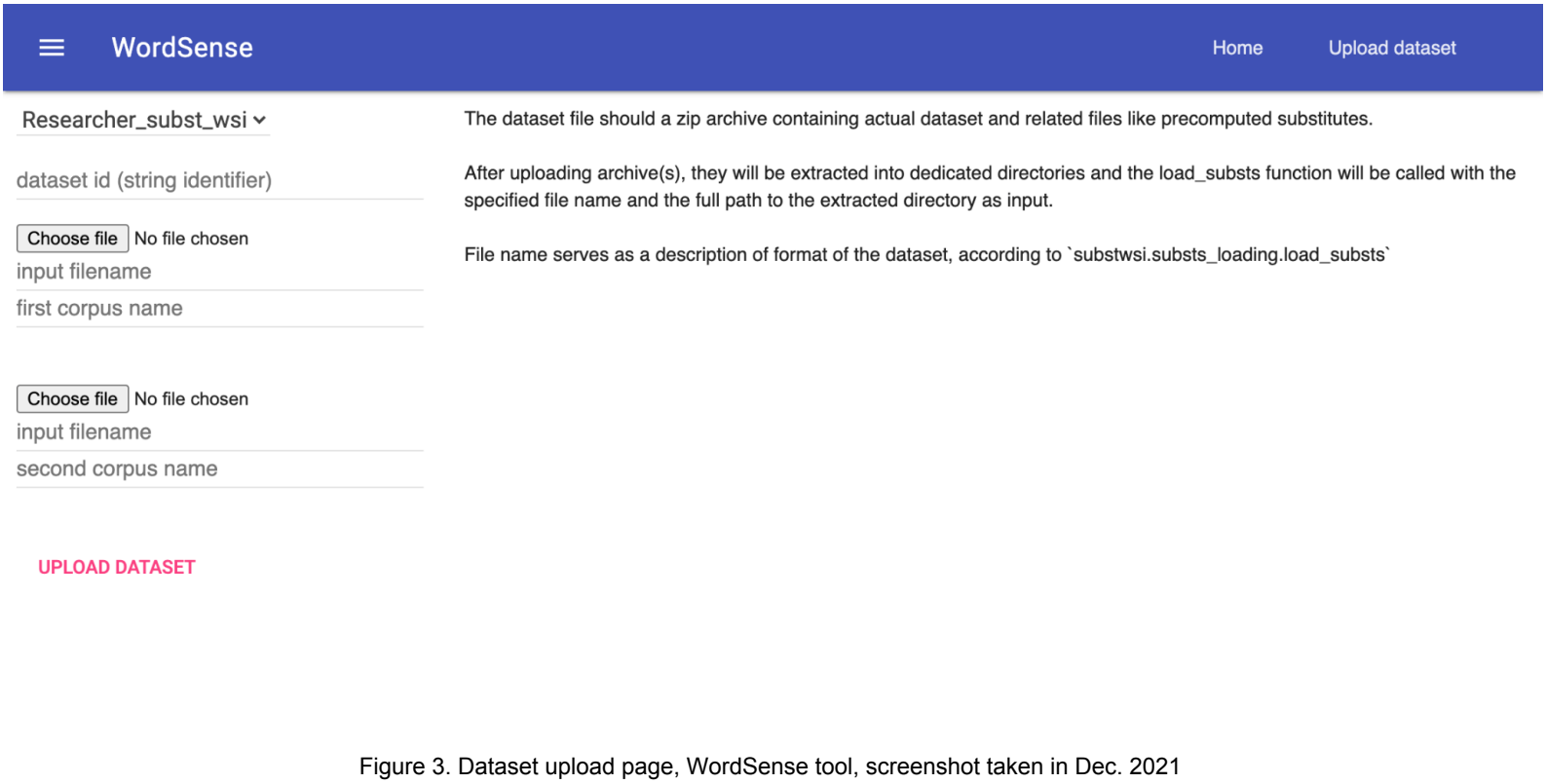


Figure 3. Dataset upload page, WordSense tool, screenshot taken in Dec. 2021

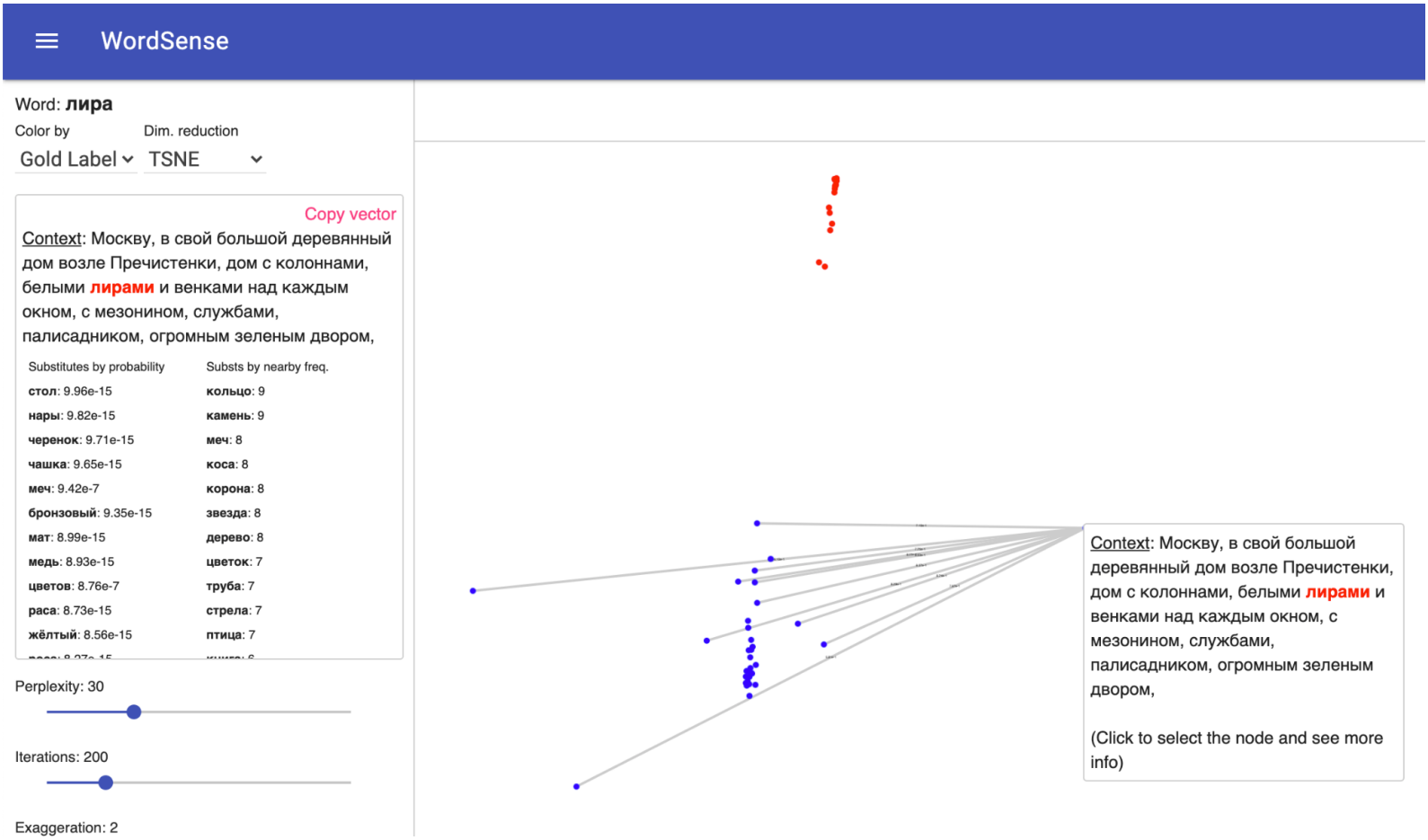


Figure 4. Word in context scatterplot for word ‘лира’, demonstrating the selection of a single point, screenshot taken in Dec. 2021

Figure 5. Word in context scatterplot for word ‘лира’, demonstrating the selection of multiple points, screenshot taken in Dec. 2021

Figure 6. Clustering summary for ‘пакет’, pre soviet and post soviet corpora, screenshot taken in Dec. 2021

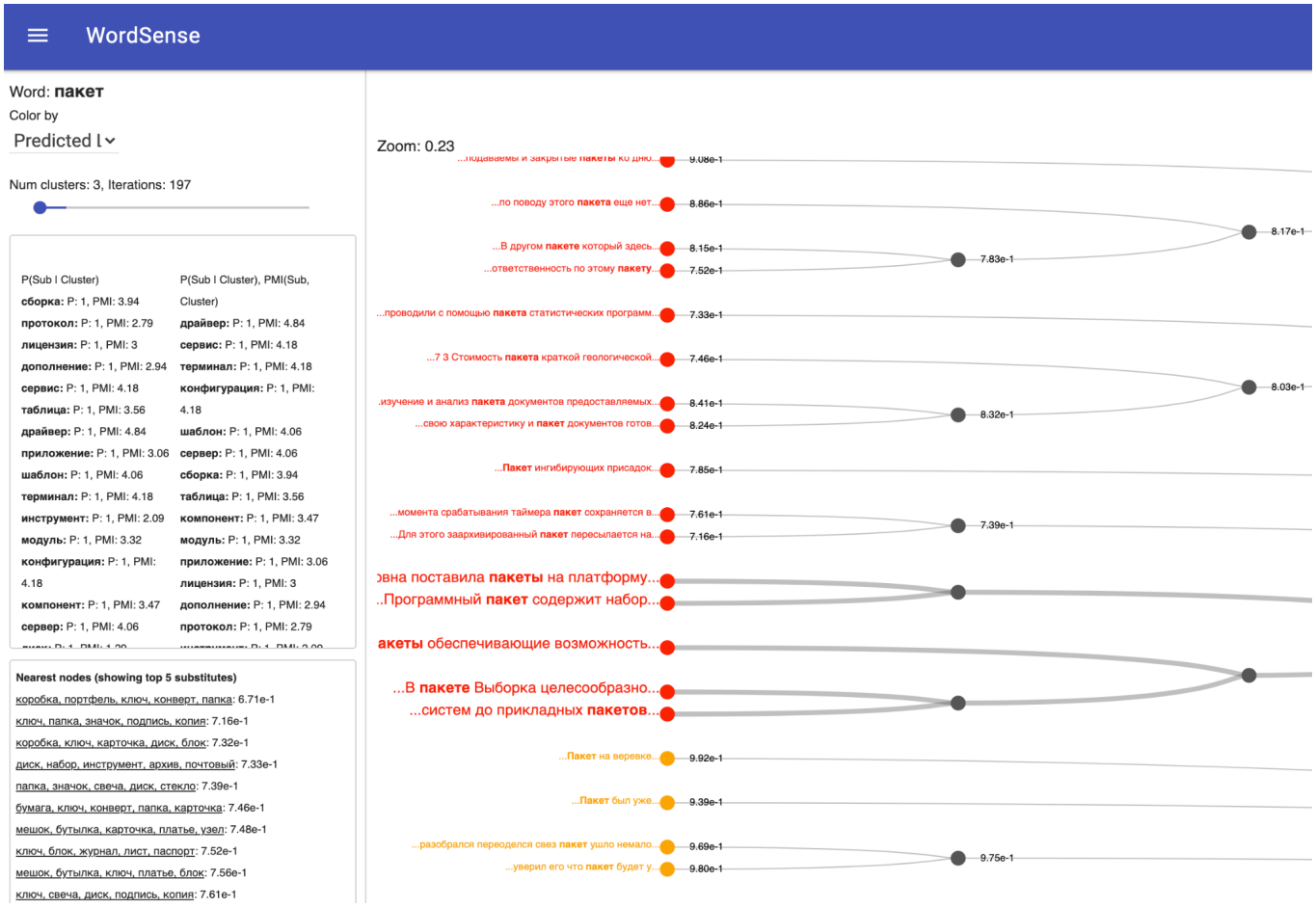


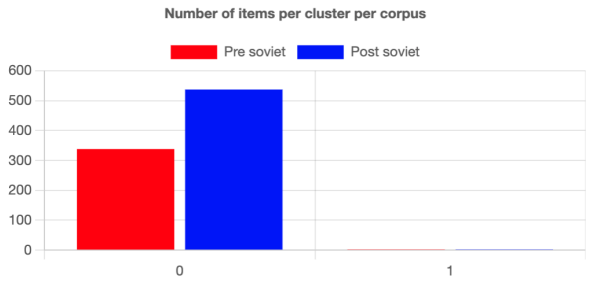
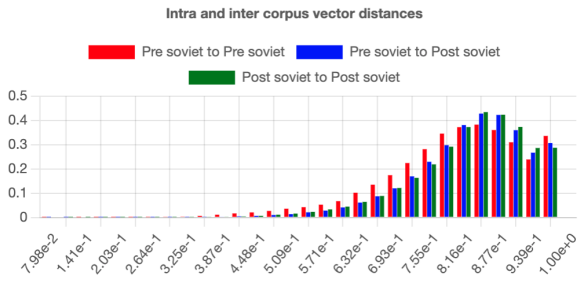
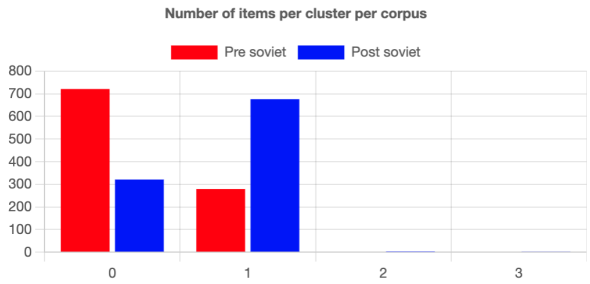
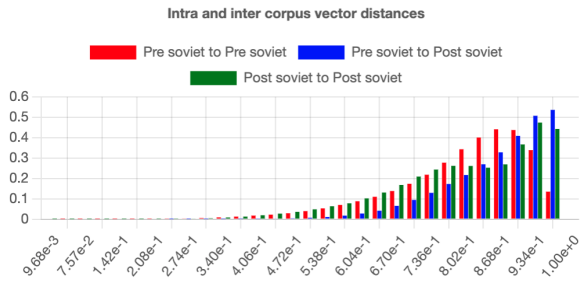
Figure 7. A fragment of the clustering dendrogram for ‘пакет’, demonstrating the selection of a cluster node, screenshot taken in Dec. 2021

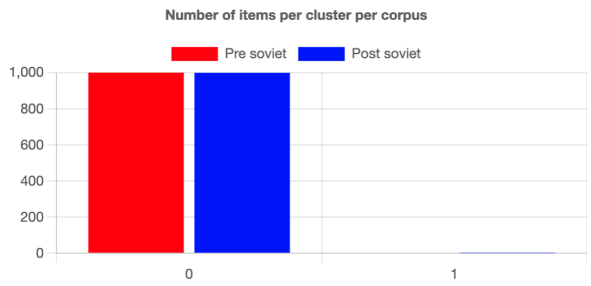
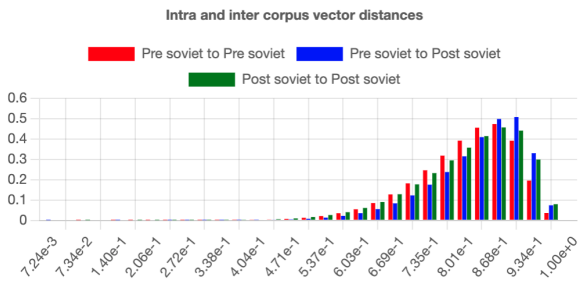
Figure 8. A fragment of the clustering dendrogram for ‘пакет’, demonstrating the selection of a sample node, screenshot taken in Dec. 2021

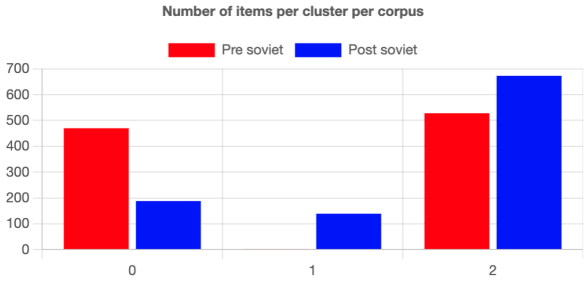
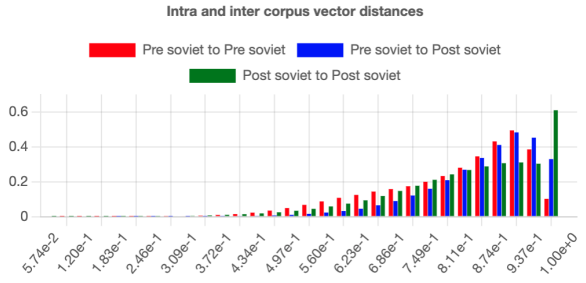
Figure 9. A fragment of the clustering dendrogram for ‘пакет’, demonstrating the selection of a pair of nodes, screenshot taken in Dec. 2021

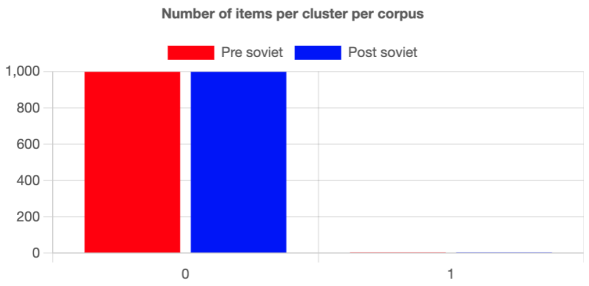
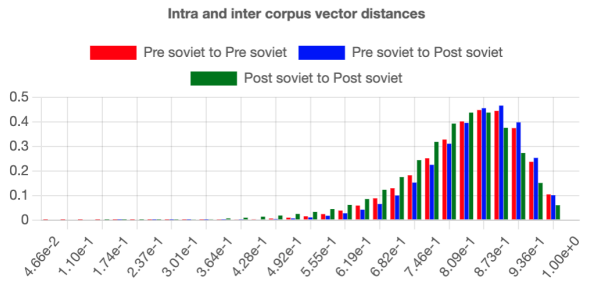
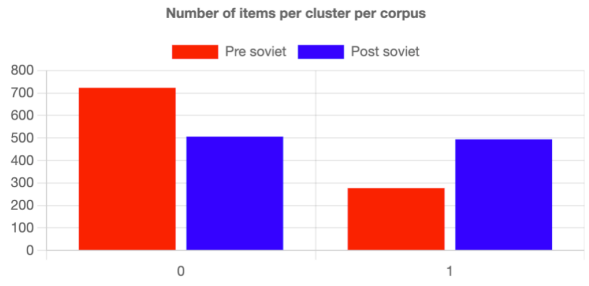
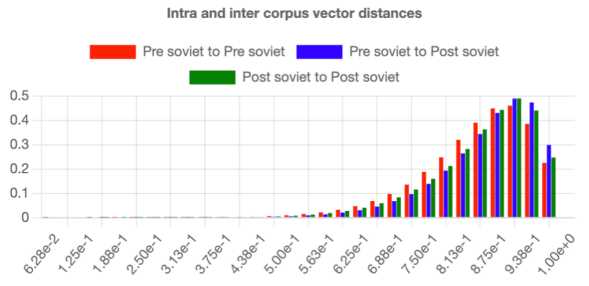
Figure 10. A fragment of the sense inspection tool for ‘пионер’, pre-soviet and post-soviet corpora, screenshot taken in Dec. 2021

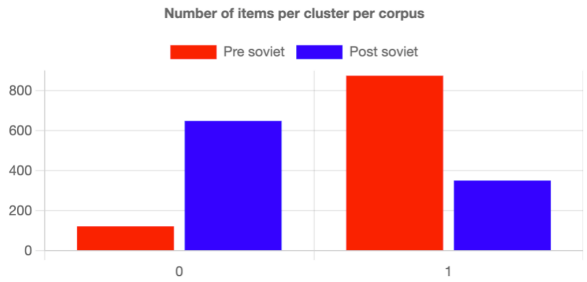
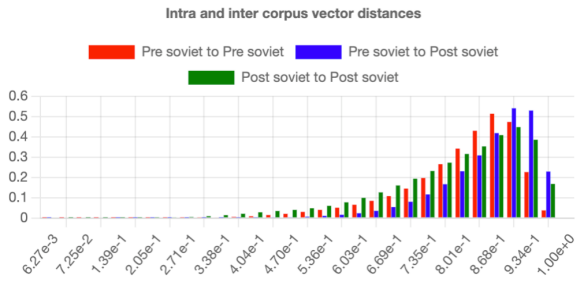
Appendix C

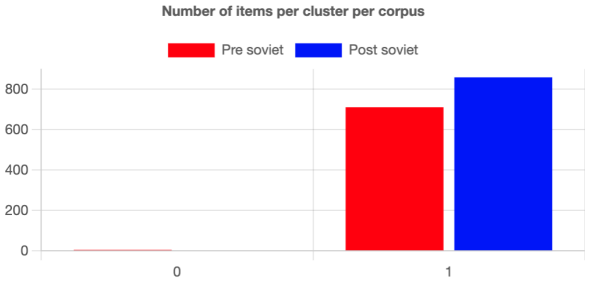
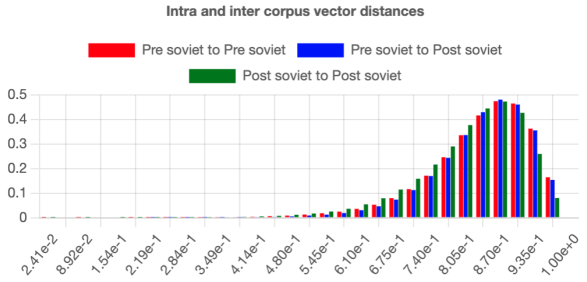
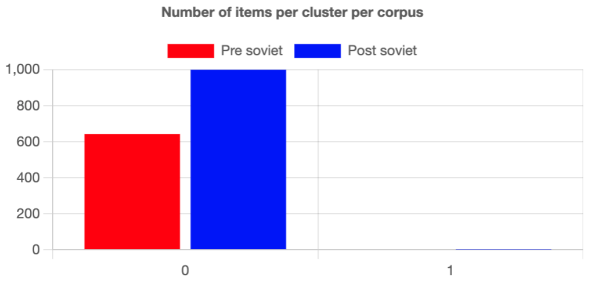
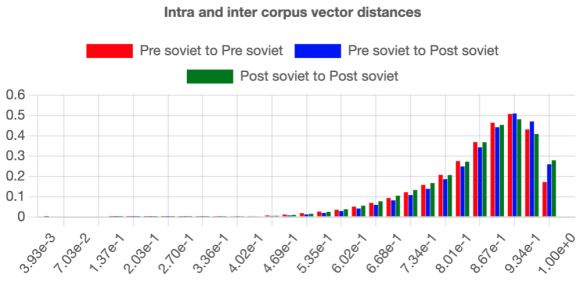
Word	Automatic results, senses and scores			
кануть	 			
	Sense 1	Gold	Утонуть, упасть на дно	Stable
		Pred	**Description:** Collapse, fall into water (often mentioned as fall into the river Lethe) **Signature:** падать, лететь, ехать, рухнуть, уйти, провалиться, полететь, попасть, катиться, сойти	Stable
	Sense 2	Gold	Исчезнуть из виду	New
		Pred	**Description:** Vanish, disappear, leave **Signature:** пропасть, покинуть, проснуться, скрыться, уйти, сбежать, кончить, развестись, прибыть, падать	Stable
	Sense 3	Gold	Пройти, минуть, исчезнуть	Stable
		Pred	Description: Die in figurative sense, get lost in time and space **Signature:** погибнуть, быть, умереть, идти, лечь, вернуться, пройти, мёртвый, потерять, уходить	Stable
	Sense 4	Gold	Капнуть	Stable
		Pred	Not identified	N/A
	Scores	Sense identification recall: 0.75 % of correctly identified lexical semantic change patterns: 66.6%		
классный	 			
	Sense 1	Gold	Not linked	N/A
		Pred	**Description:** Classroom, a place for taking classes in the school **Signature:** школьный, личный, семейный, частный, учительский, детский, кабинетный, групповой,	Stable

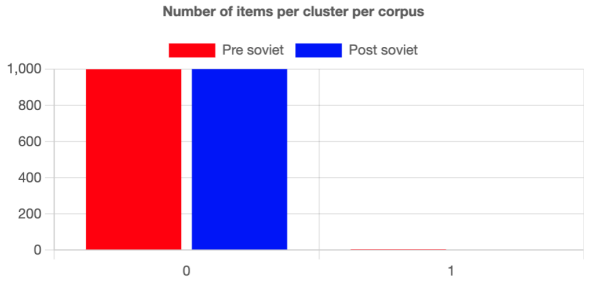
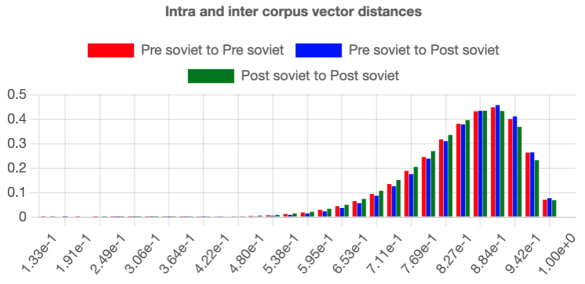
			педагогический, учёный	
	Sense 2	Gold	Not linked	N/A
		Pred	**Description:** A person on a authoritative position in the school, female - ‘классная дама’ **Signature:** советский, молодой, мужской, образованный, красивый, знаменитый, модный, старший, умный, добрый	Stable (Freq. declined)
	Sense 3	Gold	Хороший, отличный	New
		Pred	**Description:** Cool, great, excellent **Signature:** замечательный, интересный, здоровый, крутой, удачный, умный, забавный, модный, любить, красивый	New
	Sense 4	Gold	Имеющий класс (разряд)	Stable
		Pred	Not identified	N/A
	Sense 5	Gold	Имеющий отношение к школьному обучению	Stable
		Pred	Not identified	N/A
	Scores	Sense identification recall: 0.33 % of correctly identified lexical semantic change patterns: 100%		
молодец	 			
	Sense 1	Gold	Мужчина	Stable
		Pred	**Description:** Man, male **Signature:** старик, мужик, девочка, слуга, мальчик, старец, брат, сестра, товарищ, офицер	Stable
	Sense 2	Gold	Похвала (in the reference this sense is split into two, since at first it was used only for male and later adapted to female, that was not spotted in the experiment)	Stable
		Pred	**Description:** Appreciation, recognition **Signature:** красивый, замечательный, добрый, умница, красавица, гость, знаменитость, дурак, дочка, дама	Stable
	Sense 3	Gold	Служащий	Stable
		Pred	Not identified	N/A
	Scores	Sense identification recall: 0.66		

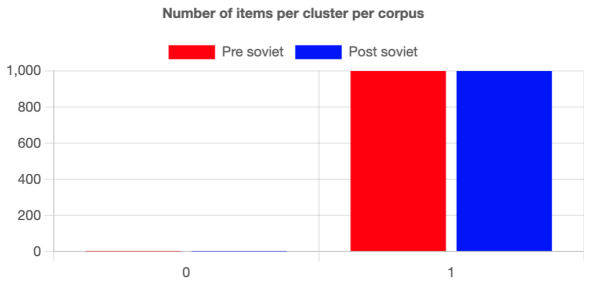
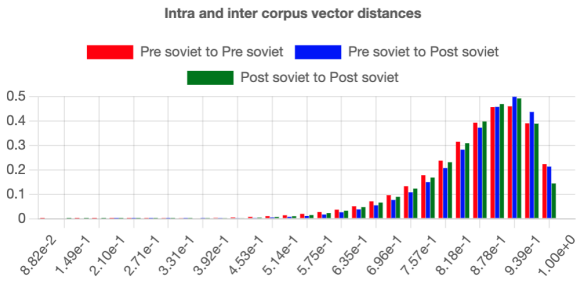
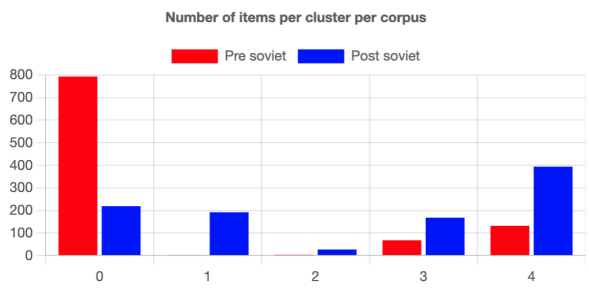
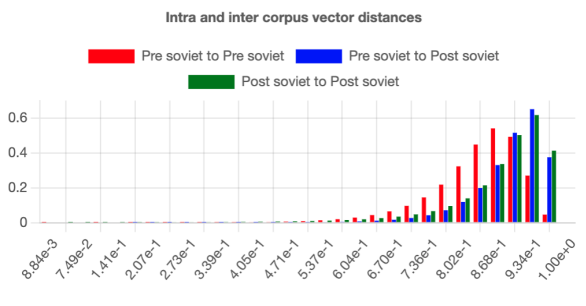
		% of correctly identified lexical semantic change patterns: 100%		
передовой	 			
	Sense 1	Gold	Прогрессивный	Stable
		Pred	**Description:** Progressive person **Signature:** опытный, умный, образованный, независимый, талантливый, советский, активный, известный, честный, богатый	Stable
	Sense 2	Gold	Статья (in the reference this sense is split into two, since at first it was used only for male and later adapted to female, that was not spotted in the experiment)	Stable
		Pred	**Description:** Popular publication/article **Signature:** популярный, авторский, публицистический, интересный, советский, оригинальный, революционный, спорный, авторитетный, важный	Stable
	Sense 3	Gold	Расположенный в авангарде (о военных действиях)	Stable
		Pred	**Description:** War frontline **Signature:** следовой, военный, тыл, боевой, запасный, главный, разведка, лагерь, вооружение, оборона	Stable
	Sense 4	Gold	Расположенный впереди	Stable
		Pred	**Description:** Located in the front or in front of something **Signature:** центральный, крайний, задний, средний, фронтовой, второй, ходовой, следовой, вести, правый	Stable
	Sense 5	Gold	Посланник, гонец	Stable
		Pred	Not identified	N/A
	Sense 6	Gold	Идущий впереди	Stable
		Pred	Not identified	N/A
	Sense 7	Gold	Not linked	N/A
		Pred	**Description:** Best, leading **Signature:** эффективный, перспективный, базовый, современный, отечественный, качественный, технологический, специализированный, специализировать, альтернативный	New
	Scores	Sense identification recall: 0.66 % of correctly identified lexical semantic change patterns: 100%		

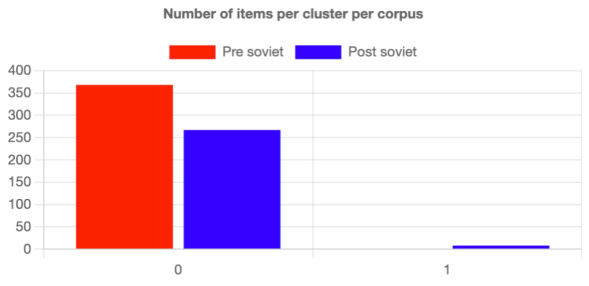
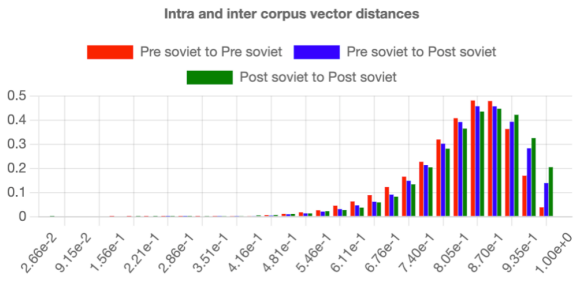
пожалуй	 			
	Sense 1	Gold	Возможно	Stable
		Pred	**Description:** Possibly, perhaps **Signature:** предположительный, предполагать, определённый, возможно,, вероятно,, наверное,, согласиться, полагать, поверить, пожалуй	Stable
	Sense 2	Gold	Будь добр	Lost
		Pred	**Description:** Please, in the sense of asking to do something **Signature:** слушать, господь, давать, смотреть, дорогой, советовать, шутить, верить, послушать, думать (substitutes are not descriptive, however small clusters with non-ambiguous example usages are present)	Lost
	Sense 3	Gold	Хороший, отличный	New
		Pred	**Description:** Cool, great, excellent **Signature:** замечательный, интересный, здоровый, крутой, удачный, умный, забавный, модный, любить, красивый	New
	Sense 4	Gold	Склоняюсь к тому, что...	Stable
		Pred	Not identified	N/A
	Sense 5	Gold	Ладно, согласен	Lost
		Pred	Not identified	N/A
	Scores	Sense identification recall: 0.5 % of correctly identified lexical semantic change patterns: 100%		
пока	 			
	Sense 1	Gold	Союз с фоновым значением (‘в то время как’)	Stable
		Pred	**Description:** While, for example while the water was boiling, the they played a chess game **Signature:** след, впереди, лето, пора, надолго, покуда,	Stable

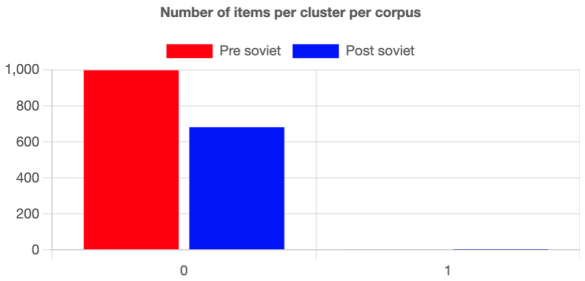
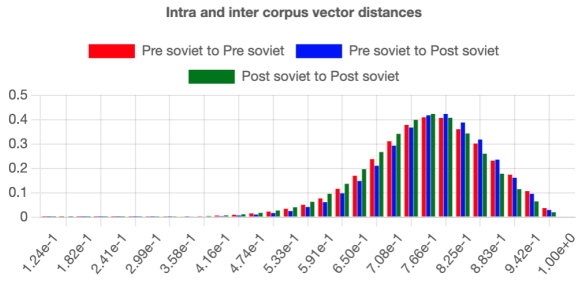
			отчий, доколе, сколь, предположительный (substitutes are not descriptive, however small clusters with non-ambiguous example usages are present)	
	Sense 2	Gold	Наречие — ‘в течение некоторого времени’, ‘до сих пор еще’	Stable
		Pred	**Description:** Yet, for example “I can’t tell that yet” **Signature:** определённый, надолго, предположительный, приблизительный, объективный, будущий, намеренный, предполагать, отчасти, теоретически (substitutes are not descriptive, however small clusters with non-ambiguous example usages are present)	Stable
	Sense 3	Gold	Элемент формулы прощания	Stable
		Pred	Not identified	N/A
	Sense 4	Gold	Этикетное слово — ‘до свидания’	Stable
		Pred	Not identified	N/A
	Scores	Sense identification recall: 0.5 % of correctly identified lexical semantic change patterns: 100%		
привет	 			
	Sense 1	Gold	1. Здравствуйте 2. Формула конца письма с выражением внимания к адресату письма 3. Формула конца письма с выражением внимания к третьему лицу	Stable
		Pred	**Description:** Greetings **Signature:** пожелание, благодарность, гость, знакомство, извинение, респект, милость, дружба, подарок, приветствие	Stable
	Sense 2	Gold	Дружелюбное, ласковое обращение с кем-либо	Stable
		Pred	Not identified	N/A
	Sense 3	Gold	Словесное или несловесное выражение внимания к собеседнику	Stable
		Pred	Not identified	N/A
	Scores	Sense identification recall: 0.6		

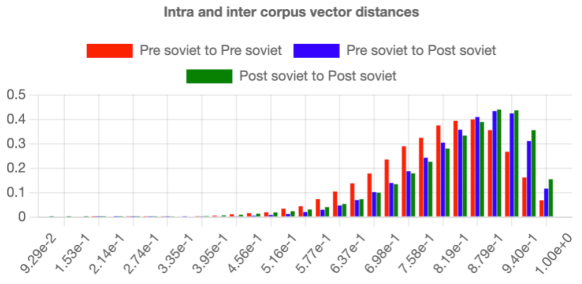
		% of correctly identified lexical semantic change patterns: 100% * Three senses were identified correctly, however they were mapped to a single English word		
пружина	 			
	Sense 1	Gold	Прямое значение	Stable
		Pred	**Description:** Spring in the sense of device **Signature:** опора, колесо, ручка, кольцо, кнопка, коробка, механизм, болт, свеча, двигатель	Stable
	Sense 2	Gold	Метафора причины	Stable
		Pred	Not identified	N/A
	Sense 3	Gold	Метафора сжатости (движения)	Stable
		Pred	Not identified	N/A
	Scores	Sense identification recall: 0.33 % of correctly identified lexical semantic change patterns: 100%		
сволочь	 			
	Sense 1	Gold	Not linked	N/A
		Pred	**Description:** Name of a river **Signature:** земля, река, солнце, гора, волга, огонь, долина, озеро, скал, суша	New
	Sense 2	Gold	Подлец	Stable
		Pred	**Description:** Low quality person, in a sense of insult **Signature:** идиот, враг, бандит, раб, собака, слуга, глупец, мужик, предатель, девочка	Stable
	Sense 3	Gold	Вольница, военный сброд	Stable
		Pred	Not identified	N/A

	Sense 4	Gold	Маленькие люди, чернь	Stable
		Pred	Not identified	N/A
	Sense 5	Gold	Сброд	Stable
		Pred	Not identified	N/A
	Sense 6	Gold	Экспрессивное восклицание	Stable
		Pred	Not identified	N/A
	Scores	Sense identification recall: 0.25 % of correctly identified lexical semantic change patterns: 100%		
публика	 			
	Sense 1	Gold	Читатели	Stable
		Pred	**Description:** Reader, the group of consumers for a certain written content, article, newspaper, book. **Signature:** читатель, общественность, журналист, писатель, критик, автор, профессионал, учёный, зритель, слушатель	Stable
	Sense 2	Gold	Аудитория, зрители, слушатели	Stable
		Pred	**Description:** Audience, in a theater, concert, cinema, TV, radio. **Signature:** зритель, журналист, поклонник, профессионал, критик, аудитория, жюри, организатор, мастер, автор	Stable
	Sense 3	Gold	Народец	Stable
		Pred	**Description:** Public, in the sense of open group of random people **Signature:** чиновник, общественность, журналист, избиратель, читатель, учёный, эксперт, пресса, наблюдатель, зритель	Stable
	Sense 4	Gold	Пассажиры	Stable
		Pred	Not identified	N/A
	Sense 5	Gold	Уличная толпа, масса	Stable
		Pred	Not identified	N/A
	Sense 6	Gold	Светское общество	Lost

		Pred	Not identified	N/A
	Scores	Sense identification recall: 0.5 % of correctly identified lexical semantic change patterns: 100%		
стиль	 			
	Sense 1	Gold	Способ летоисчисления	Stable (Freq. declined)
		Pred	**Description:** The way of tracking time/calendar **Signature:** календарь, весна, год, столетие, праздник, дата, старый, образец, рок, знак	Stable (Freq. declined)
	Sense 2	Gold	1. Черты, свойственные конкретному человеку (например, деятелю искусства) 2. Особенности направления архитектуры	Stable
		Pred	**Description:** A particular way of producing something, often art, architecture, poetry. **Signature:** дух, цвета, образ, жанр, материал, тон, вариант, масштаб, облик, контекст	Stable
	Sense 3	Gold	Not linked	N/A
		Pred	**Description:** A particular way of doing something. **Signature:** почерк, вектор, манер, выражение, стилистика, колорит, техника, видение, приём, контур	Stable
	Sense 4	Gold	Манера, совокупность особенностей по отношению к широкому кругу явлений (поведение, одежда, взгляды, внешность, интерьер)	Stable
		Pred	Not identified	N/A
	Sense 5	Gold	Характеристика языковых средств	Stable
		Pred	Not identified	N/A
	Scores	Sense identification recall: 0.6 % of correctly identified lexical semantic change patterns: 100%		
тройка	 			

	Sense 1	Gold	Лошади	Stable
		Pred	**Description:** Horses or a carriage with horses **Signature:** четвёрка, лошадь, пятка, машинка, пешком, лыжа, пара, конь, вдвоём, мотоцикл	Stable
	Sense 2	Gold	Костюм	New
		Pred	**Description:** Cloth **Signature:** шапка, майка, майк, футболка, куртка, маска, сумка, шуба, платье, блузка	Stable (Freq. increased)
	Sense 3	Gold	1. Количество, сумма из трех единиц 2. Три человека	Stable (Freq. increased)
		Pred	**Description:** Something of quantity three, including people **Signature:** четвёрка, двойка, пятка, третий, четырехка, семёрка, шестик, сотня, одиночка, второй	New
	Sense 4	Gold	Оценка в учебе	Stable
		Pred	Not identified	N/A
	Sense 5	Gold	Птица-тройка	Stable
		Pred	Not identified	N/A
	Sense 6	Gold	Игральная карта с тремя очками	Stable
		Pred	Not identified	N/A
	Sense 3	Gold	Not linked	N/A
		Pred	**Description:** Named entity ("Тройка Диалог", "Тройка Самара") **Signature:** группа, четвёрка, втб, альфа, альф, партнёр, третий, питер, сеть, титан	New
	Scores	Sense identification recall: 0.57 % of correctly identified lexical semantic change patterns: 25%		
червяк	 			
	Sense 1	Gold	Маленькое беспозвоночное животное	Stable
		Pred	**Description:** Worm, in the sense of animal **Signature:** червь, птица, животный, зверь, гриб, собака, кошка, камень, рыба, цветок	Stable
	Sense 2	Gold	Ничтожное создание	Stable
		Pred	**Description:** Weak, small create, used in a negative sense	Stable

			Signature: нищий, грешник, чертяка, мумия, человек,, жид, мертвец, гнома, гном, трус	
	Sense 3	Gold	Внутренний паразит	Stable
		Pred	Not identified	N/A
	Sense 4	Gold	Техническое приспособление	Stable
		Pred	Not identified	N/A
	Scores	Sense identification recall: 0.5 % of correctly identified lexical semantic change patterns: 100%		
знатный	 			
	Sense 1	Gold	Хороший	Stable
		Pred	**Description:** Good, decent person **Signature:** червь, птица, животный, зверь, гриб, собака, кошка, камень, рыба, цветок	Stable
	Sense 2	Gold	Принадлежащий к знати, высокий по чину	Stable
		Pred	**Description:** High rank person, respectable and powerful **Signature:** царский, известный, высокий, почётный, частный, светский, знаменитый, богатый, церковный, духовный	Stable
	Sense 3	Gold	Знаемый, известный, видимый	Stable
		Pred	**Description:** Well known, famous person **Signature:** известный, знаменитый, богатый, красивый, крупный, честный, видный, уважаемый, образованный, умный	Stable
	Sense 4	Gold	Существенный, серьезный (усилитель)	Stable
		Pred	**Description:** Good, decent thing, can be both material thing or not, for example time **Signature:** достойный, полезный, великий, редкий, добрый, богатый, значительный, хороший, сильный, приличный	Stable
	Sense 5	Gold	Техническое приспособление	Stable
		Pred	Not identified	N/A
	Sense 6	Gold	Выдающийся в труде	New
		Pred	Not identified	N/A

	Scores	Sense identification recall: 0.66 % of correctly identified lexical semantic change patterns: 100%		
пионер	 			
	Sense 1	Gold	Первооткрыватель	Stable
		Pred	**Description:** First in something, founder, leader **Signature:** организатор, автор, лидер, мастер, основатель, идеолог, ведущий, вести, наследник, носитель	Stable
	Sense 2	Gold	Сапер	Stable
		Pred	**Description:** Military profession **Signature:** полицейский, советник, солдат, охотник, командир, помощник, врач, старик, офицер, старшина	Lost
	Sense 3	Gold	Член детской организации	New
		Pred	**Description:** Member of student/school organisation **Signature:** школьник, учёный, воспитатель, мальчик, учитель, офицер, девочка, ветеран, волонтёр, пожарный	New
	Sense 4	Gold	Первый поселенец на какой-либо территории	Stable
		Pred	Not identified	N/A
	Scores	Sense identification recall: 0.75 % of correctly identified lexical semantic change patterns: 66%		