Interpretable approach to detecting semantic changes based on generated definitions

Tatarinov Maksim & Demidovsky Aleksandr

HSE University, Nizhny Novgorod

tatarinovst0@gmail.com, monadv@yandex.ru

GitHub

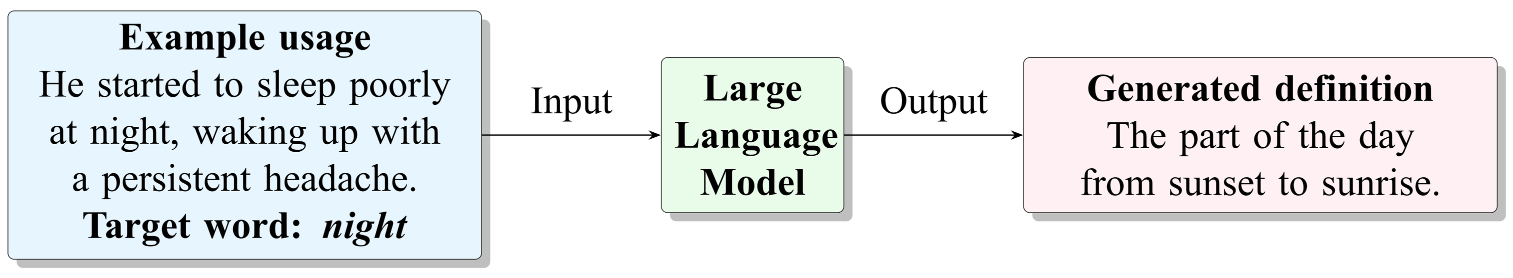
Introduction

Static and contextual embeddings excel at capturing semantic relationships for detecting semantic change, but lack human-readable word descriptions. Advancements in recent research involve definition generation with language models, which offer more illustrative descriptions (Giulianelli et al., 2023; Fedorova et al., 2024). It could aid historical linguists and lexicographers in creating dictionaries and language history studies, such as Dobrushina and Daniel' (2018). However, the practical evaluation of this approach remains limited.

The primary **objective** of this study is to **assess the effectiveness of language models in detecting semantic changes in words through the generation of definitions**. It would include:

- **quantitative metrics from a shared task**;
- **qualitative analysis by reproducing a linguistic analysis of words known to have undergone semantic shifts**.

Definition Modeling



Model and Data

FRED-T5-1.7B was chosen due to its performance in processing Russian (Zmitrovich et al., 2024). At the time of selection, it was the top performer on the RussianSuperGLUE benchmark (Shavrina et al., 2020), with a score of 0.762.

FRED-T5-1.7B was trained with the task of generating an accurate definition on a dataset where each entry contained a word, its context, and a corresponding definition.

Data

The training dataset was based on "Small academic dictionary" (Evgenyeva, 1981 1984). It was cleaned to:

- remove usage labels;
- entries without usage examples;
- entries without informative definitions, such as *Состояние по знач. глаг. линять* [*State by the meaning of the verb "to shed"*];
- those that provided grammatical rather than lexical information, such as *наречие к причастию приглашающий* [*Adverb to the participle "inviting"*].

This resulted in 122,350 entries, with 90% in the training part.

Shared Task Evaluation

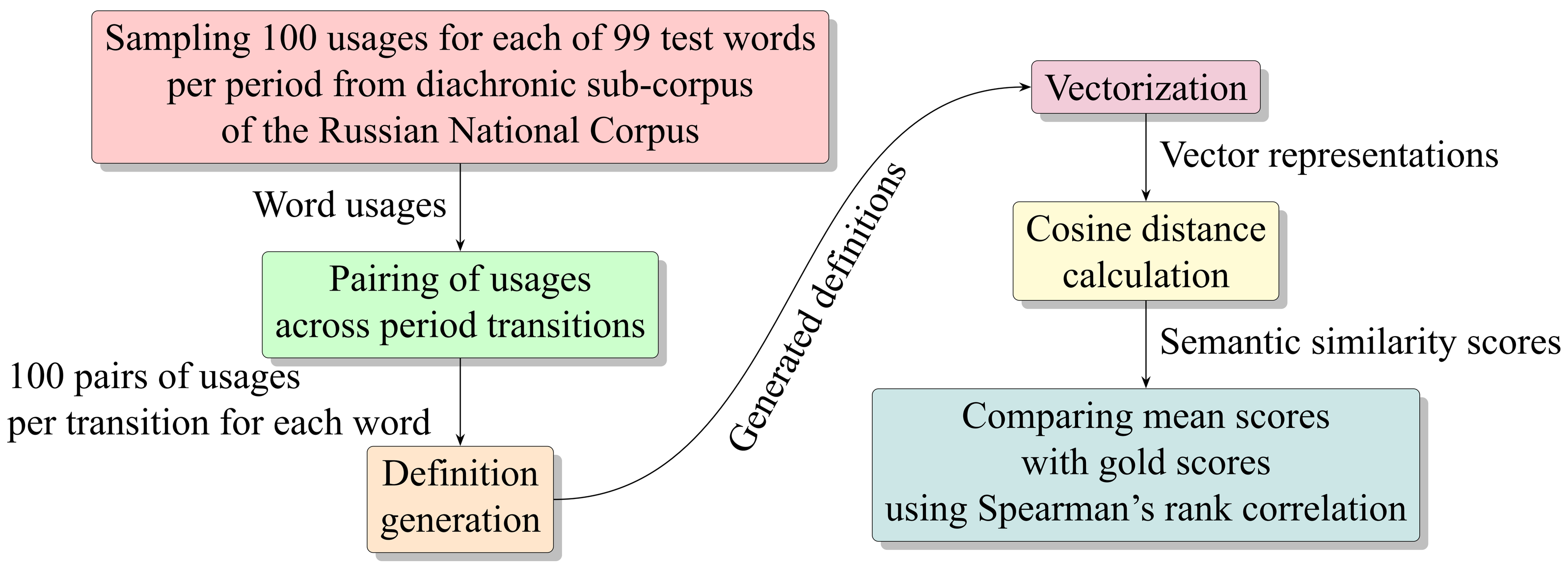
Data

The *RuShiftEval* task's test set (Kutuzov and Pivovarova, 2021) was utilized for evaluation. The task focuses on detecting semantic changes in Russian nouns across three historical transitions:

- RuShiftEval-1 (Pre-Soviet:Soviet);
- RuShiftEval-2 (Soviet:Post-Soviet);
- RuShiftEval-3 (Pre-Soviet:Post-Soviet).

The task provided a test set of gold change scores for 99 Russian nouns corresponding to the transitions.

Pipeline



Results

The paraphrase-multilingual-mpnet-base-v2 model (Transformers, 2023), additionally fine-tuned on RuSemShift, a similar dataset (Rodina and Kutuzov, 2020), was used to vectorize definitions.

Team	Average	Word Representation Type	Model Used
DeepMistake (post-competition)	**0.850**	**Contextual Emb.**	**XLM-R**
Proposed Approach	**0.815**	**Generated Definitions**	**FRED-T5-1.7B**
GlossReader	0.802	Contextual Emb.	XLM-R
DeepMistake	0.791	Contextual Emb.	XLM-R
Other 11 Teams	0.720-0.178	…	…

Table 1: Algorithm Results Compared to RuShiftEval Teams

The proposed approach outperforms most entries in the RuShiftEval task, as shown in Table 1.

Method	RuShiftEval-1	RuShiftEval-2	RuShiftEval-3	Base Model
Proposed Approach without vectorizer fine-tuning	**0.722**	**0.763**	**0.749**	**FRED-T5-1.7B**
Fedorova et al. (2024)	0.488	0.462	0.504	MT0-XL

Table 2: Comparison with Definition Generation Approaches

A shown in Table 2, the proposed approach significantly outperforms the results of Fedorova et al. (2024). The vectorizer fine-tuning step was omitted to ensure compatability of the results.

Fedorova et al. (2024) appears to retain unhelpful definitions in the training data, possibly resulting in their model reproducing non-informative patterns and the lower performance of their approach.

Qualitative Evaluation

20 words exhibiting shifts from *Two Centuries in Twenty Words* (Dobrushina and Daniel', 2018) are selected: *знатный [noble]*, *кануть [to disappear]*, *классный [classy/cool]*, *мама [mom]*, *машина [machine/car]*, *молодец [young man/attaboy]*, *пакет [bag/package]*, *передовой [advanced]*, *пионер [pioneer]*, *пожалуй [perhaps]*, *пока [until/bye]*, *привет [hello]*, *пружина [spring]*, *публика [public]*, *свалка [landfill/fight]*, *сволочь [bastard]*, *стиль [style]*, *тётка [aunt]*, *тройка [three/a set of three]*, *червяк [worm]*. 300 usage examples per word per period are obtained from the diachronic subcorpus of Russian National Corpus (Savchuk et al., 2024). The trained model generates definitions for each word usage. To obtain meanings and changes in their usages, a visualization algorithm is applied. The definitions from the visualization are compared with information from semantic descriptions of words, written based on Sternin and Rudakova (2017) method of generalizing multiple dictionary definitions, and classified according to the error types. Finally, changes in the frequency of meanings over time provided by the visualizations are compared with historical data from *Two Centuries in Twenty Words*.

Visualization

Generated definitions are transformed into vector representations using a vectorizer and clustered using DBSCAN. Each cluster is represented by a prototypical definition closest to its centroid.

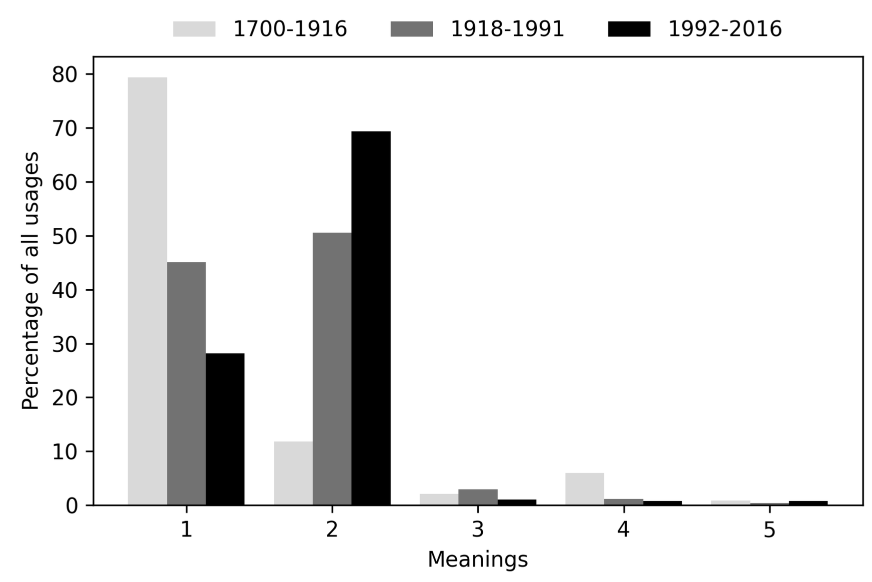


Figure 1: Semantic Shift of the Word *машина [machine/car]* (Parameters: eps=0.14, min_samples=5)

Meanings for *машина [machine/car]*:

1. A device or instrument for a specific task.
2. An automobile or vehicle.
3. An aircraft or helicopter.
4. A mechanically or thoughtlessly acting person.
5. A system, a collection of institutions or organizations.

Results

As a result of generalizing dictionary definitions, 121 meanings were compiled for 20 words. 83 definitions were obtained using the proposed approach. Thus, excluding 5 incorrect definitions, out of 121 meanings 64.4% were identified.

Type of Definition	Count	Percentage
Correct	57	68.67%
Close	10	12.04%
Incorrect	5	6.02%
Insufficiently Specific	3	3.61%
Redundancy or Excessive Use of General Phrases	4	4.81%
Close, Redundancy or Excessive Use of General Phrases	1	1.20%
Overly Specific	3	3.61%
Self-reference	0	0.00%
Opposite Meaning	0	0.00%
Incorrect Part of Speech	0	0.00%

Table 3: Types of Definitions and Their Counts

The majority of definitions are correct without any errors or shortcomings (68.67%).

Results	Word amount
Correct	12
Partially correct	4
Incorrect	1
Full analysis not possible	3

Table 4: Results of Statistical Analysis of Semantic Shifts

Overall, main meaning changes consistent with the book's data were identified in 12 out of 20 words. Additionally, changes partially aligned in 4 other words. A full analysis is not feasible for *публика [public]*, *кануть [to disappear]* and *сволочь [bastard]* due to *Two Centuries in Twenty Words* not having sufficient diagrams or detected meanings falling under one in the book.

Conclusions

- The study demonstrated the effectiveness of definition modeling in detecting semantic shifts.
- A FRED-T5-1.7B model, fine-tuned on the "Small academic dictionary", was used to generate context-based word definitions.
- The model performed among the top solutions of RuShiftEval and outperformed Fedorova et al. (2024).
- A visualization algorithm was developed to represent semantic changes over time, allowing for reproducing a manual effort of studying semantic changes for a set of 20 words.
- Qualitative analysis revealed that 68.67% of generated definitions were fully correct, with main changes fully detected in 12 out of 17 words available for analysis and partial alignment in 4.

References

N.R. Dobrushina and M.A. Daniel'. 2018. *Dva veka v dvadtsati slovakh [Two Centuries in Twenty Words]*. Izdatel'skiy dom Vysshey shkoly ekonomiki, Moskva, 2 edition.

A.P. Evgenyeva. 1981-1984. *Slovar' russkogo yazyka: V 4-kh t. [Dictionary of the Russian Language: In 4 Volumes]*. Russkiy yazyk, Moskva, 4-e izd., ispr. i dop [4th ed., corrected and supplemented] edition. V 4-kh tomakh [In 4 volumes].

Mariia Fedorova, Andrey Kutuzov, and Yves Scherrer. 2024. Definition generation for lexical semantic change detection. // Lun-Wei Ku, Andre Martins, and Vivek Srikumar, *Findings of the Association for Computational Linguistics: ACL 2024*, P 5712–5724, Bangkok, Thailand, August. Association for Computational Linguistics.

Mario Giulianelli, Iris Luden, Raquel Fernandez, and Andrey Kutuzov. 2023. Interpretable word sense representations via definition generation: The case of semantic change analysis. // Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, P 3130–3148, Toronto, Canada, July. Association for Computational Linguistics.

Andrey Kutuzov and Lidia Pivovarova. 2021. Rushifteval: a shared task on semantic shift detection for russian. // *Computational linguistics and intellectual technologies: Papers from the annual conference Dialogue*, P 533–545.

Julia Rodina and Andrey Kutuzov. 2020. RuSemShift: a dataset of historical lexical semantic change in Russian. // Donia Scott, Nuria Bel, and Chengqing Zong, *Proceedings of the 28th International Conference on Computational Linguistics*, P 1037–1047, Barcelona, Spain (Online), December. International Committee on Computational Linguistics.

S.O. Savchuk, T.A. Arkhangelskiy, A.A. Bonch-Osmolovskaya, O.V. Donina, Yu.N. Kuznetsova, O.N. Lyashevskaya, B.V. Orekhov, and M.V. Podryadchikova. 2024. Natsionalny korpus russkogo yazyka 2.0: novye vozmozhnosti i perspektivy razvitiya. *Voprosy Yazykoznaniya*, 2:7–34.

Tatiana Shavrina, Alena Fenogenova, Emelyanov Anton, Denis Shevelev, Ekaterina Artemova, Valentin Malykh, Vladislav Mikhailov, Maria Tikhonova, Andrey Chertok, and Andrey Evlampiev. 2020. RussianSuperGLUE: A russian language understanding evaluation benchmark. // *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics.

I.A. Sternin and A.V. Rudakova. 2017. *Slovarnye definicii i semanticheskiy analiz [Dictionary Definitions and Semantic Analysis]*. Istoki, Voronezh.

Sentence Transformers. 2023. paraphrase-multilingual-mpnet-base-v2. https://huggingface.co/sentence-transformers/paraphrase-multilingual-mpnet-base-v2. Retrieved April 19, 2024.

Dmitry Zmitrovich, Aleksandr Abramov, Andrey Kalmykov, Vitaly Kadulin, Maria Tikhonova, Ekaterina Taktasheva, Danil Astafurov, Mark Baushenko, Artem Snegirev, Tatiana Shavrina, Sergei S. Markov, Vladislav Mikhailov, and Alena Fenogenova. 2024. A family of pretrained transformer language models for Russian. // Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, P 507–524, Torino, Italia, May. ELRA and ICCL.