ISSN 2007-9737

Prediction of Lexical-Semantic Relations in a Low Resource Language: From Word2Vec To LLM

Lizaveta Astapenka[1], Olga Mitrofanova[2,*]

[1] Independent Researcher,
Belarus

[2] St. Petersburg State University,
Department of Mathematical Linguistics,
Russia

li.astapenka@gmail.com, o.mitrofanova@spbu.ru

Abstract. This paper examines the ability of language models to capture semantic relations between words in a low-resource language. We describe experiments on automatic prediction of lexical-semantic relations in Belarusian using models of Word2Vec, BERT, and LLM families which differ in neural architecture, feature types and NLP applications. Training and fine-tuning of the models was carried out on the datasets compiled for our study: Belarusian corpora with UD POS tagging and a database of synonyms and antonyms extracted from Belarusian dictionaries. Model performance was evaluated by pseudo-disambiguation test (Word2Vec CBOW and skip-grams) as well as by expert assessments (roberta-small-Belarusian, Gemini 2.5 Pro). The results proved to be valid and can be applied to create and enrich lexical databases, to analyse word co-occurrence, to improve machine translation, paraphrasing, summarization, and other systems related to automatic processing of the Belarusian language.

Keywords. Lexical-semantic relations, the Belarusian language, Word2Vec, BERT, large language models.

1 Introduction

Complex neural network architectures underlying modern language models enable them to capture not only morphological and syntactic, but also semantic properties of language units. Automatic extraction of lexical-semantic relations from text corpora is considered to be important in regard to the systematic description of vocabulary especially for low-resource languages. The Belarusian language is represented in a set of corpus resources and NLP tools. There are various online dictionaries for the Belarusian language (for example, Verbum, Slounik), morphological and syntactic taggers (Stanza, UD-Pipe). In 2013, the lemmatizer YABC_Tagger was implemented as part of the "Yet Another Belarusian Corpus" project [31].

Various tools for tagging, parsing and information extraction for Belarusian (lemmatization, text summarization, etc.) are available on www.corpus.by platform. Belarusian grammar database and the National Corpus ("Belaruski N-corpus") are developed in the Institute of Linguistics named after Yakub Kolas [13]. NLP tools for Belarusian can be found on the Github repository (navalnica/be_nlp_speech_resources). However, to date, language models trained on Belarusian corpora for lexical semantics tasks are not yet publicly available. Such models can be used in machine translation, linguodidactics, information retrieval, as well as in text simplification, automatic text style transfer, sentiment analysis, etc.

Our main objective was to train and evaluate language models of the Belarusian language capable of predicting paradigmatic and syntagmatic relations observed in text corpora. Thus, our study fills the gap in the Belarusian NLP. The paper is structured as follows: Section 2 is devoted to related work in lexical-semantic prediction, Section 3 dwells upon dataset collection and annotation, Section 4 presents model choice and experimental design, Section 5 contains discussion of results, and Section 6 summarizes our research.

2 Related Work

Multiple studies in computational semantics focused on text and meaning embeddings point out which semantic relations between lexemes are captured by language models and how they can be differentiated. In our research, we consider three main types of models: Word2Vec [17], BERT [6], and LLMs (Large Language Models).

Distributional semantic models identify both syntagmatically and paradigmatically similar words [1] [3] [30]. The study [30] showed that the output of models for the target word includes both its synonyms, antonyms, hyponyms, hyperonyms, meronyms, and holonyms. However, the disadvantage of these models is the lack of any differentiation between relations. Both synonyms and antonyms, for example, receive high cosine similarity due to the contextual resemblance. Nevertheless, researchers proposed several methods for distinguishing semantic relations in models.

One possible approach is to use patterns derived from syntactic parse trees. These patterns are vectorized using a recurrent neural network with long short-term memory (LSTM) and fed into the classifier to determine a specific semantic relation (synonymy or antonymy). The authors also propose a second option, in which a vector representation is formed by combining word vectors and a pattern vector [19]. However, the disadvantage of the described approach is that it requires a large amount of training sentences that simultaneously include both words from a pair.

[7] focus on changing the loss function. They combine three functions: the first is the initial Skip-gram loss function, which is modified so that the cosine measure for synonyms should be close to 1, and for antonyms to -1; the second function uses information about synonyms and antonyms added from an external thesaurus; the third function concerns sentiment information from the SentiWordNet resource. As an alternative approach to differentiating between synonymy and antonymy, it was proposed to use the Siamese neural network [9] [26]. It consists of two identical neural subnetworks that have identical sets of weights. This architecture makes it possible to compare feature vectors of any two objects and determine whether they are similar. There is also another type of such neural networks, when the input is not pairs of objects, but triplets, including anchor, positive (similar to the target), and negative objects. The study [9] describes the idea of two-stage Siamese neural network training. [26] presents a simplified version of this algorithm leading to improved embeddings that take into account semantic properties of words. In this case ”word — synonym — antonym” triplets are used as training data. In our work, we implement this method, since it is applied to vectors of an already trained baseline model (Word2Vec, etc.) and it does not require a large amount of any special training data.

Many researchers investigate how well BERT models are able to capture morphological, syntactic, and semantic relations between words in a language [24] [32]. In [29], models were tested on eight different tasks: POS-tagging, named entity recognition, semantic roles and syntactic dependencies tagging, coreference resolution, etc. The authors conclude that contextualized embeddings do encode information related to semantic roles, relations, entity types, etc., but models with such vector representations are not always superior to other types of models in semantics-related tasks, although they are significantly better at establishing syntactic relations. The study [25] considers the possibility of fine-tuning BERT models to classify semantic relations such as ”is a”, ”part of”, ”property of”, ”made of”, etc. The study [27] shows that the attention mechanism enables BERT models to identify not only synonymous and antonymic relationships, but also factual information.

To automatically select synonyms for a particular word in a particular sentence, researchers [10] used a method for comparing both the embeddings of the target word and its possible substitutes, as well as the distances in the vector space between the original sentence and the sentence with the completed substitution. In [34], it is proposed to partially mask the target word before the stage of comparing embeddings. The study [8] examines the possibility of using BERT to predict syntagmatic relations, in particular collocations and lexical functions. The authors conducted several experiments: with masking the collocation in a sentence, with adding the original sentence without a mask as a context, as well as with additional training on data of the type <sentence — collocation — lexical function>. The results showed that the model does predict collocations of the form "verb + noun", but there are significant problems with combinations of "adjective + noun".

In our study, we predict using BERT synonymous substitutes of nouns, adjectives, verbs, and adverbs in the context, since the architecture of such models seems to be the most suitable for this task.

Large language models (LLM) are capable of processing and generating not only texts, but also video, audio, and images. Numerous studies are conducted to evaluate the capabilities of models to solve problems related to various fields, including lexical semantics. For example, in [18] researchers analyze the abilities of the LLaMA-2 and Mistral models in constructing taxonomies, in particular in predicting hyponyms, hyperonyms, etc. The paper [33] compares how different types of models (Positive Pointwise Mutual Information (PPMI), Skip-gram with negative sampling (SGNS), BERT, GPT-4) determine the change in word meanings over time. As for PPMI, SGNS, and BERT, the cosine measure for the target word vectors was calculated at different time intervals, and then the correlation with human annotation was determined. For GPT-4, a prompt was created in such a way that the model returned either 0 or 1, depending on whether there is a change in the meaning in this context. The authors conclude that GPT-4 is superior to all other methods considered. Multilingual LLMs work particularly well with English, but they turn out to be less effective in processing low-resource languages due to the amount of training data [22].

In our study, we evaluate the ability of LLMs to generate synonyms for words in Belarusian.

3 Research Data

3.1 Corpus of the Belarusian language

A Belarusian corpus was compiled to train Word2Vec models and provide training data for BERT. The text collection includes three segments: the corpus of Belarusian from the Universal Dependencies (UD) project; Belacorpus; the corpus of Belarusian from the collection of the University of Leipzig (Web 300K).

— The UD corpus (305,417 tokens, 25,231 sentences) is a fragment of a parallel Belarusian-Russian subcorpus within the Russian National Corpus (RNC) and covers texts of various genres: fiction (authors: Frantsishak Bagushevich, Yanka Kupala, Maxim Garetsky, Vasil Bykau, Ivan Melezh); fragment of the novel "The Lord of the Rings" by J. R. R. Tolkien translated into Belarusian; news articles; messages from social networks (from Telegram channels (artsiadziba, belarusian_history, etc.)); texts from Wikipedia [28].

— The second part of our research corpus is Belacorpus [15]. The corpus size is 246 text files, 1,535,047 tokens. It includes texts of various genres for the period 1987-2010: fiction, articles from magazines and newspapers, legal texts, etc.

— Texts from the collection of the University of Leipzig were used as the third fragment of the corpus. To achieve greater representativeness, the Web-300K-2015 corpus was selected. It includes 300 000 sentences and is compiled on the basis of various websites in the Belarusian language.

Belacorpus and Web-300K-2015 were lemmatized and annotated by using the Stanza NLP tool. The quality of the morphological annotation was checked manually. The accuracy of the annotation is quite high ($\geq$ 0.85), however, in some cases, the lemma or part of speech is still incorrectly determined. Mistakes are most often found in words with alternation. For example, in the third-person singular verb form «вернецца» ("(he/

ISSN 2007-9737

she/ it) will return"), the lemma was automatically defined as «вернуцца_VERB», although the correct variant is «вярнуцца_VERB» with an accent on the second syllable. Similarly, the alternation is ignored when annotating some nouns. For example, the lemma «пчола_NOUN» is incorrect (the correct lemma is «пчала_NOUN» ("bee")). In addition, the system probably truncates some lemmas by analogy with a number of other words («шчак_NOUN» instead of «шчака_NOUN» ("cheek")). Errors also occur when trying to lemmatize complex words (the algorithm leaves the form «бібліятэкі-філіяла_NOUN» instead of «бібліятэка-філіял_NOUN»). As for POS tagging, proper names are sometimes not recognized, so they are assigned the tag _NOUN instead of _PROPN (for example, «ірэначка_NOUN» instead of «Ірэначка_PROPN»).

After preprocessing and combining all the files into a single corpus, tokens with the PUNCT, SYM, and X tags (i.e. punctuation marks, various symbols (for example, %), and foreign language elements) were removed. At this stage, it was decided not to filter the corpus by form words in order to preserve the original relations between words in sentences, but to remove them only if they affect the performance of the model. In addition, with the help of regular expressions, the corpus was cleared of dates, bibliographic references, and strings consisting only of numbers or Latin alphabet characters. Short lines that were less than four words long were also deleted. The final size of the corpus is 413,012 sentences (5,662,829 tokens).

3.2 Datasets with Synonyms and Antonyms

In addition to the corpus, our research requires datasets representing dictionaries of synonyms and antonyms in a machine-readable format. For the Belarusian language, such datasets have not yet been publicly available, so it was decided to parse web pages that contain dictionary materials.

sing the BeautifulSoup library data was extracted from the Verbum website (an online dictionary of synonyms) [12]. The dataset includes 1939 synonymic rows. The dictionary of antonyms is not available online, so for parsing, the

Table 1. Example of the output data for the word «мова» ("language")

Skip-gram, ADJ
англійскі_ADJ ("English")
рускі_ADJ ("Russian")
украінскі_ADJ ("Ukrainian")
царкоўнаславянскі_ADJ ("Church Slavonic")
старабеларускі_ADJ ("Old Belarusian")

«Слоўнік лексічных формаў (сінонімы, амонімы, антонімы, паронімы, амографы, амафоны)» [11] was downloaded in DJVU format from the website of the Belarusian electronic library «Беларуская Палічка». Using online services, the file was converted to pdf format, then we applied optical character recognition (OCR) technology to it. Using regular expressions and the PyPDF2 library, a total of 597 antonymic pairs were extracted from the dictionary. After that, the final dataset was checked manually. The resulting list of antonyms was expanded by automatically selecting synonyms for them from the data tables we compiled. The total number of antonymic pairs increased to 1624.

4 Models and Experimental Setup

4.1 Word2Vec Models

Several Word2Vec models were trained with different hyperparameters (vector_size, window, min_count, epochs). To obtain the best model, we relied primarily on the pseudodisambiguation method.

To differentiate between paradigmatic and syntagmatic relations, we propose POS-filters. For example, for nouns, we can automatically select verbs and adjectives that are combined with them in contexts. Tables 1 and 2 show the semantic associates provided by the skip-gram and CBOW models for the word «мова» ("language").

In conditions of limited resources (both computational and linguistic), trained Word2Vec models, despite sufficiently high accuracy in evaluation, may have a number of disadvantages. For

Table 2. Example of the output data for the word «мова» ("language")

CBOW, NOUN
правапіс_NOUN ("spelling")
філалогія_NOUN ("philology")
літаратура_NOUN ("literature")
арфаграфія_NOUN ("orthography")
маўленне_NOUN ("speech")

example, if the frequency of the word for which we want to get semantic associates turns out to be low in our corpus, then the model may make irrelevant predictions. In addition, vectors of words that are considered synonymous can be significantly distant from each other due to the small number of sentences indicating their similarity to the model. Both synonyms and antonyms can receive high cosine values due to similar contexts in which words are used. To improve the quality of predictions, we decided to focus our research on a method involving the use of additional lexical databases and Siamese neural networks. The Siamese Semantic Vectors algorithm [14] was used as a basis. It implements the training of a Siamese neural network for the English language using the WordNet thesaurus and the GloVe model. In our case, the Word2Vec Skip-gram model was used because, compared to CBOW, it is more context-sensitive, which allows it to better capture the semantic relations between words [17].

The training dataset consisted of lists of synonyms and antonyms for the Belarusian language based on dictionary materials. Thus, the total number of pairs of synonyms was 4038, and the number of antonyms made up 1624. The data was then divided into training and test samples in a percentage ratio of 80:20.

Keras library was used to train the neural network. The algorithm transfers the vector representations of the model to a new space in which the cosine measure will be closer to 1 between synonyms and to -1 between antonyms. The Siamese neural network consists of two subnets with the same weights that process data in parallel. The loss function tracks the degree of error between the predicted and actual target values. The input data are vectors of synonymous, antonymic, and neutral pairs (random combinations of words that are neither synonyms nor antonyms). Actual values are -1 (for antonyms), 1 (for synonyms), 0 (for neutral pairs). After training the neural network, the transformation is applied to all vectors of the dictionary.

4.2 BERT

The ability of BERT models to construct contextualized embeddings is particularly useful in the task of selecting synonymous substitutes of words in sentences. The task of predicting synonyms by the model can be reduced to the task of masked language modeling, when BERT suggests the most likely words in place of a special token [MASK] in the sequence. Therefore, the model should predict for a specific context not just possible options for filling the position of a masked token, but synonymous ones.

We used "word—sentence" pairs as training data, and the contexts were duplicated for lexemes from the same synonymic rows so that the model can understand that these words are interchangeable in this sentence. At first, we selected contexts from our corpus for the target words, and then manually identified the appropriate synonyms. The result was a new table with three columns (word — context — synonyms), which was then converted into two json format files with duplicate contexts. Thus, the amount of training data was 1203 pairs. The test dataset included 100 sentences (25 contexts for 4 main parts of speech: nouns, adjectives, verbs, adverbs).

Before the fine-tuning process, we tested the ability of several models to predict words in the position of a masked token. There are several multilingual and monolingual pre-trained models for the Belarusian language, among which we selected two that most correspond to the goals of our research, namely **xml-roberta-large** (94 languages, 561 million parameters) and **roberta-small-Belarusian** (1 language, 15.7 million parameters). The disadvantage of the first

ISSN 2007-9737

model is that it often predicts irrelevant words from Russian instead of Belarusian. In this regard, monolingual models turn out to be better, despite a very small number of parameters. Thus, the second model was used in all further experiments. We set up hyperparameters (num_train_epochs = 6, per_device_train_batch_size = 16, learning_rate = 5e-5) and trained two models: on data with lemmas and on data with word forms. The results obtained using the second model turned out to be more suitable.

Since our capabilities to expand the dataset and train a new model with an increased number of parameters are limited, it was necessary to use additional methods to improve the quality of predictions. Following [10] and [34], we decided to use the **Sentence Transformer (SBERT)** [23] for sentence embedding comparison, but before that, to add the synonyms indicated in the dictionary to the list of candidates already proposed by the model. Since there are no SBERT models for the Belarusian language, the multilingual models **LaBSE** (110 languages, 471 million parameters) and **paraphrase-multilingual-MiniLM-L12-v2** (50 languages, 118 million parameters) were selected. It was verified on the test sample that the first model copes with the task better. Thus, the final algorithm includes several stages. At first, the pre-trained roberta-small-Belarusian model receives a sentence with a special token [MASK] and predicts possible words that could stand in its place. Candidates from the dictionary are also added to the list. Then, in place of the mask, words from the list are inserted into the sentence in turn. The LaBSE model compares the sentence embeddings and ranks candidates based on the value of the cosine measure. In addition, we added filters to avoid punctuation marks in the results, and also introduced a limit on the minimum number of characters per word (len(word) ≥ 3).

4.3 Large Language Models (LLMs)

LLMs are actively used to solve a wide variety of tasks. We decided to evaluate how well the initial models can predict synonyms in the Belarusian language without any additional training. The work of the models was tested on a test dataset. Among the multilingual LLMs, the most well-known and currently available ones were selected, namely **DeepSeek-V3**, **GPT-4o**, **Claude 3.5 Haiku**, **Gemini 2.0 Flash**, **Gemini 2.5 Pro**. Then the basic prompt in the Belarusian language was compiled. The model was supposed to suggest substitutes for certain words in sentences, following the example.

Spelling and grammar errors are one of the most common problems (for example, «рэшучыя»* instead of «рашучыя», «рознастайныя»* instead of «разнастайныя», attempts to form the plural of «палымя»* in its absence, etc.). In addition, models often generate calques from Ukrainian and Russian («шчырасэрдна»* — ukr. «щиросердно», «душэраздзіральны»* — rus. «душераздирающий» etc.). Also, models sometimes propose stylistically colored words as synonymous substitutes that are not suitable for a neutral context («дурны» (silly) — colloq. «ідыёцкі»). Thus, based on the analysis of errors and correspondence between the proposed variants and the dictionary data, the **Gemini 2.5 Pro** model was selected for further expert evaluation.

5 Results and Discussion

5.1 Word2Vec Models

The Word2Vec model performance was assessed using the pseudo-disambiguation procedure [2] [4] [5] [20] [21] for collocations ”ADJ + NOUN” (adjective + noun). The test dataset was compiled manually and included 2 sets of trigrams (target word — correct collocate — incorrect collocate):

— A target noun (дзень_NOUN day) — an adjective that occurs in the same context with the given noun in the corpus (новы_ADJ new) — an adjective that does not combine with the given noun (незалежны_ADJ independent);

— A target adjective (e.g. вядомы_ADJ famous) — a noun that occurs in the same context with the given adjective in the corpus (e.g. фільм_NOUN movie) — a noun that does not combine with the given adjective (сярэдзіна_NOUN middle).

100 most frequent nouns and 100 most frequent adjectives in the corpus were taken as target words. All sentences containing correct collocations were removed from the training corpus. The task of the model was to determine correct collocations.

According to the results of pseudodisambiguation, we chose two models that achieve high accuracy values for lists with both the target noun and adjective (78% and 90% for the skip-gram model, 83% and 92% for the cbow model): 1 — vector_size = 250, window = 5, min_count = 5, epochs = 5, sg = 1 (skip-gram algorithm), 2 — vector_size = 300, window = 5, min_count = 5, epochs = 5, sg = 0 (CBOW algorithm). To visualize the performance of the models a web-application[1] was developed on the Hugging Face platform.

To evaluate the models before and after training the neural network, the metric was used that measured how much of all synonyms (antonyms) the model had determined correctly. All predictions of the cosine measure from 0.6 to 1 for synonyms and from -0.2 to -1 for antonyms were considered correct.

Figure 1 shows the distribution graph after the differentiation between synonymous and antonymic relations. Figure 2 shows the graphs of the distribution of cosine similarity values for pairs of synonyms and antonyms obtained before training the neural network.

The graphs demonstrate that it is possible to distinguish between synonymy and antonymy using the Siamese neural network. While initially the cosine measure for pairs of antonyms took only positive values, then after training the model, the distribution shifts to a range with negative numbers. The proportion of pairs for which the cosine measure exceeds 0.6 reached only 0.4 (40%) for both training and test data before adding the neural network.

[1] https://huggingface.co/spaces/lizaastapenka/lexical-semantic-calculator-for-the-Belarusian-language

5.2 BERT

Since it is difficult to assess automatically how suitable a particular substitute is, it was decided to use the expert assessment method. We compiled two questionnaires to evaluate the model before[2] and after[3] the fine-tuning procedure. All data was randomly divided into five samples (20 sentences with 5 predictions for each). To assess how well the model captures the semantic relations between lexemes, the experts should give scores from 0 to 5 for each word according to the following criteria:

5 — the target word;

4 — synonym (including if it is a word that is synonymous only in this context);

3 — an antonym (including if it is a word that is an antonym only in this context);

2 — other paradigmatic relationships (for example, part/whole relations; words included in the same associative field, etc.);

1 — the word is related to the target word in terms of word formation;

0 — there is no connection between the words + all cases of incorrect output (for example, fragments of words that do not make sense).

Native speakers of the Belarusian language (mostly students of Belarusian universities: Yanka Kupala State University of Grodno, Belarusian State University of Foreign Languages) were experts. The total number of informants for each of the questionnaires was 25 people. Based on the results obtained, the experts' consistency, average scores and relative values were determined. The maximum possible score for a single sentence was 21 points, meaning the model had to predict the target word and its four synonyms.

Therefore, the maximum score for the entire sample was 420 points. The experts' consistency was calculated using the Kappa Fleiss formula. Tables 3 and 4 show the results for the model before and after training (along with the additional operations).

In Figure 3, the diagram shows the difference in the average scores for each of the samples.

[2] https://forms.gle/QasBKseLBZhxznEs7
[3] https://forms.gle/DPwqAac8YNcgTcs19

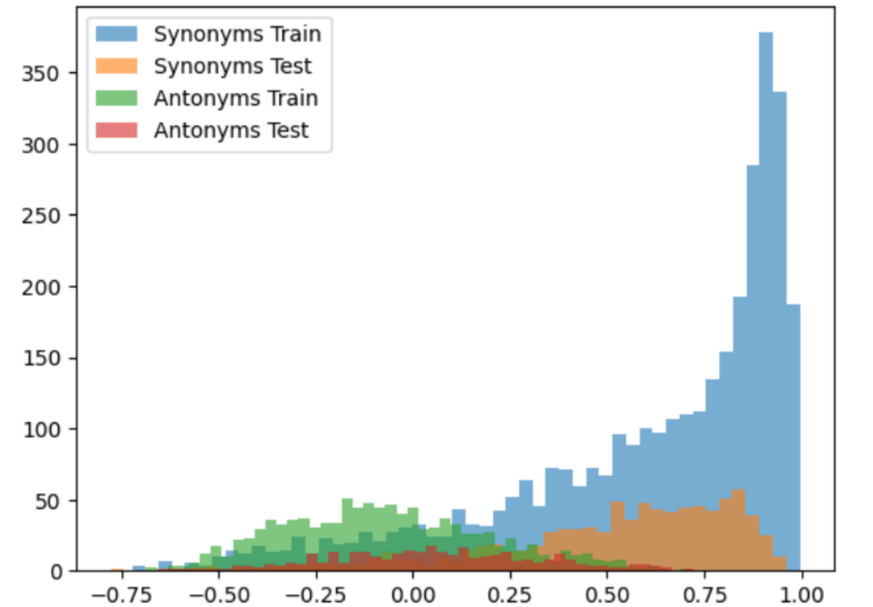

Fig. 1. Distribution of cosine similarity values for pairs of synonyms and antonyms **after** training the neural network

Table 3. Model evaluation before fine-tuning

Sample	Number of people	Average score	Relative value	Consistency of experts' opinions
1	5	118,6	0,28	0,83
2	5	87	0,21	0,81
3	5	127,6	0,30	0,87
4	5	84,4	0,20	0,79
5	5	98	0,23	0,88

The accuracy of predictions has increased from an average of 24% to 75%. The Wilcoxon signed rank test was applied to verify the statistically significant difference between average scores for each sentence (before and after the fine-tuning process).

The results ($Z = -8.54$, $p\_value < 0.001$) proved that there is a significant difference between the rates. The consistency value varies from 0.75 to 0.88. It is considered significant in the humanities [16].

In the fifth sample, the final model shows the highest scores (352 out of 420). This can be explained by the fact that in this group of sentences, there were many adjectives and adverbs as target words. It is easier for the model to identify synonyms for them.

The evaluation results show that the task of predicting synonyms in the context of the Belarusian language using the BERT model is solved with fairly high accuracy.

Nevertheless, several disadvantages can be noted. Due to the peculiarities of BERT tokenization and the relatively small number of parameters of the pre-trained model, in some cases the output data may contain fragments rather than whole words («натхня» instead of «натхняць»).

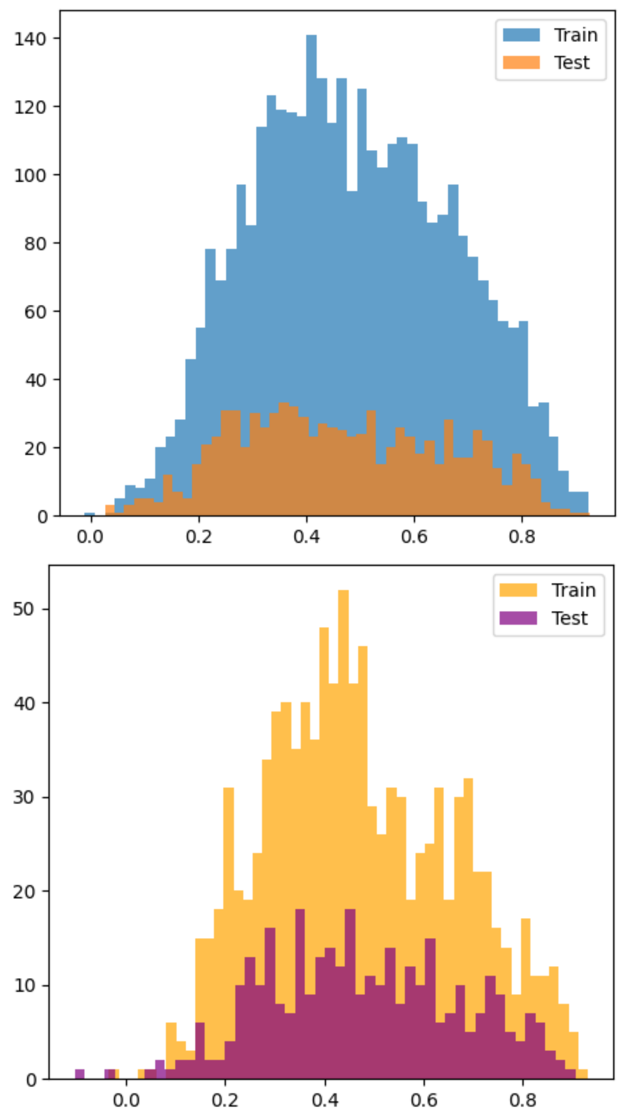

Fig. 2. Distribution of cosine similarity values for pairs of synonyms (above) and antonyms (below) **before** training the neural network

In addition, lemmas from the dictionary are also included in the list of predictions in their initial form, regardless of the context. To solve this problem, we need a system of morphological synthesis for the Belarusian language.

There is also a limitation on the quality of the sentences themselves. If the context is too general, it will be difficult for the model to guess even the target word, so the results may contain inappropriate substitutes.

5.3 Large Language Models (LLMs)

We compiled a questionnaire[4] to assess how appropriate the synonyms generated by Gemini

[4]https://forms.gle/UUq8NzpoFCB8PYCF6

Table 4. Model evaluation after fine-tuning

Sample	Number of people	Average score	Relative value	Consistency of experts' opinions
1	5	304	0,72	0,8
2	5	305,8	0,73	0,81
3	5	308,4	0,73	0,85
4	5	308	0,73	0,77
5	5	352	0,84	0,75

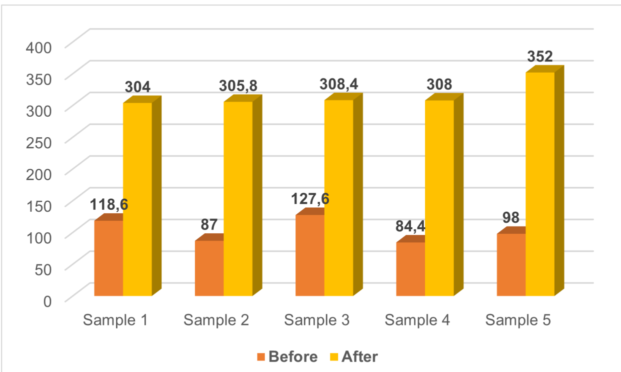

Fig. 3. Average scores (before/ after fine-tuning)

2.5 Pro are in terms of meaning. The dataset was divided into 5 samples of 20 sentences each. The experts were asked to evaluate each response as follows:

0 — this word is NOT a synonym and CANNOT be used as a substitute (including cases where the word is stylistically inappropriate) / such a word does not exist in the Belarusian language / the word is misspelled;

1 — this word is a synonym (including contextual) and can replace the target word without losing its meaning in this sentence.

We considered the maximum possible score to be 100 points (5 points for all correct substitutions in one sentence). For each sample, the average score was calculated, as well as the consistency of the expert opinions (using the Kappa Fleiss coefficient, as in the case of BERT models). Table 5 shows the results obtained.

The expert group included native speakers of the Belarusian language (84 people). Consistency takes values in the range from 0.49 to 0.59, which is a reliable result for the humanities [16]. The average score varies between 50 and 60 points for all samples. It indicates that the model does not fully cope with the task of generating synonyms in contexts for the Belarusian language.

Although the focus has recently been on large language models, they still do not demonstrate high accuracy when working with languages whose training data is presented in a much smaller volume than for English.

However, such models can be used as an auxiliary tool, for example, to increase the size of lexical databases. Researchers also point to the effectiveness of the joint work of generative language models and BERT models. It can be considered as a possible direction for our further research.

6 Conclusion

In this paper, we conducted several experiments on training and fine-tuning language models with different architectures for the Belarusian language and evaluated them in the task of predicting paradigmatic and syntagmatic relations which are typical for text corpora.

POS-filters were added to differentiate between paradigmatic and syntagmatic relations. Using the Siamese neural network, a distinction was drawn between synonymous and antonymic relations in a vector space, which is reflected in the graphs of the cosine similarity distribution.

We fine-tuned the BERT model for the Belarusian language to determine the synonyms of words in the context. The final algorithm

Table 5. Model evaluation after fine-tuning

Sample	Number of people	Average score	Relative value	Consistency of experts' opinions
1	8	51,88	0,52	0,59
2	12	58,83	0,59	0,51
3	36	61,64	0,62	0,55
4	21	57,1	0,57	0,49
5	7	56,57	0,57	0,54

also includes accessing vocabulary data and comparing sentences in terms of their similarity using the Sentence Transformer. The expert evaluation showed that the accuracy of the model's predictions increased from an average of 24% to 75%. The Wilcoxon test determined that there is a statistically significant difference between the rates.

In addition, we analyzed the errors of generative language models when working with the Belarusian language and conducted an expert assessment of how suitable the synonymous substitutes proposed by the Gemini 2.5 Pro model are. The values of the average score for each of the five samples (from 50 to 60 with a maximum score of 100 points) indicate that LLMs currently do not reach a high level in tasks related to the field of lexical semantics of the Belarusian language.

At the same time, such models can offer relevant, but not listed in dictionaries, options, which is important for increasing the volume of lexical databases.

Since synonymy is key and one of the most common among lexical-semantic relations, the model training for the task of its prediction allows us to lay the foundation for the subsequent development of more complex linguistic systems for the Belarusian language, covering a variety of semantic relations.

The models obtained cannot only be integrated into systems of that kind, but also applied in such areas as information retrieval, machine translation, linguodidactics, etc. The models, program code, and datasets are placed in the GitHub repository[5].

[5]https://github.com/lizaveta190/models-datasets-Belarusian

References

1. **Ali, M. A., Sun, Y., Zhou, X., Wang, W., Zhao, X. (2019).** Antonym-synonym classification based on new sub-space embeddings.

2. **Antipenko, A., Mitrofanova, O. (2020).** Comparative study of word associations in social networks corpora by means of distributional semantics models for russian. International Journal of Open Information Technologies, Vol. 8, No. 1, pp. 27–33.

3. **Bonch-Osmolovskaya, A. (2024).** Distributivnye modeli v proektsii teoreticheskikh podkhodov leksicheskoj semantiki (po dannym NKRYA) [distributive models in the projection of theoretical approaches of lexical semantics (according to the RNC)]. Trudy IRYA RAN [Proceedings of the V. V. Vinogradov Russian Language Institute], , No. 4, pp. 200–242. DOI: https://doi.org/10.31912/pvrli-2024.4.11.

4. **Bukia, G., Protopopova, E., Mitrofanova, O. (2015).** Korpusnaya otsenka stepeni blizosti edinits v leksicheskikh konstruktsiyakh [corpus estimating of association strength between units in lexical constructions]. In **Gerd, A., Nikolaev, I.**, editors, Strukturnaya i prikladnaya lingvistika [Structural and applied linguistics], 11. St Petersburg University Publishing House, pp. 252–270.

5. **Clark, S., Weir, D. (2002).** Class-based probability estimation using a semantic hierarchy. Computational Linguistics, Vol. 28, No. 2, pp. 187–206. DOI: 10.1162/089120102760173643.

6. **Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2019).** BERT: Pre-training of deep bidirectional transformers for language understanding. **Burstein, J., Doran, C., Solorio, T.**, editors, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, pp. 4171–4186. DOI: 10.18653/v1/N19-1423.

7. **Dou, Z., Wei, W., Wan, X. (2018).** Improving Word Embeddings for Antonym Detection Using Thesauri and SentiWordNet: 7th CCF International Conference, NLPCC 2018, Hohhot, China, August 26–30, 2018, Proceedings, Part II. pp. 67–79. DOI: 10.1007/978-3-319-99501-4$_6$.

8. **Espinosa Anke, L., Codina-Filba, J., Wanner, L. (2021).** Evaluating language models for the retrieval and categorization of lexical collocations. **Merlo, P., Tiedemann, J., Tsarfaty, R.**, editors, Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Association for Computational Linguistics, Online, pp. 1406–1417. DOI: 10.18653/v1/2021.eacl-main.120.

9. **Etcheverry, M., Wonsever, D. (2019).** Unraveling antonym's word vectors through a Siamese-like network. **Korhonen, A., Traum, D., Màrquez, L.**, editors, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, pp. 3297–3307. DOI: 10.18653/v1/P19-1319.

10. **Glenski, M., Sealy, W. I., Miller, K., Arendt, D. (2021).** Improving synonym recommendation using sentence context. **Moosavi, N. S., Gurevych, I., Fan, A., Wolf, T., Hou, Y., Marasović, A., Ravi, S.**, editors, Proceedings of the Second Workshop on Simple and Efficient Natural Language Processing, Association for Computational Linguistics, Virtual, pp. 74–78. DOI: 10.18653/v1/2021.sustainlp-1.9.

11. **Khvalej, Y., Sharpla, U. (2004).** Slounik leksichnyh formau (sinonimy, amonimy, antonimy, paronimy, amografy, amafony) [Dictionary of lexical forms (synonyms, homonyms, antonyms, paronyms, homographs, homophones)]. Paradoks, Minsk.

12. **Klyshka, M. (1993).** Slounik sinonimau i blizkaznachnykh slou [Dictionary of synonyms]. Vyshejshaya shkola, Minsk, 2nd edition.

13. **Koshchanka, U. (2018).** Natsyyanal'ny korpus belaruskaj movy u kantekstse korpusnaj lingvistyki slavyanskikh krain [The National Corpus of the Belarusian Language in the context of Slavic countries corpus linguistics]. XVI International Congress of Slavists, Belgrade, pp. 60–79.

14. **Maxwell, J. (2019).** Siamese semantic vectors. https://github.com/maxwelljohn/word-vector-remapping. Accessed: 2025-11-14.

15. **Mazzitelli, L. F. (2021).** Belacorpus. https://github.com/Belarusian-Corpus. Accessed: 2025-02-15.

16. **McHugh, M. (2012).** Interrater reliability: The kappa statistic. Biochemia medica: časopis Hrvatskoga društva medicinskih biokemičara / HDMB, Vol. 22, pp. 276–282. DOI: 10.11613/BM.2012.031.

17. **Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013).** Efficient estimation of word representations in vector space.

18. **Moskvoretskii, V., Panchenko, A., Nikishina, I. (2024).** Are large language models good at lexical semantics? a case of taxonomy learning. **Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., Xue, N.**, editors, Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), ELRA and ICCL, Torino, Italia, pp. 1498–1510.

19. **Nguyen, K. A., Schulte im Walde, S., Vu, N. T. (2017).** Distinguishing antonyms and synonyms in a pattern-based neural network. **Lapata, M., Blunsom, P., Koller, A.**, editors, Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, Association for Computational Linguistics, Valencia, Spain, pp. 76–85.

20. **Pereira, F., Tishby, N., Lee, L. (1993).** Distributional clustering of English words. 31st Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Columbus, Ohio, USA, pp. 183–190. DOI: 10.3115/981574.981598.

21. **Petrushenko, L., Mitrofanova, O. (2023).** Predicting style-dependent collocations in russian text corpora. Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2023, pp. 79–89.

22. **Qin, L., Chen, Q., Zhou, Y., Chen, Z., Li, Y., Liao, L., Li, M., Che, W., Yu, P. S. (2024).** Multilingual large language model: A survey of resources, taxonomy and frontiers.

23. **Reimers, N., Gurevych, I. (2019).** Sentence-BERT: Sentence embeddings using Siamese BERT-networks. **Inui, K., Jiang, J., Ng, V., Wan, X.**, editors, Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, pp. 3982–3992. DOI: 10.18653/v1/D19-1410.

24. **Rogers, A., Kovaleva, O., Rumshisky, A. (2020).** A primer in BERTology: What we know about how BERT works. Transactions of the Association for Computational Linguistics, Vol. 8, pp. 842–866. DOI: 10.1162/tacl$_{a0}0349$.

25. **Roussinov, D., Sharoff, S., Puchnina, N. (2023).** Fine-tuning language models to recognize semantic relations. Language Resources and Evaluation, Vol. 57, pp. 1463–1486. DOI: 10.1162/tacl$_{a0}0349$.

26. **Samenko, I., Tikhonov, A., Yamshchikov, I. P. (2021).** Intuitive Contrasting Map for Antonym Embeddings. IOS Press. DOI: 10.3233/faia210282.

27. **Serina, L., Putelli, L., Gerevini, A. E., Serina, I. (2023).** Synonyms, antonyms and factual knowledge in BERT heads. Future Internet, Vol. 15, No. 7. DOI: 10.3390/fi15070230.

28. **Shishkina, Y., Lyashevskaya, O. (2021).** Sculpting enhanced dependencies for belarusian. Analysis of Images, Social Networks and Texts: 10th International Conference, AIST 2021, Tbilisi, Georgia, December 16–18, 2021, Revised Selected Papers, Springer-Verlag, Berlin, Heidelberg, pp. 137–147. DOI: 10.1007/978-3-031-16500-9$_1 2$.

29. **Tenney, I., Xia, P., Chen, B., Wang, A., Poliak, A., McCoy, R. T., Kim, N., Durme, B. V., Bowman, S. R., Das, D., Pavlick, E. (2019).** What do you learn from context? Probing for sentence structure in contextualized word representations.

30. **Timofeeva, M. (2018).** Typology of semantic relations extracted by the instrument rusvectōrēs. Nauchnyi dialog, Vol. 8, pp. 74–87. DOI: 10.24224/2227-1295-2018-8-74-87.

31. **Volchek, O., Poritski, V. (2013).** An experimental corpus of belarusian: its present and future. Trudy Mezhdunarodnoy konferentsii ”Korpusnaya lingvistika -2013” [Proceedings of the international conference ”Corpus linguistics-2013”], St. Petersburg, Russia, pp. 231–239.

32. **Vulić, I., Ponti, E. M., Litschko, R., Glavaš, G., Korhonen, A. (2020).** Probing pretrained language models for lexical semantics. **Webber, B., Cohn, T., He, Y., Liu, Y.**, editors, Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, pp. 7222–7240. DOI: 10.18653/v1/2020.emnlp-main.586.

33. **Wang, R., Choi, M. (2023).** Large language models on lexical semantic change detection: An evaluation.

34. **Zhou, W., Ge, T., Xu, K., Wei, F., Zhou, M. (2019).** BERT-based lexical substitution. **Korhonen, A., Traum, D., Màrquez, L.**, editors, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Florence, Italy, pp. 3368–3373. DOI: 10.18653/v1/P19-1328.

Article received on 15/11/2025; accepted on 26/12/2025.
**Corresponding author is Olga Mitrofanova.*