

# An Interpretable Approach to Lexical Semantic Change Detection with Lexical Substitution

Arefyev N. V.

Lomonosov Moscow State University  
Samsung R&D Institute Russia  
HSE University  
Moscow, Russia  
nick.arefyev@gmail.com

Bykov D. A.

Lomonosov Moscow State University  
Moscow, Russia  
dima13051998@gmail.com

## Abstract

In this paper we propose a new Word Sense Induction (WSI) method and apply it to construct a solution for the RuShiftEval shared task on Lexical Semantic Change Detection (LSCD) for the Russian language. Our WSI algorithm based on lexical substitution achieves state-of-the-art performance for the Russian language on the RUSSE-2018 dataset. However, our LSCD system based on it has shown poor performance in the shared task. We have studied mathematical properties of the COMPARE score employed in the task for measuring the degree of semantic change, as well as the discrepancies between this score and our WSI predictions. We have found that our method can detect those aspects of semantic change, which the COMPARE metric is not sensitive to, such as appearance or disappearance of a rare word sense. An important property of our method is its interpretability, which we exploit to perform the detailed error analysis.

**Keywords:** Lexical Substitution, Lexical Semantic Change Detection, Word Sense Induction

**DOI:** 10.28995/2075-7182-2021-20-31-46

## Интерпретируемый подход к обнаружению семантических сдвигов с помощью лексических подстановок

Арефьев Н. В. ◊ △ ▽

Москва, Россия  
nick.arefyev@gmail.com

Быков Д. А. ◊

Москва, Россия  
dima13051998@gmail.com

Московский государственный университет имени М. В. Ломоносова ◊  
Московский Исследовательский Центр Самсунг △

Национальный исследовательский университет «Высшая школа экономики» ▽

## Аннотация

В данной статье мы предлагаем новый метод решения задачи выделения значений слов (WSI) и строим на его основе решение задачи обнаружения семантических сдвигов (LSCD). Наш алгоритм выделения значений слов превосходит по качеству предыдущие методы на датасете RUSSE-2018 для русского языка. Однако наша система решения задачи LSCD показывает низкие результаты на RuShiftEval. Мы изучили математические свойства метрики COMPARE, используемой в данной задаче для оценки степени семантического сдвига, а также расхождения между данной метрикой и предсказаниями нашего метода. Мы обнаружили, что наш метод может определять такие семантические сдвиги, к которым данная метрика не чувствительна, например, появление или исчезновение редкого значения. Другим отличительным свойством нашего метода является интерпретируемость, которую мы использовали при анализе ошибок.

**Ключевые слова:** Задача обнаружения семантических сдвигов, лексические подстановки, выделение значений слов.

## 1 Introduction

Lexical semantic change detection (LSCD) is a problem of detecting changes in word meaning over time. Semantic change is a complex phenomenon which is hard to define or formalize. This results in different data annotation schemes, a bunch of complementary metrics, and also different LSCD systems sensitive

to different aspects of the whole phenomenon. The idea behind our approach is first to discover word senses using a Word Sense Induction (WSI) method based on clustering of lexical substitutes. Then we can study the discovered senses and find out if any new sense appeared or an old one disappeared between two time periods. Unlike popular approaches to LSCD that build a single vector for each word in each time period [11], our approach is much more interpretable because it can display examples of each discovered sense, or label corresponding clusters with the words related to those senses. Following this approach, we developed a system for the RuShiftEval [7] competition on LSCD for the Russian language.

Our WSI model is based on the method proposed in [2] for English, which generates lexical substitutes for all occurrences of a particular target word with BERT [6], and then clusters bag-of-word representations of those substitutes. Lexical substitutes are usually different for target word occurrences with different senses and similar for those with the same sense. For instance, the word *mouse* will receive substitutes like *keyboard*, *monitor* in the phrase *connect the mouse to your PC*, while in the phrase *laboratory mouse* substitutes like *frog*, *rabbit*, *monkey* will be generated. We make the base method multilingual by replacing English BERT [6] model with XLM-R [5] model trained on texts in 100 different languages. However, simply replacing BERT with XLM-R to generate substitutes does not work well for the Russian language because the base method generates only single subword substitutes, which are either functional words or word pieces. This results in poor inventory of meaning-bearing substitutes and very poor WSI performance. Thus, we propose a technique to generate multi-subword substitutes, which boosts WSI performance for the Russian language (and likely other languages too, though we did not test it yet). Additionally, we propose and experiment with different Hearst-like patterns specific for the Russian language, and also their combinations. Generating multi-subword substitutes and selecting the best combination of Hearst-like patterns results in new SOTA on the *bts-rnc* dataset from RUSSE-2018 WSI competition for the Russian language.

During the RuShiftEval competition period we did not manage to achieve good values of the COMPARE metric used in the competition, and also could not outperform our strong but less interpretable baseline based on the orthogonal Procrustes alignment of SGNS vectors, which was among three best performing methods in the SemEval-2020 Task 1 LSCD competition [11], though that competition employed different annotation scheme and metrics. In the post-competition period we performed a mathematical analysis of the COMPARE metric and discovered that unlike our WSI-based method, it is not sensitive to the appearance or disappearance of rare senses, but rather estimate the change in the relative frequency of the most frequent sense. We show that for those words that have gained or lost one of their senses the COMPARE metric can be either higher or lower than for those words that did not undergo any change, which raises a question about the applicability of this metric for LSCD.

Based on our analysis of the COMPARE metric, after the competition we decided to try the same substitution-based distance metric employed for clustering in our WSI method, but skipping the clustering itself to adapt to the properties of the COMPARE metric. This adaptation gave large improvement of the competition metric and outperformed the strong baseline. Finally, we performed error analysis of our WSI-based method. It has revealed that there are actually cases when our method discovers the appearance of a new rare sense, which the COMPARE metric is not sensitive to.

## 2 Related work

RuShiftEval [7] is the first LSCD shared task for the Russian language. There are three time periods in the RuShiftEval competition (Soviet, pre-Soviet, and post-Soviet). For the evaluation, they were grouped into three pairs of periods. The participants were asked to estimate the semantic shift of each test word between all pairs of periods by providing 3 scores for each word. The data annotation scheme and the evaluation metric follow those proposed as a part of the Diachronic Usage Relatedness (DURel) dataset for German [12]. The performance of the participating systems was evaluated by calculating the Spearman correlation between the predicted word scores and the values of the COMPARE [12] metric. To calculate the COMPARE metric during the dataset annotation, for each word and each pair of time periods several dozens sentence pairs were randomly sampled. In these pairs the first sentence was sampled from the first period and the second sentence from the second period. Each sentence pair was annotated

by humans using 1-4 scale where larger values corresponded to more similar senses of the target word. Finally, annotations of sentence pairs were averaged to produce word scores. The COMPARE metric is known to confuse polysemy and meaning change [12], in this paper we describe other interesting properties of this metric.

An alternative approach to annotation and evaluation was proposed in SemEval-2020 Task 1 [11], where the authors tried to account for the appearance or disappearance of relatively rare word senses, which the COMPARE metric is not sensitive to. They basically clustered word occurrences based on human judgements about similarity of these occurrences by meaning. The transitivity of the same-sense relation is exploited to improve annotation efficiency. Based on the obtained gold standard clustering corresponding to word senses, two subtasks were proposed. The first subtask required binary classification determining if the set of senses of a particular word has changed. The second subtask required ranking words according to the change in relative frequencies of their senses.

Our solution is based on a substitution-based approach to WSI [1, 2, 4] which exploits lexical substitutes to distinguish word senses. Following [2] we used dynamic patterns and develop specific patterns for the Russian language. We developed a lexical substitution model based on the XLM-R multilingual masked language model [5], which is trained on texts in 100 languages. As a baseline we used the LSCD method [10] based on orthogonal Procrustes alignment of SGNS word embeddings [8]. Similar methods based on SGNS embeddings and different alignment techniques were the best performing methods in SemEval-2020 Task 1 [11].

### 3 Methods of WSI and LSCD

Our LSCD system retrieves examples for the target word from each time period and performs WSI based on lexical substitution for all those examples. The result of WSI is a clustering of word occurrences. Our solution submitted to the competition analyzed the number of examples from each time period in each cluster and made the predictions based on those numbers. The clusters can be labeled with distinctive substitutes. Both these labels and the examples in each cluster can be examined, which makes the proposed method interpretable.

After the competition, we developed a simpler LSCD method inspired by the revealed properties of the evaluation metric. This method employs the same measure of similarity between lexical substitutes that was used for clustering in the WSI algorithm, but skips clustering itself. This significantly improves the evaluation metric, but makes the predictions less intuitive and interpretable.

#### 3.1 Lexical Substitution and Word Sense Induction

Our WSI method employs lexical substitutes generated using the XLM-R masked language model (MLM) [5] to represent word meaning manifested in a particular context. We tried several approaches to generate substitutes. One of them is replacing the target word in the given text fragment with a special token `<mask>` and asking the MLM to recover it. This results in contextually plausible words, but often they are not related to the target word and do not describe its sense in any way. To solve this problem we used dynamic patterns proposed in [1], which are Hearts-like patterns applied to the inputs of the MLM. For instance, we can replace the target word `T` with "`T and <mask>`" or "`<mask> and T`", resulting in words that can stand in a co-ordinated row with the target, mostly co-hyponyms of the target word. We compared a number of such patterns and discovered that the best performance is achieved by two symmetric patterns "`T (а также <mask>)`" and "`<mask> (а также T)`" (literally translated as "`T (and also <mask>)`" and "`<mask> (and also T)`"), which were both selected for our final solution.

However, the XLM-R vocabulary contains only functional and very frequent Russian words, most content words are split into several subwords. We found that single-subword substitutes alone result in poor performance of WSI for the Russian language and extended the base approach by generating multi-subword substitutes. To achieve this, we insert several `<mask>` tokens in the pattern ("`T and <mask><mask><mask>`", for instance). Then we generate substitutes from left to right using beam search, i.e. we take  $K$  most probable subwords predicted by XLM-R for the first `<mask>` token, for each

of them predict  $K$  most probable continuations resulting in  $K^2$  sequences of two subwords filling two `<mask>` tokens, and from them leave only  $K$  most probable sequences, etc. Unlike traditional language models, a masked language model is not an autoregressive model. Thus, the described procedure is not a mathematically sound way to deal with the distribution learnt by the model, though it gives empirically good results. Some examples are presented in table 1. For our final solution we generated substitutes consisting of 1 to 3 subwords.

Next, in order to represent the lexical meaning of the target word and not its grammatical form, the generated substitutes are lemmatized. For each lemma the sum of the predicted probabilities of all of its forms and across two symmetric patterns with different number of masks is calculated. Finally,  $K$  substitutes with the highest sum of probabilities are taken for each example. Following [2], we build TF-IDF bag-of-words vectors of the generated substitutes and clustered these vectors using agglomerative clustering with cosine distance and complete linkage. For WSI experiments in section 4.1, the number of clusters is either selected for each word individually using the silhouette score similarly to [4], or is selected the same for all words based on training set metrics. For LSCD we iterated over several values to obtain different clusterings (see section 3.2). Finally, for the qualitative analysis in section 4.4 only the silhouette score was used.

cluster ID	sentence	top 10 substitutes
0	дополнительных вертикально расположенных винта (а также <code>&lt;mask&gt;&lt;mask&gt;</code> ). С целью обеспечения устойчивости вертолета плоскости вращения несущих винтов были немного наклонены внутрь, этим достигался эффект, аналогичный эффекту поперечного V крыла	<b>крыла:</b> 2e-2 <b>колеса:</b> 9e-3 <b>двигателя:</b> 9e-3 <b>опоры:</b> 8e-3 <b>хвоста:</b> 7e-3 <b>пружины:</b> 6e-3 <b>других элементов:</b> 6e-3 <b>привода:</b> 5e-3 <b>диска:</b> 5e-3 <b>двух дополнительных:</b> 5e-3
0	настоящее время разворачивается его серийное производство. Основными особенностями машины является отсутствие хвостового винта (а также <code>&lt;mask&gt;&lt;mask&gt;</code> ) и модульная компоновка. Преимущества соосной схемы обеспечили вертолету следующие потребительские свойства: – высокую	<b>двигателя:</b> 5e-2 <b>крыла:</b> 1e-2 <b>привода:</b> 1e-2 <b>колеса:</b> 1e-2 <b>амортизатора:</b> 1e-2 <b>хвоста:</b> 8e-3 <b>кузова:</b> 8e-3 <b>его отсутствие:</b> 7e-3 <b>тормозов:</b> 6e-3 <b>других элементов:</b> 6e-3
1	после десяти часов и садились закусывать. Одни ужинали, другие играли в скромные винт (а также <code>&lt;mask&gt;&lt;mask&gt;</code> ) и преферанс, третьи проигрывались в «железку» и штрафами покрывали огромные расходы Кружка.	<b>покер:</b> 1e-2 <b>рулетку:</b> 9e-3 <b>в карты:</b> 9e-3 <b>карточные:</b> 7e-3 <b>лото:</b> 5e-3 <b>тайм:</b> 5e-3 <b>спиннинг:</b> 4e-3 <b>теннис:</b> 4e-3 <b>экспресс:</b> 3e-3 <b>шашки:</b> 3e-3
1	к своему сослуживцу Шешковскому, у которого каждый день собирались чиновники играть в винт (а также <code>&lt;mask&gt;&lt;mask&gt;</code> ) и пить холодное пиво. «Своею нерешительностью я напоминаю Гамлета. – думал Лаевский	<b>в карты:</b> 8e-2 <b>шахматы:</b> 8e-2 <b>мяч:</b> 2e-2 <b>рулетку:</b> 2e-2 <b>курить:</b> 1e-2 <b>стрелять:</b> 9e-3 <b>теннис:</b> 8e-3 <b>хоккей:</b> 8e-3 <b>шутить:</b> 7e-3 <b>другие игры:</b> 6e-3

Table 1: An example of inputs to the XLM-R model, generated substitutes and clusters assigned.

### 3.2 Semantic Change Detection

Our WSI method can be directly used to decide that a new sense appeared or an old disappeared between two time periods if there are clusters containing only examples from one of the periods. However, for the RuShiftEval competition the words had to be ranked according to some scores reflecting the strength of semantic change. To calculate the score of semantic change for a particular word and two time periods, we sampled  $A$  sentences containing this word from each time period and clustered all  $2A$  sentences several times using our WSI method with different number of substitutes and different number of clusters<sup>1</sup>. Then from the clustered sentences we randomly generated  $B$  pairs of sentences with the first sentence from an old period, and the second from a new one. Each sentence was sampled from the corresponding  $A$  sentences with replacement. Then the final score was calculated as the average across all clusterings of the indicators that both sentences fell into the same cluster. The intuition behind this method is the following.

<sup>1</sup>Our best submitted solution iterated over [50,60,70,80,90,100,200,300] substitutes and [2,3] clusters.

If two word occurrences are identical in meaning, they are likely to be grouped together in many different clusterings. However, if the meanings are different, they can be grouped together by chance or mistake, but hopefully not too often. For the best submitted solution we set  $A = 60$  and  $B = 120$ .

### 3.3 Cluster labeling

During WSI the clustering algorithm groups together those target word occurrences that have similar substitutes. Thus, we can find those substitutes that are specific to each cluster. Following [2], we calculate the Pointwise Mutual Information (PMI) for a given substitute and a given cluster. To label each cluster, we select substitutes having the maximum PMI value with that cluster. Despite many noisy substitutes appearing in such labels, they are still useful as compact representations of clusters. Thus, we employ these labels for the qualitative analysis.

### 3.4 Cosine Similarity Averaging

After the competition we developed a new algorithm using the same distances between substitute vectors, but skipping the clustering procedure. For each word we sampled  $A = 80$  examples from each period. For each pair of periods and each word we generated  $B = 120$  pairs of sentences in which the first sentence was from an earlier period and the second sentence was from a later period. Then, for each pair of sentences we calculated the cosine distance between substitute vectors. Finally, for each target word the average distance was calculated, producing the predicted word scores.

### 3.5 SGNS+OP+CD

As a baseline in our study we used SGNS vectors with orthogonal Procrustes alignment and cosine distance proposed in [10]. We got the best results with the following hyperparameters: windowsize=5, k=5, t=0.001, dimensions=300, minCount=0, iters=5.

## 4 Results and Discussion

### 4.1 Experiments with Word Sense Induction

Table 2 compares our WSI method to the previous best results during and after RUSSE-2018 competition on bts-rnc dataset [9] using the official metric of that competition, which is the Adjusted Rand Index (ARI), and official evaluation scripts. All experiments and hyperparameter selection were performed on the training set, while the final results in Table 2 are reported on the public and the private parts of the test set.

model	public test ARI	private test ARI
<b>RUSSE-2018 results</b>		
AdaGram baseline [9]	0.262	0.213
RUSSE-2018 best result [9]	0.351	0.338
RUSSE-2018 2nd best result [3]	0.281	0.281
<b>other published results</b>		
RuBERT semantic fingerprint [13]	0.21 <sup>△</sup>	
bayes-comb-silnc [4]	0.502	0.451
bayes-comb-fixnc [4]	0.464	0.438
<b>our results</b>		
andalso-2subwords-silnc	0.525	0.507
andalso-2subwords-fixnc	<b>0.564</b>	<b>0.573</b>

Table 2: Comparison with the published results on the RUSSE-2018 bts-rnc dataset. Selecting the number of clusters for each word using silhouette score (silnc) or as a hyperparameter on train (fixnc). <sup>△</sup>The subset of bts-rnc used to evaluate this method is not specified in [13].

To find optimal hyperparameters of the lexical substitution model, we employed maxARI metric introduced in [4], which is the maximum possible ARI achieved by agglomerative clustering with different number of clusters. First, we compare Hearst-like patterns that potentially can give good substitutes for the Russian language. In Figure 1, we consider patterns with only one <mask> token, which is to the left from the target word. We selected the best performing patterns "<mask> и Т" ("<mask>

and T"), "<mask> или T" ("<mask> or T"), "<mask> (а также T)" ("<mask> (and also T)"), "<mask> (в том числе T)" ("<mask> (including T)").

For the selected patterns we considered two possible orders of the target word and the <mask> token. After the competition we also experimented with symmetric combinations obtained by multiplying probability distributions over possible substitutes. This is equivalent to leaving only substitutes that are probable in both patterns. Figure 1 shows that this symmetric combinations consistently improve the results for all patterns, especially when multi-subword substitutes are generated. Interestingly, putting <mask> to the right of the target, like in "T and <mask>", works bad when generating multi-subword substitutes because multi-word expressions such as *other varieties*, *in other places* are often generated, which are not very related to any particular word senses. Also we notice that generating two-subword substitutes significantly improve the performance for symmetric combinations, while generating three-subword substitutes seems to be an overkill.

Finally, Table 2 compares our WSI method using two-subword substitutes and symmetric combination of "X (а также X)" ("X (and also X)") to the best previous results on the *bts-rnc* dataset, where it achieves the new state-of-the-art performance. The application of dynamic patterns sometimes produces ungrammatical sentences. In Appendix D we provide some preliminary analysis of this problem.

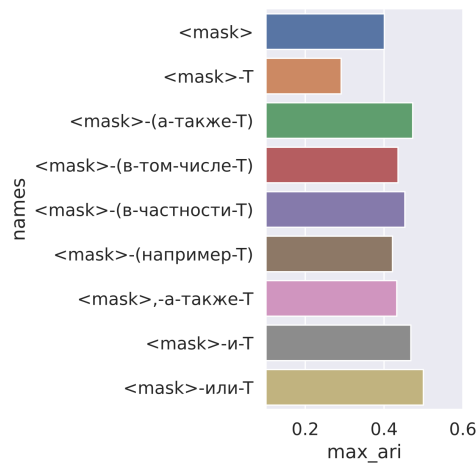


Figure 1: Comparison of different patterns with a single <mask> token.

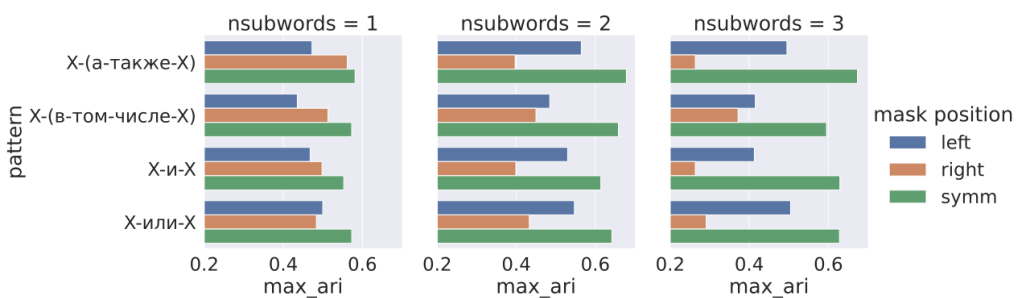


Figure 2: Comparison of patterns with different number and positions of <mask> tokens.

## 4.2 LSCD Results for RuShiftEval

In the RuShiftEval the systems were ranked by the Spearman correlation between their scores and the values of the COMPARE metric for each word. This metric is the average of annotator predictions for a set of sentence pairs, where the first sentence refers to an earlier time period, and the second to a later one. Unfortunately, during the evaluation period we did not have the cosine averaging approach yet. The results in Table 3 show that the cosine averaging method gives word scores that are much better correlated

with the COMPARE metric compared to the submitted WSI-based method and SGNS+OP+CD baseline. If submitted, the cosine averaging method could give us 4th place among all teams.

methods	pre-Soviet:Soviet	Soviet:post-Soviet	pre-Soviet:post-Soviet	average
<b>best competition results</b>				
GlossReader	0.781	0.803	0.822	0.802
DeepMistake	0.798	0.773	0.803	0.791
vanyatko	0.678	0.746	0.737	0.720
<b>our submissions</b>				
WSI based approach	0.274	0.202	0.307	0.261
SGNS + OP + CD	0.2690	0.2240	0.4230	0.3050
<b>our post-evaluation results</b>				
Cosine Similarity Averaging	0.671	0.677	0.658	0.66

Table 3: LSCD results on RuShiftEval, Spearman’s rank correlation.

### 4.3 The COMPARE metric analysis

In this section we describe the properties of the COMPARE metric explaining what kind of semantic changes does it reflect and what kind of changes it is not sensitive to. To simplify our analysis, we will discuss only cases when a word has unrelated senses and each pair of its occurrences receive annotations of 1 (entirely different senses) or 4 (same sense).

We start with a simple example. Suppose there is a word (a homonym) having two unrelated senses in the old corpus with probabilities of 0.9 for sense1 and 0.1 for sense2. First, suppose the least frequent sense disappeared in the new corpus and only the most frequent is left. Then the second sentence sampled from the new corpus always has the first sense of the target word. Thus, we sample a pair of sentences with the same word meaning and receive the score 4 from annotators with probability of 0.9, while sampling unrelated word meanings and obtaining the score 1 has the probability of 0.1 (Table 4 shows all possible outcomes and their probabilities for this and the following cases). Thus, the expected value of the COMPARE metric is  $4 * 0.9 + 1 * 0.1 = 3.7$ . Now suppose that the most frequent sense disappeared. The same calculations result in the COMPARE metric of  $4 * 0.1 + 1 * 0.9 = 1.3$ . Finally, suppose nothing changed, then the COMPARE metric will be  $4 * (0.81 + 0.01) + 1 * (0.09 + 0.09) = 3.46$ . Thus, if a sense has disappeared (or appeared, as the calculations are symmetric), the COMPARE metric can be both near its maximum of 4 or its minimum of 1 depending on the sense frequencies. And when nothing changed, the result is in between. Thus, the metric does not reflect appearance or disappearance of word senses, but rather frequency distribution of those senses.

outcome			outcome probability		
sentence1	sentence2	annotation	only sense1 left	only sense2 left	nothing changed
sense1	sense1	4	0.9	0	0.81
sense2	sense2	4	0	0.1	0.01
sense1	sense2	1	0	0.9	0.09
sense2	sense1	1	0.1	0	0.09
			<b>COMPARE=3.7</b>	<b>COMPARE=1.3</b>	<b>COMPARE=3.46</b>

Table 4: Possible outcomes, their probabilities and corresponding annotations for our two sense example.

Now let us study the metric in a more general case. Suppose there are several unrelated senses of a word with probabilities  $p = (p_1, \dots, p_n)$  in the old corpus and  $q = (q_1, \dots, q_n)$  in the new one. For convenience, let us order senses such that  $p_1 > p_2 > \dots > p_n$ . The COMPARE metric is the average over annotations for sampled sentences. Thus, it is an estimate of the expected value of annotation for a pair of sentences sampled from uniform distributions over the old and the new corpus, which is the sum of probabilities of sampling the same sense<sup>2</sup> or different senses<sup>3</sup> of the target word from two corpora multiplied by the corresponding annotations (4 or 1):

<sup>2</sup>which is the probabilities of sampling i-th sense from both corpora summed over all senses

<sup>3</sup>it can be simply calculated as the probability of NOT sampling the same sense

$$E_{s_1 \sim p, s_2 \sim q} \text{annot}(s_1, s_2) = 4 * \sum_i p_i q_i + 1 * (1 - \sum_i p_i q_i) = 3 * \sum_i p_i q_i + 1$$

Since  $q_i$  are probabilities, this sum can be treated as a weighted average of  $p_i$  with weights  $q_i$ . This weighted average achieves maximum when  $q_1 = 1$  (only the most frequent sense left) and minimum when  $q_n = 1$  (only the least frequent sense survived). When nothing has changed, it is somewhere in between.

This analysis suggests that a method that predicts high semantic change scores for those words which obtained or lost a rare sense will have low Spearman correlation with the COMPARE metric simply due to the shown properties. At the same time, a method that is sensitive only to the change in the frequency of the most frequent sense may be considered as a good LSCD method by this metric.

#### 4.4 Qualitative and error analysis

In this section we study the discrepancies between the results of our WSI algorithm and the COMPARE metric values. For each word we sampled 80 examples from the pre-Soviet period and 80 from the post-Soviet period from the Russian National Corpus. From the test set we selected two groups of words, including 50 words with COMPARE > 3 into the first group and 12 words with COMPARE < 2 into the second group. According to the COMPARE metric, the second group shall contain words with strong semantic shift while the first group shall not.

We applied our WSI algorithm to all of these words. We predicted that a word has acquired a new sense according to our algorithm if there was a cluster containing at most two examples from the pre-Soviet period and at least 4 examples from the post-Soviet period (such bounds compensate for mistakes of the clustering algorithm). From the first group (high COMPARE) we took 10 words that acquired a new sense according to our predictions, while from the second group (low COMPARE) we took 10 words which presumably did not acquire or lose any senses<sup>4</sup> (Table 5). Those words have the largest discrepancy between our predictions and the values of the COMPARE metric in the gold standard.

first group			second group		
word	COMPARE	discrepancy type	word	COMPARE	discrepancy type
бригада	3.08	correct new sense (OK)	дух	1.88	a lot of senses (OK)
жесть	3.41	incorrect new sense	наложение	1.78	a lot of senses (OK)
обоснование	3.58	incorrect new sense	полоса	1.41	all senses in one cluster
ранец	3.38	incorrect new sense	роспись	1.57	changed MFS (OK)
сверстник	3.82	incorrect new sense	ссылка	1.93	changed MFS (OK)
список	3.05	incorrect new sense	тачка	1.89	all senses in one cluster
стол	3.25	incorrect new sense	хрен	1.6	changed MFS (OK)
тупик	3.14	incorrect new sense	центр	1.87	a lot of senses (OK)
увольнение	3.32	senses are combined	ядро	1.47	changed MFS (OK)
углеводород	3.2	correct new sense (OK)	ясли	1.9	a lot of senses (OK)

Table 5: Discrepancies between predictions of our method and the COMPARE metric. Discrepancies that are not errors of our method are marked with (OK).

Appendix A shows clusters for the words from the first group, which obtained a new sense according to our predictions. The first type of discrepancies (**correct new sense**), are those words, that really obtained a new sense at least in our sampled subset of RNC, but were not detected by the COMPARE metric. There are two words, *бригада* (*military brigade / team of workers*) and *углеводород* (*hydrocarbon as a class of organic chemical compounds / oil and natural gas as economic resources*), for which our predictions seem to be correct. In Appendix C random examples containing these words from two time periods and the assigned clusters are shown. For the first word *бригада* WSI found the most frequent sense associated with the army, which was overwhelming in the pre-Soviet corpus, and the second sense

<sup>4</sup>Sense frequencies still might have changed, it can be easily detected from cluster sizes if required. However, it is an open question whether LSCD methods shall detect only words that acquired or lost some sense, or words with relative frequencies of senses changed also. To make our qualitative analysis more simple and clear, we selected the first option to retrieve words for analysis.



(a team of workers), which was found in the post-Soviet corpus only and is still relatively rare compared to the first one. For the second word, WSI found the correct new sense associated with oil and natural gas as economic resources, and an existing one associated with chemical compounds. Almost all examples for the "oil and natural gas" sense are from the post-Soviet corpus, the value of COMPARE metric are relatively high probably due to the existing sense still dominating. The second type of discrepancies (**incorrect new sense**) are the words *жесть* (*tinplate*), *ранец* (*satchel*), *стол* (*table*), *обоснование* (*justification*), *тупик* (*deadlock*), *список* (*list*), *сверстник* (*coeval*). Our method divided some senses into more than one cluster. Thus, the prediction is wrong. The third type (**senses are combined**) consists of one word *увольнение*, the senses of "vacation" and "dismissal" were merged into a single cluster, but a few examples of the sense "dismissal" were put into a separate cluster.

In the second group (Appendix B), there are words with low COMPARE metric for which our method has not found any new or lost senses. We found three types of such words. The first type (**all senses in one cluster**) consists of the words *полоса* (*line / time interval / forest belt*), *мачка* (*car / cart*), for which our method mistakenly found only one frequent sense not counting outliers. For the rest of the words, the clustering results seem to be correct. Moreover, despite no senses were obtained or lost, the observed distribution of sense frequencies explains low values of COMPARE metric. For the words *ядро* (*nucleus / cannonball*), *ссылка* (*deportation / reference*), *хрен* (*horseradish / man*), *ропись* (*list / painting / signature*), the most frequent sense changed (**changed MFS**). The words *наложение* (*imposition / overlay*), *дых* (*spirit / ambience / ghost*), *центр* (*downtown / middle*), *ясли* (*manger / nursery / crib*) are highly polysemous, thus, receive low COMPARE values, which is a known problem for COMPARE metric [12].

The results of our error analysis are shown in Table 5. We have found 10 words out of 20, for which our method correctly predicted that the set of word senses has changed or has not changed. The discrepancy with the COMPARE metric for these words can be explained by the fact that this metric reflects the change in sense frequencies rather than the appearance or disappearance of senses, and also gives low scores to highly polysemous words. Indeed, according to the analysis of the clusters built by our WSI algorithm there was a significant change in relative frequency of the most frequent senses of several words that have not obtained or lost senses. Also we have found highly polysemous words with low values of the COMPARE metric. For two words with relatively high COMPARE values the appearance of a new sense was correctly predicted, but this sense has not become the most frequent sense, which explains the discrepancy. The most frequent error of our method is splitting a single sense into multiple clusters, which we will try to overcome in the future work.

## 5 Conclusion

We proposed a WSI method achieving new state-of-the-art for the Russian language on the *bts-rnc* dataset. Also we proposed two approaches for solving LSCD task. The first approach is based on our WSI method. This approach has an important advantage of being interpretable. One can look at the clusters that contain sentences from only one time period and understand which sense appeared or disappeared, or if this is just a mistake of the algorithm. Our second method is less interpretable, but it achieves higher results according to the competition metric.

## References

- [1] Asaf Amrami and Yoav Goldberg. Word Sense Induction with Neural biLM and Symmetric Patterns. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4860–4867, 2018.
- [2] Asaf Amrami and Yoav Goldberg. Towards better substitution-based word sense induction. *arXiv e-prints*, pages arXiv–1905, 2019.
- [3] Nikolay Arefyev, Pavel Ermolaev, and Alexander Panchenko. How much does a word weigh? weighting word embeddings for word sense induction. In *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue”*, pages 68–84, Moscow, Russia, 2018. RSUH.
- [4] Nikolay Arefyev, Boris Sheludko, and Tatiana Aleksashina. Combining Neural Language Models for Word Sense Induction. In *Analysis of Images, Social Networks and Texts*, page 105–121. Springer International Publishing, 2019.
- [5] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Édouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, 2020.
- [6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *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)*, pages 4171–4186, 2019.
- [7] Andrey Kutuzov and Lidia Pivovarova. RuShiftEval: a shared task on semantic shift detection for Russian. *Komp’yuternaya Lingvistika i Intellektual’nye Tekhnologii: Dialog conference*, 2021.
- [8] Tomas Mikolov, Kai Chen, G.s Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. *Proceedings of Workshop at ICLR*, 2013, 01 2013.
- [9] Alexander Panchenko, Anastasia Lopukhina, Dmitry Ustalov, Konstantin Lopukhin, Nikolay Arefyev, Alexey Leontyev, and Natalia Loukachevitch. RUSSE’2018: A Shared Task on Word Sense Induction for the Russian Language. In *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue”*, pages 547–564, Moscow, Russia, 2018. RSUH.
- [10] Dominik Schlechtweg, Anna Häty, Marco Del Tredici, and Sabine Schulte im Walde. A Wind of Change: Detecting and Evaluating Lexical Semantic Change across Times and Domains. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 732–746, 2019.
- [11] Dominik Schlechtweg, Barbara McGillivray, Simon Hengchen, Haim Dubossarsky, and Nina Tahmasebi. SemEval-2020 task 1: Unsupervised lexical semantic change detection. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1–23, Barcelona (online), December 2020. International Committee for Computational Linguistics.
- [12] Dominik Schlechtweg, Sabine Schulte im Walde, and Stefanie Eckmann. Diachronic usage relatedness (DUREl): A framework for the annotation of lexical semantic change. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 169–174, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.
- [13] Aleksandr Slapoguzov, Konstantin Malyuga, and Evgenij Tsopa. Word Sense Induction for Russian Texts Using BERT. In *Conference of Open Innovations Association, FRUCT*, number 28, pages 621–627. FRUCT Oy, 2021.

## Appendix A Words with high COMPARE and a new sense predicted

In Figure 3 we show 10 words from RuShiftEval dataset with COMPARE > 3, for which our method predicted that a new sense appeared in the post-Soviet time period, i.e. returned a cluster containing at least 4 examples from the post-Soviet period and at most 2 examples from the pre-Soviet period.

Words	Cluster names	OLD COUNT	NEW COUNT
<b>бригада</b>	0 2, армия, батальон, батарея, войско, два, дивизия, дивизя	0 79	0 61
	1 служба, машина, они, человек, полиция, организация, я, весь	1 1	1 19
<b>жесьть</b>	0 страх, я, лицо, глаз, форма, силой, красота, кровь	0 13	0 22
	1 бетон, бронза, глина, железо, из, картон, лист, масло	1 65	1 53
	2 любовь, красота, сила	2 2	2 5
<b>обоснование</b>	0 построение, изложение, практика, толкование, понимание, выражение	0 30	0 21
	1 создание, разработка, осуществление, проектирование, использование, реализация	1 11	1 18
	2 аргумент, факт, в, другой, причина, решение, аргументация, доказательство	2 6	2 17
	3 проектирование, расчёт, разработка, другой, содержание, решение	3 0	3 9
	4 признание, основа, оправдание, подтверждение, смысл, установление	4 29	4 12
	5 результат, другой, основа, выражение, содержание, основание	5 4	5 3
<b>ранец</b>	0 вода, в, друг, другой, я, дом, она, предмет	0 2	0 7
	1 автомат, без, велосипед, весь, вещей, глаз, голова, деньги	1 66	1 72
<b>сверстник</b>	0 враг, коллега, мой, не, отец, соратник, ученик, родные	0 78	0 56
	1 самый, взрослый, старший, с, школьник, весь, семья, ребёнок	1 2	1 24
<b>список</b>	0 письмо, документ, текст, дать, предложение, книга, лист, фамилия	0 41	0 17
	1 это, память, остальной, в, лист, журнал, рейтинг, резюме	1 13	1 22
	2 остальной, рейтинг, состав, они, по, программа, этот, другой	2 1	2 4
	3 описание, по, номер, название, перечень, некоторый, портрет, фото	3 24	3 35
	4 программа, он, свой, реестр, число, текст, в, отчёт	4 1	4 2
<b>стол</b>	0 блюдо, стакан, к, другой, ребёнок, в, он, стол	0 2	0 8
	1 внутри, голова, дверь, книга, комната, кровать, кухня, мебель	1 77	1 68
	2 другой, на, и, к, в, перед, зал, кресло	2 1	2 4
<b>тупик</b>	0 дом, угол, дорога, конец, ряд, один, в, обратно	0 14	0 27
	1 паника, затруднение, сомнение, ужас, шок, точка, хаос, наоборот	1 61	1 22
	2 -, ,, 3, </s>, автомат, велосипед, ключ, пистолет	2 1	2 14
	3 кризис, опасность, проблема, положение, он, ситуация, колея, только	3 4	3 17
<b>увольнение</b>	0 арест, болезнь, выход, должность, заключение, заявление, измена, изменение	0 79	0 75
	1 в, на, работа, отпуск, освободить, после, я, он	1 1	1 5
<b>углеводород</b>	0 водород, воздух, кислота, окисел, он, пара, раствор, с	0 78	0 42
	1 ресурс, сырьё, природный, нефтяной, минеральный, топливо, и, нефтепродукт	1 2	1 38

Figure 3: Words with high COMPARE that acquired a new sense according to our predictions

## Appendix B Words with low COMPARE and no new senses predicted

In Figure 4 we show 10 words from RuShiftEval dataset with COMPARE < 2, for which our method did not find any new senses in the post-Soviet period compared to the pre-Soviet one, i.e. there is no clusters with at least 4 examples from the post-Soviet period and at most 2 examples from the pre-Soviet period.

Words	Cluster names	OLD COUNT	NEW COUNT
<b>дух</b>	0 физический, душить, ум, сердце, воля, здоровье, разум, тело	51	41
	1 стиль, форма, вид, природа, настрой, образ, характер, закон	17	21
	2 свой, себя, состояние, форма, сознание, настроение, настрой, сила	4	3
	3 ангел, бог, человек, я, закон, другой, он, себя	5	6
	4 кровь, я, другой, себя, слово, свой, в, мир	3	9
<b>наложение</b>	0 воздействие, помощь, с, без, действие, предупреждение, и, себя	5	2
	1 административный, администрирование, взыск, взыскание, возложение, возмещение	39	61
	2 воздействие, действие, нанесение, установка, наличие, использование	3	2
	3 с, и, без, изменение, наложение	1	1
	4 установка, изменение, создание, использование, удаление, и, нанесение, наличие	10	11
	5 снятие, без, не, нанесение, удаление, помощь, использование, установка	22	3
<b>полоса</b>	0 большой, в, весь, вокруг, волна, глаз, гора, граница	78	80
	1 город, они, поль, один, ряд, и, другой	2	0
<b>роспись</b>	0 бюджет, отчёт, сведение, список, статья, запись, имя, закон	73	20
	1 архитектура, декор, живопись, композиция, скульптура, украшение	7	60
<b>ссылка</b>	0 автор, адрес, имя, информация, описание, ответ, текст, указание	30	59
	1 казнь, лагерь, обратно, отпуск, смерть, убийство, арест, война	50	21
<b>тачка</b>	0 авто, автобус, автомашина, автомобиль, в, вагон, велосипед, весь	79	80
	1 и, другой	1	0
<b>хрен</b>	0 без, вода, гриб, кофе, масло, молоко, мясо, мёд	48	15
	1 кто, что, на, какой, мы, нет, ничто, бог	5	58
	2 брат, мужик, пёс, мой, ребёнок, идиот, друг, человек	27	7
<b>центр</b>	0 комплекс, дом, парк, место, символ, город, пол, и	8	11
	1 цель, главное, источник, частью, основа, главный, символ, граница	37	11
	2 верх, угол, конец, сторона, на, вне, пол, напротив	11	7
	3 столица, регион, район, область, напротив, ряд, вокруг, город	21	25
	4 институт, университет, другой, комплекс, регион, россия, дом, орган	3	26
<b>ядро</b>	0 внутри, днк, корень, молекула, основа, основание, поверхность, структура	38	71
	1 патрон, стрела, бомба, меч, гранат, ракета, рука, камень	42	9
<b>ясли</b>	0 детсад, детский, интернат, класс, лагерь, санаторий, училище, учреждение	24	60
	1 грудь, рука, вода, слово, человек, он, земля, дверь	25	5
	2 дорога, ворот, стол, дверь, двор, домик, комната, пол	16	8
	3 церковь, храм, комната, пол, он, грудь, рука, земля	14	6

Figure 4: Words with low COMPARE and no new senses predicted.

## Appendix C Examples of sentences for the words *бригада* and *углеводород*

Tables 6, 7 provide examples of sentences for two words with high COMPARE metric for which our method found a new sense. Each table contains 10 randomly sampled sentences from the pre-Soviet and post-Soviet time periods. From these examples we conclude that our algorithm correctly detected that a new sense appeared at least in the sampled examples. This sense is surely related to the old sense, but in our opinion it is not equal to it.

cluster ID	sentence
<b>Pre-Soviet</b>	
0	Нигде у нас нет резерва , и государь с <b>бригадою</b> пехоты стоит почти на аванпостах в середине пустого пространства ( верст с 50 до 60 ) , разделяющего войска тырновские ( Николай Николаевич ) от войск наследника .
0	Батальоны образцовые , но командир <b>бригады</b> генерал-майор Эллис так прирос к гвардейским порядкам , что не может свыкнуться с мыслью , что он не в Красном Селе : за два часа перед смотром и накануне весьма продолжительно производил репетиции прохождением церемониальным маршем с музыкою и хоровыми солдатскими ответами на начальнические приветствия .
0	На правом берегу Днепра действовала и II-ая конная армия под командой бывшего войскового старшины Миронова , в составе 2-ой , 16-ой и 21-ой кавалерийских дивизий и особой конной <b>бригады</b> .
1	В Скорцару и Янке : 7-я и 8-я легкая батареи 15-й артиллерийской <b>бригады</b> .
0	В поле не потеряется и возложенную на него задачу выполнит с успехом ; к подчиненным строго требователен , беспристрастен и справедлив ... ” и так много , и потом заключение : “ Достойн быть командиром неотдельной <b>бригады</b> ” .
0	На Кавказе , поступив в Куринский полк , я застал Бярятинского уже полковником , командовавшим одним из батальонов Кабардинского полка , который вместе с Куринским составлял <b>бригаду</b> , и хотя полки эти стояли в разных местностях , но их частям , кроме экспедиций , нередко приходилось встречаться в крепости Грозной , где была штаб-квартира командиров бригады и дивизии , причем последний был в то же время и начальником левого фланга Кавказского корпуса , и когда вместо генерала Фрейтаг [ а ] прибыл в Грозную начальником левого фланга генерал Нестеров , то , как знакомый ранее во Владикавказе с его семейством , я посещал его каждый раз по приходе в Грозную колонн Куринского полка из крепости Воздвиженской , причем мне приходилось встречаться в доме Нестерова и с князем Бярятинским .
0	Перед фронтом генерала Бабиева были обнаружены все части II-ой конной армии Миронова ( 2-ая , 16-ая и 21-ая кавалерийские дивизии и отдельная кавалерийская <b>бригада</b> ) .
0	Декабря 4 я назначен был командовать бригадою резервных рот 2 артиллерийской дивизии , и 23 артиллерийской бригады , я принял бригаду генваря 1829 г . 1829 год .
0	Командиры полков второй <b>бригады</b> так же , как и сам Вульферт , редко бывали с нами , потому что они шли сзади , на один переход , и являлись к Скобелеву , только когда догоняли нас на дневках .
0	Подлинный подписал : Начальник Главного штаба , артиллерии генерал-майор Ермолов Приказ по 1-й Западной армии 88 25 августа 1812 года Главная квартира В позиции при Бородине По воле главнокомандующего : Артиллерии генерал-майору Костенецкому поручается в командование <b>бригада</b> состоящая из Либавского и Софийского полков .
<b>Post-Soviet</b>	
0	В годы службы полковника в <b>бригаде</b> подчиненные не раз докладывали ему о том , что “ солдатское радио ” сообщало о потайных лазах в фундаменте гарнизонного клуба , через которые неуставленные военнослужащие якобы ходили в “ самоволку ” .
0	Короче , пока Витек пил в обкомовской гостинице , пытаюсь забыть девку , сбежавшую из профессорской семьи на строительство Камского гиганта , старшеклассницу , изуродованную передовой <b>бригадой</b> маляров коммунистического труда , онемевшую от боли , когда ее сунули руками в известь и держали так , потому что медленно работала , и бригаду из-за нее лишили премии и переходящего из рук в руки , как крановщица Кларка , вымпела .
0	Приезжала следственная <b>бригада</b> .
0	В ВВС и ПВО включены также три авиабазы , зенитная ракетная <b>бригада</b> , другие части и подразделения .
1	В тот год огромный их дом , похожий на бастион , был точно бы взят приступом и покорился орде наловатых , спешащих строительных <b>бригад</b> .
1	<b>Бригада</b> Усмана выезжала отныне только на подмогу обывателям при серьезных “ разборках ” , “ стрелках ” или “ наездах ” .
0	3-й ( Румелийский ) корпус : Пехота : 7 линейных полков ( 21 батальон ) , 7 стрелковых батальонов , Боснийская <b>бригада</b> ( 6 батальонов ) , Греческий волонтерский пограничный полк ( 3 батальона ) , Боснийский волонтерский пограничный полк ( 4 батальона ) , Никшичский албанский волонтерский батальон , Герцеговинский пограничный батальон .
1	За животными ухаживает целая <b>бригада</b> женщин , одна из которых – жена Майка , Джанин .
0	Не исключено , что в соседской квартире <b>бригада</b> выполняла отдельные работы .
1	Англичане специальную <b>бригаду</b> в Забайкалье прислать собираются , чтобы все тщательно подсчитать и внести читинский “ Локомотив ” в Книгу рекордов Гиннеса .

Table 6: Examples of sentences with the word *бригада*

cluster ID	sentence
<b>Pre-Soviet</b>	
0	– По изложении взгляда на изомерию ароматических <b>углеводородов</b> , там поясняется, “ что и в формуле бензола нет необходимости допущения реального существования остатков ” ( стр. 23 ) .
0	Из 8 г диметилангеликалактона получено около 2 г перегона, состоявшего из тиофенового <b>углеводорода</b> и тиофенового фенола .
0	Теперь я хотел бы особенно подчеркнуть, что такое различие единиц сродства атома углерода даже не во всех случаях необходимо для объяснения изомерии предельных <b>углеводородов</b> , если последняя действительно имеет место .
0	Лучшему выходу препятствует образование буро-красной смолы и <b>углеводородов</b> , которые всегда происходят во время реакции йодистого бутила на цианистую соль, и также образование бутиламина при действии соляной кислоты на продукт .
0	В кометных хвостах, которые, как известно, состоят преимущественно из газообразных <b>углеводородов</b> , мы имеем дело с отдельными молекулами, радиус которых $r \approx 10^8$ см, а плотность $d \approx 10$ , как показал Ф. Ехнер; в этом случае, однако, наша формула ( 5 ) неприменима во всей строгости, так как отдельные молекулы не суть абсолютно черные тела и радиус их мал сравнительно с длиной волны падающего на них света; поэтому мы можем только утверждать, что отталкивание хвостов, во много раз превышающее притяжение их, притом различное для различных веществ хвоста и обратно пропорциональное квадрату расстояния от Солнца, не противоречит нашей формуле .
0	Грэм нашел в воздухе двух копеек в Нью-Кестле 82,5-94,5 % легких <b>углеводородов</b> , 4,5-16,5 % азота и 1,0-1,3 % кислорода .
0	В первый раз было взято на 2 г <b>углеводорода</b> 5 г марганцовокислого калия, 8 г едкого натра и 1000 мл воды .
0	Такое превращение имеет место для эфиленна и для бутилена де Люинна; сюда же принадлежат и описанные отношения бутилена, открытого мною, но есть, однако, пример, что из <b>углеводорода</b> получается алкоголь, не тождественный, а изомерный с первоначальным; таковы отношения обыкновенного амильного алкоголя, амилена и амиленгидрата .
0	В обоих случаях мы должны искать причину изомерии кислот в изомерии аллиленов или двубромленных пропиленов; но если принять для пропиленна химическое строение, выраженное формулой, то углеродные атомы этого <b>углеводорода</b> не находятся все в одинаковых химических условиях, как это имеет место для углеродных атомов этилена .
0	Для превращения этого <b>углеводорода</b> в триметилкарбинол можно, однакоже, как мною указано уже и прежде, употребить серную кислоту, разведенную до некоторой степени водой ( около 5 ч. кислоты на 1 ч. воды ) .
<b>Post-Soviet</b>	
0	Нефтедобыча на водосборной площади р. Вах, служащей водоисточником г. Нижневартовска, стала причиной повышения предельно допустимого уровня содержания нефтяных <b>углеводородов</b> и тяжелых металлов в речной и питьевой воде [ 20 ] .
1	Суркова и А.А. Трофимук ( 1994 ) более 85 % запасов нефти в бассейне связаны с баженовско-неокомской нефтяной генерационно-аккумуляционной системой, характеризующейся превосходным по потенциалу источником <b>углеводородов</b> – баженовской кремнисто-глинистой толщей верхнеюрского возраста и морскими песчаниками неокомского возраста с высокими коллекторскими свойствами .
0	На второй стадии осуществляется очистка ВМС от остатков эмульсий жидких <b>углеводородов</b> в водной фазе .
1	По прогнозу Хьюджеса, пик разработки сланцевых <b>углеводородов</b> придется на 2017 год, после чего начнется падение, в результате которого за два года добыча упадет до уровня 2012-го .
1	В результате в России сейчас практически нет даже сколько-нибудь устоявшихся данных по запасам и ресурсам таких <b>углеводородов</b> , которые не вызывали бы сомнений и споров у экспертов ( а такие цифры есть, например, по США ) .
0	Так, проведенными ранее работами в различных климатических зонах было показано, что при температуре, близкой к оптимальной, утилизация <b>углеводородов</b> достигала 90-100 % всего за несколько дней ( table ) .
1	В условиях низких цен на <b>углеводороды</b> инструменты выживания ищут все .
0	В промышленных выбросах содержатся смертельно ядовитые вещества: окись углерода, двуокись азота, <b>углеводороды</b> .
0	В качестве первичного акта формирования залежи рассматривалось спонтанное газовыделение из растворов в области их закритического метастабильного насыщения низкокипящими <b>углеводородами</b> ( прежде всего, метаном ) .
1	– Судя по всему, Азербайджан обречен быть нашим главным конкурентом на рынке <b>углеводородов</b> .

Table 7: Examples of sentences with the word *углеводород*

## Appendix D Ungrammatical sentences after application of dynamic patterns

After application of dynamic patterns some sentences may become ungrammatical. This may affect the quality of generated substitutes. In this section we describe some preliminary study of this problem. We consider the disagreement in number between a verb and its subject, which is one type of grammatical errors that can be the result of application of dynamic patterns. We selected 3 real examples from the

RUSSE-2018 bts-rnc dataset and also made up one simple sentence ourselves. Each example has a singular verb and a singular subject. After one of the dynamic patterns is applied to the singular subject, it becomes plural and the agreement is broken. In tables 8, 9, 10 we show 10 most probable substitutes (after combining substitutes consisting of 1, 2 and 3 subwords) generated for this broken sentence and for the same sentence fixed manually.

Table 8 shows examples for the pattern "Т и <mask>". For our made-up example and the first real example the substitutes are syntactically plausible words, but mostly not related to the target word. After fixing the agreement more co-hyponyms appear among substitutes, which is better for representing the meaning of the target word. For the last two examples there are no significant changes in the substitutes after fixing the agreement. For the pattern "Т или <mask>" (table 9) substitutes differ less for correct and incorrect versions of sentences and seem to describe the meaning equally well. The pattern "<mask> (а также Т)" (Table 10) seems to be the most robust to incorrect agreement.

The inspected examples show that in some cases substitutes generated for a grammatical and ungrammatical versions of the same sentence can be significantly different. It seems that this effect depends on sentence complexity and dynamic pattern used. We plan to explore this effect in more details in our future work.

sentence	substitutes
Мальчик и <mask> ел шоколадное мороженное.	не: 0.51 сам: 0.12 никогда не: 0.08 с удовольствием: 0.06 вовсе не: 0.04 так: 0.03 правда: 0.03 больше не: 0.02 не только: 0.02 в детстве: 0.02
Мальчик и <mask> ели шоколадное мороженное.	девочка: 0.33 девушка: 0.26 его друзья: 0.23 мама: 0.10 женщина: 0.04 мать: 0.04 сестра: 0.03 его мама вместе: 0.03 она: 0.03 мальчик: 0.02
Пантюхин в Склифе сейчас. Он выползти на улицу успел, а на Золоткова балка и <mask> обрушилась. Эх, душой компании парень был! 28-летний Геннадий так и не	вовсе: 0.53 вовсе не: 0.09 крыша: 0.09 не: 0.07 так: 0.04 сама: 0.04 гора: 0.04 на него: 0.03 она: 0.02 без того: 0.02
Пантюхин в Склифе сейчас. Он выползти на улицу успел, а на Золоткова балка и <mask> обрушились. Эх, душой компании парень был! 28-летний Геннадий так и не	крыша: 0.13 гора: 0.12 лес: 0.06 стена: 0.05 дом: 0.05 все: 0.04 дерево: 0.03 город: 0.03 снег: 0.02 стены: 0.02
Впоследствии ее передали в Академию наук. Первая книжная лавка и <mask> была открыта в Гостином дворе на Петербургской стороне. Здесь продавали печатные указы	магазин: 0.40 библиотека: 0.25 аптека: 0.07 книгарня: 0.05 книжный магазин: 0.05 печатница: 0.05 рынок: 0.03 магазин книг: 0.02 столовая: 0.02 библиотеки: 0.02
Впоследствии ее передали в Академию наук. Первая книжная лавка и <mask> были открыты в Гостином дворе на Петербургской стороне. Здесь продавали печатные указы	магазин: 0.52 библиотека: 0.16 книжный магазин: 0.12 аптека: 0.10 рынок: 0.06 печатница: 0.05 первый магазин: 0.03 библиотеки: 0.03 магазины: 0.03 книгарня: 0.03
, нападения, грабежи Самара, 30, VIII. В селе Быковом Острого ограблена казенная винная лавка и <mask>. Похищено казенных денег 756 р., 12 ведер водки и собственных денег торговца	магазин: 0.44 торговец: 0.18 аптека: 0.06 склад: 0.05 торговый дом: 0.05 банк: 0.04 ресторан: 0.03 дом: 0.03 другой магазин: 0.02 другое имущество: 0.01
, нападения, грабежи Самара, 30, VIII. В селе Быковом Острого ограблены казенная винная лавка и <mask>. Похищено казенных денег 756 р., 12 ведер водки и собственных денег торговца	магазин: 0.45 торговец: 0.15 торговый дом: 0.09 аптека: 0.07 дом: 0.06 банк: 0.05 другой магазин: 0.05 склад: 0.04 ресторан: 0.02 почта: 0.02

Table 8: Examples with pattern "Т и <mask>"

sentence	substitutes
Мальчик или <mask> ел шоколадное мороженное.	девочка, который: 0.19 девочка не: 0.14 мальчик: 0.13 мальчика: 0.07 не: 0.07 девушка: 0.05 он: 0.04 женщина: 0.04 как: 0.03 я: 0.03
Мальчик или <mask> ели шоколадное мороженное.	мы: 0.35 девочка,: 0.13 девочка, которые: 0.09 Мы: 0.09 вы: 0.08 мальчики: 0.04 девушки: 0.03 Вы: 0.03 дети: 0.03 они: 0.02
Пантюхин в Склифе сейчас. Он выползти на улицу успел, а на Золотова балка или <mask> обрушилась. Эх, душой компании парень был! 28-летний Геннадий так и не	гора: 0.22 что: 0.06 что - то: 0.05 крыша: 0.05 дерево: 0.03 же: 0.03 дом: 0.02 камень: 0.02 как там: 0.02 стена: 0.02
Пантюхин в Склифе сейчас. Он выползти на улицу успел, а на Золотова балка или <mask> обрушились. Эх, душой компании парень был! 28-летний Геннадий так и не дворце в Летнем саду.	гора: 0.15 стена: 0.13 дом: 0.08 крыша: 0.06 мост: 0.04 дерево: 0.03 стены: 0.03 лес: 0.02 камни: 0.02 камень: 0.02
Впоследствии ее передали в Академию наук. Первая книжная лавка или <mask> была открыта в Гостином дворе на Петербургской стороне. Здесь продавали печатные указы	книжный магазин: 0.13 книжница: 0.12 рынок: 0.08 книгарня: 0.05 аптека: 0.04 салон: 0.02 базар: 0.02 книгария: 0.02 Библиотека: 0.01 библиотеко: 0.01
Впоследствии ее передали в Академию наук. Первая книжная лавка или <mask> были открыты в Гостином дворе на Петербургской стороне. Здесь продавали печатные указы	книжный магазин: 0.18 библиотека: 0.15 лавки: 0.08 книжница: 0.05 рынок: 0.04 аптека: 0.04 книгарня: 0.04 библиотеки: 0.03 Библиотека: 0.02 первый магазин: 0.01
, нападения, грабежи Самара, 30, VIII. В селе Быковом Остроге ограблена казенная винная лавка или <mask>. Похищено казенных денег 756 р., 12 ведер водки и собственных денег торговца	магазин: 0.29 склад: 0.08 дом: 0.03 заведение: 0.02 аптека: 0.02 ресторан: 0.02 торговец: 0.02 фабрика: 0.02 бар: 0.02 лавки: 0.01
, нападения, грабежи Самара, 30, VIII. В селе Быковом Остроге ограблены казенная винная лавка или <mask>. Похищено казенных денег 756 р., 12 ведер водки и собственных денег торговца	магазин: 0.31 склад: 0.05 дом: 0.05 аптека: 0.04 торговец: 0.03 ресторан: 0.03 торговый дом: 0.03 банк: 0.02 другой магазин: 0.02 магазин водки: 0.01

Table 9: Examples with pattern "Т или <mask>"

sentence	substitutes
Мальчик (а также <mask>) ел шоколадное мороженное.	мама: 0.18 его родители: 0.18 я: 0.08 папа: 0.07 девушка: 0.06 родители: 0.06 его бабушка: 0.05 отец: 0.04 мать: 0.03 мальчик: 0.03
Мальчик (а также <mask>) ели шоколадное мороженное.	его родители: 0.31 мама: 0.16 родители: 0.14 девушка: 0.09 его бабушка: 0.06 мать: 0.03 мальчик: 0.03 дети: 0.03 папа: 0.03 я: 0.02
Пантюхин в Склифе сейчас. Он выползти на улицу успел, а на Золотова балка (а также <mask>) обрушилась. Эх, душой компании парень был! 28-летний Геннадий так и не	на него: 0.11 она: 0.04 машина: 0.03 дом: 0.03 крыша: 0.02 он: 0.02 я: 0.02 дверь: 0.02 дома: 0.02 его: 0.02
Пантюхин в Склифе сейчас. Он выползти на улицу успел, а на Золотова балка (а также <mask>) обрушились. Эх, душой компании парень был! 28-летний Геннадий так и не дворце в Летнем саду.	крыша: 0.04 другие: 0.03 дома: 0.03 люди: 0.03 на него: 0.03 дом: 0.03 остальные: 0.02 город: 0.02 она: 0.02 он: 0.02
Впоследствии ее передали в Академию наук. Первая книжная лавка (а также <mask>) была открыта в Гостином дворе на Петербургской стороне. Здесь продавали печатные указы	библиотека: 0.35 магазин: 0.19 аптека: 0.15 рынок: 0.06 книжный магазин: 0.05 печатная: 0.03 книгарня: 0.03 книжница: 0.03 музей: 0.02 Библиотека: 0.02
дворце в Летнем саду. Впоследствии ее передали в Академию наук. Первая книжная лавка (а также <mask>) были открыты в Гостином дворе на Петербургской стороне. Здесь продавали печатные указы	библиотека: 0.25 аптека: 0.20 магазин: 0.12 другие: 0.07 рынок: 0.03 книжный магазин: 0.03 музей: 0.03 магазины: 0.03 другая: 0.02 библиотеки: 0.02
, нападения, грабежи Самара, 30, VIII. В селе Быковом Остроге ограблена казенная винная лавка (а также <mask>). Похищено казенных денег 756 р., 12 ведер водки и собственных денег торговца	магазин: 0.26 банк: 0.07 дом: 0.05 аптека: 0.05 другой магазин: 0.04 другие: 0.03 две другие: 0.03 почта: 0.03 банки: 0.03 деньги: 0.02
, нападения, грабежи Самара, 30, VIII. В селе Быковом Остроге ограблены казенная винная лавка (а также <mask>). Похищено казенных денег 756 р., 12 ведер водки и собственных денег торговца	магазин: 0.26 дом: 0.07 банк: 0.07 другой магазин: 0.06 аптека: 0.05 другие: 0.04 две другие: 0.04 банки: 0.04 почта: 0.03 торговец: 0.02

Table 10: Examples with pattern "Т (а также <mask>)"