# Diachronic Analysis of a Word Concreteness Rating: Impact of Semantic Change

Vladimir Bochkarev[1[0000-0001-8792-1491]], Stanislav Khristoforov[1[0000-0002-8995-6652]], Anna Shevlyakova[1[0000-0002-2659-1887]] and Valery Solovyev[1[0000-0002-7780-9823]]

[1] Kazan Federal University, 420008 Kazan, Russia
vbochkarev@mail.ru

**Abstract.** The paper analyses the correlation of change in word concreteness ratings with semantic change. To perform the analysis, we apply a neural network to diachronic data to obtain concreteness ratings of English words. As input to the model, we use co-occurrence statistics with the most frequent words extracted from the Google Books Ngram diachronic corpus. It is shown that the model, initially trained on data averaged over a long time interval, predicts the concreteness ratings with high accuracy (based on the word co-occurrence data in a particular year). The impact of lexical semantic change on the change in the concreteness rating is analyzed using 69 words borrowed from previous works. As the considered cases show, the neural network estimate of the word concreteness rating is very sensitive to changes in semantics. Among the factors that influence changes in the concreteness rating, we reveal the emergence of new meanings of a word, the competition of word meanings related to different parts of speech, the use of a word as a proper name, and the use of the word as a part of collocations. It is shown in the paper that changes in the concreteness rating can (along with changes in other properties) serve as a marker of semantic change.

**Keywords:** concreteness ratings, lexical semantic change, word co–occurrence, abstractness.

## 1 Introduction

Natural language constantly evolves, and its diachronic studies is relevant for modern science. Creation of large text corpora and development of computational tools allows one to perform large-scale analysis of language evolution. One the widely investigated spheres is semantic change.

Many works have studied evolution of semantics. Special competitions are held for computer programs to detect semantic changes [1]. By semantic changes we understand any changes in a word meaning (such as semantic shift, meaning narrowing/broadening etc). The standard approach to semantic change detection derives from the well-known distributive hypothesis [2,3] that suggests that words that are used and occur in the same contexts tend to purport similar meanings; at that, context change can be revealed by computational methods. Overviews of the state-of-the-art methods of automatic semantic change detection can be found in [4,5].

Different methods of semantic change detection have been developed in various papers. The article [5] distinguishes 3 groups of methods. The first one is static embedding, a single word representation with all semantic information. The second one is a topic model, a representation of a word that approximates different meanings. The third one is contextual embedding, the presentation of each individual use of a word with subsequent clustering of similar uses.

In all these cases, the investigation is performed in two steps. First, a certain word vector representation is built for different time intervals. Then, any change in the resulting representation over time is analyzed. However, another promising approach is possible. Thus, in the works by [6,7], it was proposed to use changes in certain properties of a word (number, case, tense) as markers of semantic change. A similar idea is developed in [8], which uses the category of animacy/inanimacy to trace semantic changes; and in [9] that discusses an algorithm for detecting new meanings of words associated with proper nouns. In this paper, we consider change of word concreteness as a marker of semantic change.

Word concreteness refers to the extent to which a word denotes an object that can be experienced by the senses [10]. Concrete nouns are words that denote objects and substances; they exist physically meaning they can be touched (for example, *book, table, cat)* and one can also get a picture of what the items are. Whereas abstract nouns are the opposite relating to things having no physical existence and non- representable (for example, *happiness, idea, truth*) [11].

Words are assigned concreteness ratings that are obtained by respondent`s survey. There are special dictionaries of concreteness ratings of words created for main languages of the world. For example, the most commonly used English dictionary is [12]. An up-to-date review of other dictionaries can be found in [13].

The issue of historical changes in the word concreteness rating was discussed in [14,15]. At the same time, [14] focuses the problem of whether the concreteness rating increases on average. In contrast to this work, we consider diachronic estimates of the concreteness rating of individual words and analyze how the concreteness rating responds to changes in word semantics. The study objective is to check the possibility of using changes in a word concreteness rating as a feature for solving the problem of lexical semantic change detection. To do this, firstly, we show how to obtain diachronic rating estimates. Concreteness ratings are estimated using word co-occurrence statistics based on the large diachronic corpus Google Books Ngram [16] by applying an algorithm proposed in a recent paper [13]. Secondly, we reveal the main factors that lead to a change in the concreteness rating, as well as perform classification of the considered examples.

## 2 Related Works

The problem of lexical semantic shift detection is widely discussed in scientific literature. A good overview of earlier works can be found in [17]. Numerous studies in the field of semantic change can be classified according to a number of criteria. As it was mentioned above in [5], change detection methods are classified depending on the used word vector representation. In [18], all works are divided into two groups: graded case when the degree of semantic change is determined; binary case when it matters whether the change occurs or not. In [19], another binary classification is proposed: 1) just revealing a change in semantics, or 2) also determining the type of change (new meaning, extension or narrowing of the meaning of a word).

The Google Books Ngram (GBN) corpus (https://books.google.com/ngrams) [20,16] and The Corpus of Historical American English described in [21] are most often used for such research. Most of the studies are performed for the English language. There are 27 datasets described in [4], most of which (24 datasets) are English.

In some papers, a change in the semantics of a words is associated with a change in its other properties. Perhaps one of the most sensitive markers of semantic change is a concreteness rating.

The concept of word concreteness plays an important role in psycholinguistic research. Concreteness ratings are used in various neurophysiological, medical, and psychological surveys to study the structure of word representations in human memory [22]. Data for dictionaries with estimates of concreteness ratings of words were obtained from participants [12,23]. For each word, at least 25 estimates are got on a 5- or 7-point scale. However, such data are not available for past centuries. Therefore, for diachronic studies, estimates of concreteness ratings can be obtained by extrapolation using various machine learning methods. An overview of works on machine estimation of concreteness ratings can be found, for example, in [13].

Diachronic corpus studies were carried out to identify changes in word concreteness ratings over time. The evolution of word semantics using contemporaneous, decade-specific computational estimates of word concreteness was studied in [14], where it was shown that distinct word types of the English language become increasingly more concrete over time and relatively concrete words tend to be used more often than abstract ones. Strict quantitative estimates have also been obtained which show that word concreteness increased by 13.5% over 150 years. To obtain diachronic estimates of word concreteness ratings, this paper uses the SentProp [24] algorithm, which combines a well-known method of label propagation with advances in word embeddings.

A valuable tool (Macroscope) for studying historical changes in language over the last two centuries was developed in [15]. Based on the traced changes in a word co-occurrence, the Macroscope provides diachronic quantitative information about changes in a word’s valence, arousal, and concreteness.

Fukugawa et al [25] evaluated the role of concreteness in predicting the direction of semantic change and found that it is more reliable and accurate than valence, and frequency. It was also found that in 70% of the attested cases of semantic change, the words became more concrete. In this paper, the analysis is carried out a posteriori for cases of lexical semantic change detected in earlier works. At that, the authors of [25] approximate the word ratings with their contemporary values extracting ratings from the database [12]. Therefore, changes in ratings over time are not considered in [25].

# 3 Data and method

## 3.1 Model Architecture

Our paper uses the method for concreteness rating estimation presented in [13]. Therefore, we will only briefly describe the method; a detailed description can be found in [13].

To train the model, we use the dictionary presented in [12] which includes 40,000 English words with concreteness ratings. To build a vector representation of a particular word, we used data on the word co-occurrence statistics in the GBN corpus. The method of co-occurrence with the most frequent words (CFW) is described in detail, for example, in [26,27,15]. According to this method, a vector that represents a word is constructed using frequency values of all bigrams including this word. There can be two types of bigrams - Wx and xW – where a target word (W) occurs before or after the context word (x). Context words are the most frequent ones selected for the study. As in [13], we choose 20,000 English context words that were the most frequent in 1900-2019, according to Google Books Ngram.

Bigram frequency data is extracted from the English corpus of Google Books Ngram. In the process of the model training, we use frequencies averaged over the period 1900-2019. Bigram frequencies have a very large dynamic range of values - from units to tens of thousands, which makes it difficult to process such data. In [28], it was proposed to use vectors composed of pointwise mutual information (PMI) values. Following [13], we use a regularized version of PMI:

$$\log_2\left(\frac{f_{wx}}{f_w f_x}+1\right)$$

Here, $f_{wx}$ is a relative frequency of a bigram consisting of the target word $W$ and the context word x, $f_w$ and $f_x$ are relative frequencies of both words.

To estimate the target parameter, a multilayer neural network of direct propagation is used. The dimension of the input vector of the neural network is equal to the doubled the number of context words. Thus, the dimension is 40,000. The architecture of the subsequent layers of the neural network was as follows: 4 hidden layers, each of which had a non-linear activation by the ELU function. The number of neurons in each layer was 128. The last output layer had a dimension of 1 and had a linear activation. The number of weights in the network is large, and most of them are in the first hidden layer. To avoid network retraining, a dropout layer [29] with a parameter of 0.1 was placed between the input and the first hidden layer. Thus, at a time, only 90% of the data from the input vector got to the input of the network in the training mode. This ensured regularization of the model and, as a result, an increase in the accuracy of estimating the concreteness rating.

The loss function was the root mean square error between the model outputs and the target values of concreteness ratings. The minimization of the functional was carried out on the basis of the stochastic gradient descent method with a batch size of 128. The Adam method [30] was used as an optimization algorithm. The optimization process was stopped based on the results of observing the dynamics of the loss function on the test set. If during 100 optimization iterations the values of the loss function did not decrease, then the optimization process was stopped and the best set of network weights was used to estimate the quality of the resulting model.

### 3.2 Learning and Testing of the Model on Synchronous Data

We selected 36,997 words from the [12] database for training and testing. These words must be used in the Google Books Ngram corpus at least in 3 different bigrams within 1900-2019 and their total frequency for the same time interval must be not lower than 30. Following [8], the set of words were divided into 6 groups, so that all forms of each lemma fell into only one of these groups. Next, the list of words was divided into training and test samples in a ratio of 2 to 1. Thus, each model was trained on two-thirds of all words. In total, 4 groups can be selected in 15 ways from the 6 groups, therefore, we obtain 15 models trained on different subsets of words. At the same time, for any word, there are 5 independently trained models for which this word belongs to the test set. Firstly, this allows for cross-validation of the training results. Secondly, it is important that we can further obtain unbiased diachronic rating estimates for an arbitrarily chosen word.

A neural network concreteness rating model with a similar architecture was tested in detail in [13]. The model used in this paper slightly differs from the model in [13]. Having trained the model, first, we tested it on synchronous data as in [13]. For each model, we calculated concreteness rating estimates on its test set, and compare the obtained values with human ratings.

The average value of the Spearman coefficient between word concreteness ratings and their estimates on the test sample was 0.8830. This almost coincides with the result obtained in [13]. We can also average the obtained concreteness ratings over the 5 models. The correlation coefficient between the word concreteness ratings and their average rating estimates was 0.8924. Thus, the use of averaging over several models allows one to obtain more accurate result than in [13].

### 3.3 Applying the Model to Annual Data

In this paper, unlike [13], we apply the model trained in the described way to diachronic data. To do this, frequencies of bigrams including target words were extracted from the Google Books Ngram corpus for each year from the interval 1880-2019. For each word and each year, we get 5 estimates of concreteness rating, and calculate the mean estimate and its standard deviation.

It is natural to expect that for input vectors built for data for one year, the accuracy will be lower than for vectors built for data averaged over a large time interval. However, the accuracy of the model on annual data turns out to be unexpectedly high. Figure 1 shows the values of the correlation coefficient between the word concreteness ratings given in [12] and their estimates calculated using the data for a given year. The Spearman correlation coefficient varies from 0.8878 to 0.8922 within the last twenty years, which is slightly lower than the coefficient value of 0.8924 obtained for the data averaged over 1900-2019.

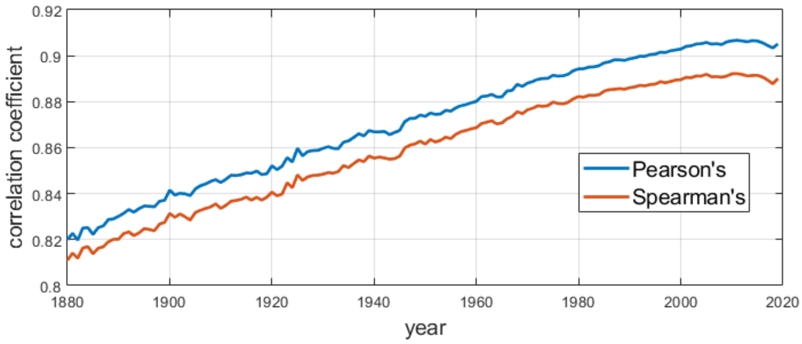


**Fig. 1.** Correlation coefficient between the word concreteness ratings given in [12] and their estimates calculated using the data for a particular year

An example of the resulting dependence (for the English word *plane*) is shown in Figure 2,A. Initially the word *plane* was a noun meaning “flat surface, simplest of all geometrical surfaces” [31]. In 1903, the first airplane was created, and the word plane (short form of airplane) started being widely used as “a powered, fixed-wing aircraft”[1]. Airplanes have become more and more widespread since the 1910s. The jump of the concreteness graph is obviously associated with the emergence of a new meaning. Obviously, aircrafts are more concrete than plane in the geometrical sense.

With the increase in the frequency of use of the word *plane* in the new meaning, its concreteness rating increased from 3.2-3.3 in the late 19th century to 4-4.2 in the second half of the 20th and early 21st centuries. Therefore, there is a correlation between the meaning change and change in the word concreteness rating. This short example was an illustration to our work.

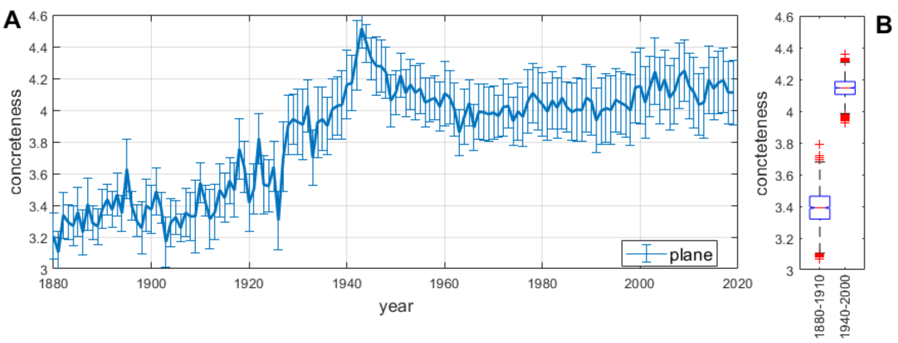


**Fig. 2.** A) Change in the concreteness rating of the word *plane* over time; B) Distribution of concreteness rating estimation for 1880-1910 and 1940-2000

In our work, we used two word sets. The first one is a dataset presented at SemEval-2020 by D. Schlechtweg et al. [1]. It is widely used for testing lexical se-

[1] https://en.wikipedia.org/wiki/Plane

mantic change detection algorithms. It is stated that 16 words out of the 37 words presented in SemEval-2020, have gained new meanings. The rest 21 words that haven`t gained new meanings are control ones. However, this markup refers only to the meanings of 37 words used in the small number of texts included in the SemEval-2020 dataset. In fact, most of the words included in the SemEval-2020 texts are polysemantic. Moreover, according to etymological dictionaries, many of them, gained new meanings in the XIX-XXI centuries. The second set of words was borrowed form [32] and consists of 32 words.

Having obtained a time series of concreteness rating for the target word, we used the algorithm described in [9] for change point detection. Having identified changes in the rating, we tested the significance of the changes using the [33] algorithm. In this algorithm, for each of the compared time intervals using bootstrapping, sets of bigram frequency vectors are generated, for each of which a concreteness rating estimate can further be calculated. Thus, we simulate the empirical distribution of the concreteness rating estimate for the compared time intervals which makes it possible to test the hypothesis about the significance of the observed changes. For example, Figure 2,B shows the ranges of estimates obtained for the time intervals 1880-1910 and 1940-2000.

## 4 Result

We calculated concreteness rating of target words over time. Then, we interpreted significant changes in the concreteness rating using data from etymological dictionaries [31,34,35], as well as using contexts of use of these words in the GBN corpus. The available examples can be roughly divided into four groups (see Table 1). In particular, concreteness rating graphs are sensitive: 1) to changes in part-of-speech to which a word belongs to, 2) to transition of a noun from the category of common nouns to the category of proper names, 3) the use of a word in fixed expressions (collocations) and 4) changes in semantics under the influence of other factors. Let us provide examples for each of the considered trends.

**Table 1.** Classification of reasons that influence the concreteness ratings of the target words

| The reason of concreteness rating change | Kulkarni et al. [32] | SemEval-2020 |
|---|---|---|
| POS Change | 14 | 12 |
| Common noun vs proper noun | 6 | 2 |
| Fixed expressions (Collocations) | 3 | 6 |
| Other factors of semantic change | 17 | 18 |
| Unclear cases | 2 | 1 |
| No significant rating change | 1 | 8 |

If one sums up the numbers in the columns of the table, it will be higher than the number of examples considered. This happens because in many cases a word undergoes a number of semantic changes during the considered time interval, and it can be assigned to more than one class.

### 4.1 POS Change

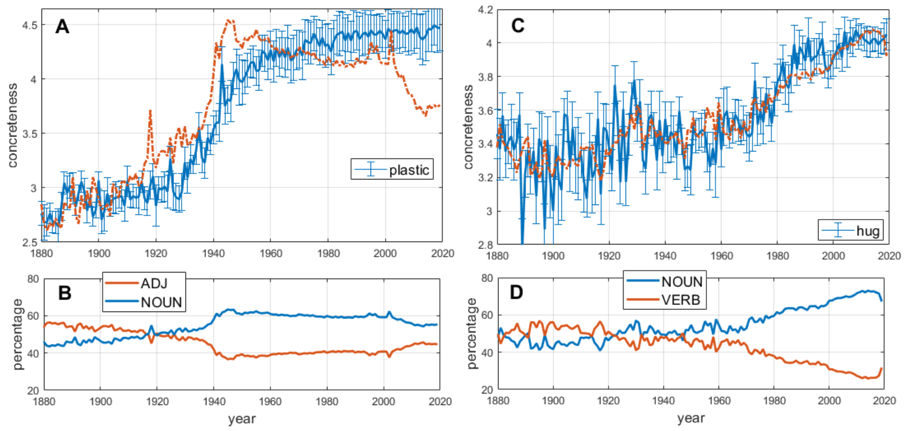


**Fig. 3.** Change in the concreteness ratings of the words *plastic* (A) and *hug* (C); Change in the percentage of use of the words *plastic* (B) and *hug* (D) as different parts of speech

The first group of words includes those that changed their grammatical meaning and shifted from one POS to another.

Let us analyze the word *plastic*. Figure 3,A shows a concreteness rating graph for this word. Information about frequency of use of *plastic* used as a particular part of speech was extracted from the GBN corpus. Figure 3,B shows the percentage of use of this word as one POS or another. According to the GBN corpus, the word *plastic* was used more frequently as an adjective until around 1920 and meant flexible, amenable [34]. In 1855, the first plastic was obtained which started being widely used after the 2nd World War. The word gained a new meaning and was used more often as a noun. The concreteness rating graph reflects this. Since about 1945-1950, one can observe a sharp increase in the word concreteness, coinciding with the increasing use of *plastic* as a noun.

We fit a linear regression model of the following form:

$$\hat{c}(t) = \sum_i a_i p_i(t) + b$$

Here, $\hat{c}(t)$ is the concreteness rating estimate, $p_i(t)$ is the percentage of use of the target word as the $i$-th part of speech, $a_i$, $b$ are constant coefficients. The regression line is shown in Figure 3,A as a dash-dotted curve. As can be seen from the figure, the model fits the concreteness rating plot well. The Pearson and Spearman correlation coefficients between the concreteness rating estimate and the percentage of plastic used as a noun are 0.869 and 0.653, respectively.

Another example is the word *hug* (see Figures 3,C and 3,D) which means the act or process of holding someone or something close to your body with your arms [35]. If one analyses the data from 1880 on the frequency graph, one will see that this word was used more often as a verb until about 1920. And starting from 1920, the tendency to use it as a noun prevails. This trend is also reflected in the concreteness graph. The concreteness of *hug* is growing. The Pearson and Spearman correlation coefficients between the concreteness rating and the percentage of hug used as a noun are 0.856 and 0.854, respectively.

These examples show that the change in a word meaning associated with the transition of a word from one part of speech to another is reflected in the concreteness graphs.

Both examples are characterized by an increase in the percentage of the use of target words as nouns, which is accompanied by an increase in the concreteness rating. In total, among the 69 examples considered, there are 9 cases of noun/verb competition and 5 cases of noun/adjective competition. In all of these examples, the concreteness rating is positively correlated with the percentage of the target word used as a noun.

### 4.2 Common Noun vs Proper Noun

The second group of words include those that have changed their grammatical meaning, in particular shifted from the category of common nouns to proper nouns (or vice versa).

The first case is the word *Bush*. *Bush* in English originally means a low plant with many branches that arise from or near the ground [34] when used as a common noun. However, this word is also used as a surname of German origin, and in particular it is the surname of a famous political dynasty in the United States. The jumps in the concreteness ratings seen in Figure 4,A is related to media coverage of George H. W. Bush and George W. Bush. Figure 4,B shows a percentage graph of the uppercase use of the word *Bush*. The Pearson and Spearman correlation coefficients between the concreteness rating estimate and the percentage of the uppercase use of the word *bush* are -0.870 and -0.874, respectively. As with the previous group of examples, we fit a regression model of the following form:

$$\hat{c}(t) = a\,p(t) + b$$

Here, $\hat{c}(t)$ is the concreteness rating estimate, $p(t)$ is the percentage of the uppercase use of the target word, a, b are constant coefficients. The regression line is shown in Figure 4,A as a dash-dotted curve.

The second example is *Windows*. *Windows* means “spaces usually filled with glass in the wall of a building or in a vehicle, to allow light and air in and to allow people inside the building to see out” [35]. It has been traditionally used as a common noun. However, in 1985, Microsoft released the first version of graphical operating system called Windows (see Figure 4,C). Since that time, the word *windows* has been widely used as a proper name. The capitalization graph reflects a growing use of the capitalized *Windows* since mid 80s with a peak in 1998 (see Figure 4,D). The concreteness rating graphs shows a respond to this change; it is seen that the concreteness of *Windows* is gradually falling at between 1985–1998. The Pearson and Spearman correlation coefficients between the concreteness rating and the percentage of the uppercase use of the word *windows* are -0.900 and -0.887, respectively. Thus, the graphs of both words show correlation between changes in grammatical meaning and concreteness ratings.

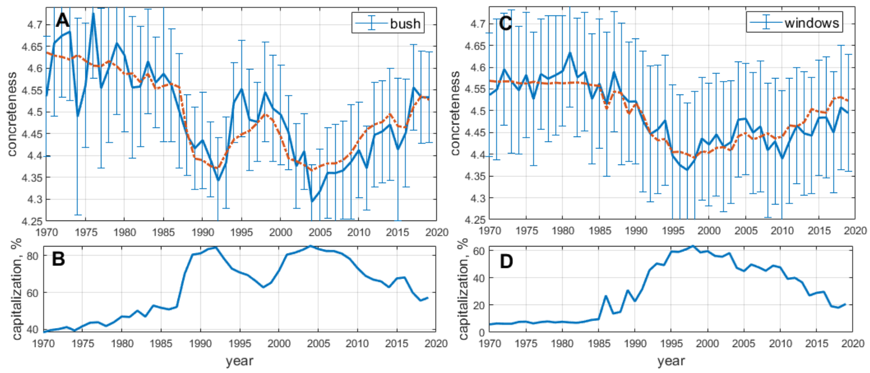


**Fig. 4.** Change in the concreteness rating estimate for the words *bush* (A) and *windows* (C); Change in the percentage of the uppercase use of *bush* (B) and *windows* (D) from the total number of uses

We have already said that one and the same word can be assigned to different groups which we distinguished above. However, there are 6 nouns that relate only to the group of words characterized by the competition between common and proper nouns. There is a negative correlation between the concreteness rating and the percentage of use of a word as a proper name observed for 2 words of the mentioned 6. The rest 4 words show positive correlation. It should be noted that the last 4 words initially had a very high (4.1-4.9) concreteness rating.

### 4.3 Collocations

The third described group of words that reflects the trends includes fixed expressions (collocations). An interesting observation is that significant changes in the concreteness rating are mostly associated with the emergence or increase in the frequency of collocations in which the target word is capitalized (for example, *Christmas Tree*, *Risk factors*, *Science Fiction*, etc.). Therefore, for examples in this group, the concreteness rating graph tends to have a high correlation with the percentage of the uppercase use of the target word. Contextually, these collocations can mean, for example, technical and other terms, cultural objects etc.

The first case is represented by the word *tree*. *Tree* basically means a tall plant that has a wooden trunk and branches growing from its upper part [35]. It is a common noun, usually written using a lower case. In this meaning, *tree* is a highly concrete noun (see Figure 5,A). However, it is seen that it`s concreteness rating graph fluctuates that is due to capitalization of the word (see Figure 5,B). We find *Tree* in some set expressions like *Christmas Tree*, *Tree Planters* etc. The Pearson and Spearman correlation coefficients between the concreteness rating estimate and the percentage of the uppercase use of the word tree are 0.862 and 0.839, respectively.

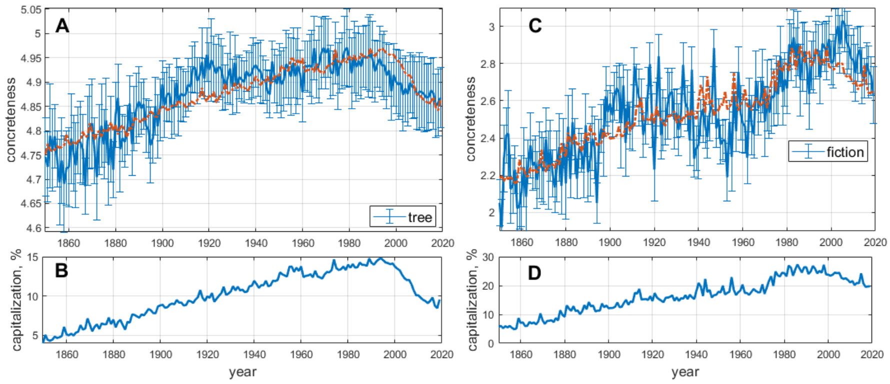


**Fig. 5.** Change in the concreteness rating estimate for the words *tree* (A) and *fiction* (C); Change in the percentage of the uppercase use of the words *tree* (B) and *fiction* (D) from the total number of uses

The second target word is *fiction* which means “the type of book or story that is written about imaginary characters and events and not based on real people and facts” or “a false report or statement that you pretend is true” [35]. The change in the concreteness rating of the target word is primarily associated with the expressions *Science Fiction*, *Modern Fiction*, *Short Fiction*, etc., which are usually written with a capital letter. The Pearson and Spearman correlation coefficients between the concreteness rating estimate and the percentage of the uppercase use of the word *fiction* are 0.805 and 0.792, respectively (see Figure 5,C, 5,D).

There is a positive correlation between the concreteness rating estimation and the frequency of collocations that include the target word observed for 6 words out of 8 belonging to this group.

## 4.4 Other Cases of Semantic Change

Let us consider changes in semantics that are not related to the previous cases. It means that a target word still belongs to one part of speech, did not shift to the category of proper names and is not a part of collocations.

The example of such words is a word *delivery* (see figure 6,A). It is a polysemantic word with various meanings (see [35]). However, the concreteness graph shows drop of concreteness ratings that can be interpreted by appearance of some new widely used meaning. To our mind, it is associated with a meaning “the act of taking goods, letters, parcels, etc. to people's houses or places of work”.

The second word is *head* (see figure 6,B). It derived from the Old English word *heafod* and it`s primary meanings are “top of the body,” also “upper end of a slope,” and “chief person, leader, ruler” [31]. Later this word gained other meanings [35]. The concreteness graph of this word shows that it has become less concrete. This can be explained by meaning competition: metaphorical, more abstract meanings start to prevail other the rest ones within the last two decades.

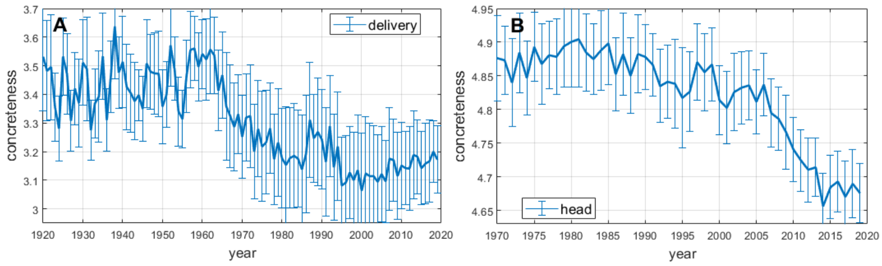


**Fig. 6.** Change the concreteness rating estimate for the words *delivery* (A) and *head* (B)

# 5 Conclusion

Semantic change has been a topic of interest for many scientists. A number of studies have shown that a change in semantics leads to changes in other word properties that can be revealed by computational methods and which, in turn, can be used as markers of semantic changes. In this article, we consider concreteness of words and demonstrate that it is very sensitive to changes in semantics. The study was carried out using deep learning neural networks based on data from the Google Books Ngram corpus. A detailed analysis of changes in the word concreteness rating over time was carried out using a set of 69 words that are commonly used as test words for evaluating algorithms for detecting semantic changes [1,32]. Except for 3 words, we managed to offer a clear interpretation of significant changes in the concreteness rating for all the considered examples.

We distinguished four classes of reasons that influence changes of concreteness ratings of words. Moreover, for the three classes, change in the concreteness rating correlates with changes in other well-defined word properties. As the analysed examples show, the concreteness rating can respond to the emergence of a new word meaning or competition between old meanings of a word. The highest sensitivity is observed when a word changes its part-of-speech attribution. When meanings of a word expressed by noun/verb or noun/adjective compete, a higher percentage of nouns increases the word concreteness rating.

Also, the concreteness rating changes if being a common noun, the word starts being used as a proper name. Finally, the concreteness rating responds to the appearance of new frequently used collocations that include a given word. It can be, for example, technical and other terms, cultural objects etc.

Thus, change in the concreteness rating can be a marker of lexical semantic change, but only in combination with an analysis of other properties of a word.

**Acknowledgements**

This research was financially supported by Russian Science Foundation, grant № 20-18-00206.

## References

1. Schlechtweg, D., McGillivray, B. Hengchen, S., Dubossarsky, H. and Tahmasebi, N.: SemEval-2020 task 1: Unsupervised lexical semantic change detection. In: Proceedings of the Fourteenth Workshop on Semantic Evaluation, pp. 1-23. International Committee for Computational Linguistics, Barcelona (2020).
2. Harris, Z.: Papers in structural and transformational Linguistics. Reidel, Dordrecht (1970).
3. Firth, J.R.: A Synopsis of Linguistic Theory, Studies in Linguistic Analysis 1930-1955. Special volume of the Philological Society, 1-32 (1957).
4. Tahmasebi, N., Borin, L. & Jatowt, A.: Survey of computational approaches to lexical semantic change detection. In: Tahmasebi, N., Borin, L., Jatowt, A., Xu, Y. & Hengchen, S. (eds.) Computational approaches to semantic change. (Language Variation 6), pp. 1–91. Language Science Press, Berlin (2021).
5. Hengchen, S., Tahmasebi, N., Schlechtweg, D., Dubossarsky, H.: Challenges for Computational Lexical Semantic Change. In: Tahmasebi, N., Borin, L., Jatowt, A., Xu, Y. & Hengchen, S. (eds.). 2021. Computational approaches to semantic change. (Language Variation 6), pp. 341–372. Language Science Press, Berlin (2021).
6. Ryzhova, A., Ryzhova, D., Sochenkov, I.: Detection of Semantic Changes in Russian Nouns with Distributional Models and Grammatical Features. Komp'juternaja Lingvistika i Intellektual'nye Tehnologii 20, 597–606 (2021).
7. Kutuzov, A., Pivovarova, L. and Giulianelli, M.: Grammatical Profiling for Semantic Change Detection. In: Proceedings of the 25th Conference on Computational Natural Language Learning, pp. 423–434. Association for Computational Linguistics (2021).
8. Bochkarev, V., Achkeev, A., Shevlyakova, A., Khristoforov, S. Diachronic Neural Network Predictor of Word Animacy. In: Pichardo Lagunas, O., Martínez-Miranda, J., Martínez Seis, B. (eds) Advances in Computational Intelligence. MICAI 2022, LNCS, vol. 13613, pp. 215-226. Springer, Cham (2022). doi: 10.1007/978-3-031-19496-2_16
9. Bochkarev, V.V., Khristoforov, S.V., Shevlyakova, A.V. and Solovyev, V.D.: Neural Network Algorithm for Detection of New Word Meanings Denoting Named Entities. IEEE Access 10, 68499-68512 (2022). doi: 10.1109/ACCESS.2022.3186681.
10. Paivio, A., Yuille, J.C. and Madigan, S.A.: Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology 76(1, Pt.2), 1–25 (1968). doi: 0.1037/h0025327
11. Paivio, A.: Mental imagery in associative learning and memory. Psychological Review 76, 241-263 (1969). doi:10.1037/h0027272
12. Brysbaert, M., Warriner, A. B. and Kuperman, V.: Concreteness ratings for 40 thousand generally known English word lemmas. Behavior Research Methods 46(3), 904–911 (2014).

13. Bochkarev, V.V., Khristoforov, S.V., Shevlyakova, A.V. and Solovyev, V.D.: Comparison of the three algorithms for concreteness rating estimation of English words. Acta Polytechnica Hungarica 19(10), 99-121 (2022).
14. Snefjella, B., Généreux, M. & Kuperman, V.: Historical evolution of concrete and abstract language revisited. Behavior Research Methods 51, 1693–1705 (2019). doi: 10.3758/s13428-018-1071-2
15. Li, Y., Engelthaler, T., Siew, C. S., & Hills, T. T.: The Macroscope: A tool for examining the historical structure of language. Behavior Research Methods 51, 1864–1877 (2019). doi: 10.3758/s13428-018-1177-6
16. Lin, Y., Michel, J.-B., Aiden, E.L., Orwant, J., Brockman, W., Petrov, S.: Syntactic Annotations for the Google Books Ngram Corpus. In: Li, H., Lin, C.-Y., Osborne, M., Lee, G.G., Park, J.C. (eds.) 50th Annual Meeting of the Association for Computational Linguistics 2012, Proceedings of the Conference, vol. 2, pp. 238-242. Association for Computational Linguistics, Jeju Island, Korea (2012).
17. Tang, X.: A state-of-the-art of semantic change computation. Natural Language Engineering 24(5), 649-676 (2018).
18. Schlechtweg, D. and Schulte im Walde, S.: Simulating Lexical Semantic Change from Sense-Annotated Data. In: The Evolution of Language: Proceedings of the 13th International Conference (EVOLANGXIII). (2020).
19. Boleda, G.: Distributional Semantics and Linguistic Theory. Annu. Rev. Linguist. 6, 213–234 (2020). doi: 10.1146/annurev-linguistics-011619-030303
20. Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., et al.: Quantitative analysis of culture using millions of digitized books. Science 331(6014), 176–182 (2011).
21. Davies, M.: Expanding Horizons in Historical Linguistics with the 400 million word Corpus of Historical American English. Corpora 7, 121-57 (2012).
22. Solovyev, V.: Concreteness/Abstractness Concept: State of the Art. In: Velichkovsky, B.M., Balaban, P.M., Ushakov, V.L. (eds) Advances in Cognitive Research, Artificial Intelligence and Neuroinformatics. Intercognsci 2020, Advances in Intelligent Systems and Computing, vol 1358, pp. 275-283. Springer, Cham (2021).
23. Coltheart, M.: The MRC psycholinguistic database. The Quarterly Journal of Experimental Psychology Section A 33(4), 497–505, (1981). doi: 10.1080/14640748108400805.
24. Hamilton, W. L., Clark, K., Leskovec, J., & Jurafsky, D.: Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 595–605. Association for Computational Linguistics, Austin, Texas (2016).
25. Fugikawa, O., Hayman, O., Liu, R., Yu, L., Brochhagen, T. and Xu, Y.: A computational analysis of crosslinguistic regularity in semantic change. Front. Commun. 8, 1136338 (2023). doi: 10.3389/fcomm.2023.1136338
26. Xu, Y. and Kemp, C.: A Computational Evaluation of Two Laws of Semantic Change. In: Proceedings of the 37th Annual Meeting of the Cognitive Science Society, CogSci 2015, Pasadena, California, USA, July 22-25, (2015).
27. Khristoforov, S., Bochkarev, V. and Shevlyakova A.: Recognition of Parts of Speech Using the Vector of Bigram Frequencies. In: van der Aalst W. et al. (eds) Analysis of Images, Social Networks and Texts. AIST 2019. CCIS, vol. 1086, pp. 132-142. Springer, Cham (2020). doi: 10.1007/978-3-030-39575-9_13
28. Bullinaria, J.A. and Levy, J.P.: Extracting semantic representations from word co-occurrence statistics: Stop-lists, stemming, and SVD. Behavior Research Methods 44(3), 890-907 (2012).

29. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 15(1), 1929-1958 (2014).
30. Kingma, D.P., Ba, J. Adam: A Method for Stochastic Optimization, arXiv preprint, arXiv:1412.6980, 2014, http://arxiv.org/abs/1412.6980.
31. Online Etymology Dictionary, https://www.etymonline.com/, last accessed 2023/05/30.
32. Kulkarni, V., Al-Rfou, R., Perozzi, B., Skiena, S.: Statistically significant detection of linguistic change. In: Proceedings of the 24th International Conference on World Wide Web, pp. 625–635. Florence, Italy (2015).
33. Bochkarev, V. and Shevlyakova, A.: Calculation of a confidence interval of semantic distance estimates obtained using a large diachronic corpus. J. Phys.: Conf. Ser. 1730, 012031 (2021). doi: 10.1088/1742-6596/1730/1/012031
34. Dictionary.com, https://www.dictionary.com/, last accessed 2023/05/30.
35. Cambridge Dictionary, https://dictionary.cambridge.org/, last accessed 2023/05/30.