

Conference Paper

Teaching the Native and Foreign Language at High School Using the Possibilities of Case Linguistics

Dmitrieva Yuliya Yurievna¹ and Yakovleva Lyubov Anatolyevna²

¹National Research University “Moscow Power Engineering Institute”, Moscow, Russia

²Technical Institute (branch) of NEFU in the city of Neryungri, Neryungri, Russia

Abstract

Corpus linguistics is a modern direction in the field of linguistics, which involves the accumulation of texts in this language with grammatical markup and the possibility of advanced search. The corpus implies the possibility of quick implementation of statistical sampling for any period of interest to the researcher, as well as the collection and analysis of contexts for the use of the studied word or group of words.

The ability to collect corpuses of texts and use them is inextricably linked with the technical progress of the second half of the twentieth century: in fact, Internet search engines use the technology of intelligent search, largely due to the achievements of corpus linguistics.

The possibilities of the various languages collecting corpuses are practically not mastered in the teaching of both the native and foreign languages. This article proposes an option to use the capabilities of the National Corpus of the Russian Language and the British Corpus in order to motivate students and perform mini-linguistic research.

The indicated technology has been repeatedly tested at bachelors and masters of linguistic and philological specialties with a positive result. The introduction of interaction with the corpus of texts into classes positively affects both the students' research competence and the development of skills in mastering modern technologies and the capabilities of the digital society.

The mini-studies can even be conducted by first-year bachelors, since they do not require deep linguistic knowledge. When presenting the results of their research, students can go on to compare and research the received statistical data. When mastering both a foreign and a native language, working with the corpus makes it possible to look at words from a new angle and demonstrate the relationship between the history of society and the history of words.

Keywords: Teaching methodology, corpus linguistics, foreign language, native language.

Corresponding Author:
 Dmitrieva Yuliya Yurievna
 yydmitrieva@mail.ru

Published: 21 January 2021

Publishing services provided by
Knowledge E

© Dmitrieva Yuliya Yurievna and
 Yakovleva Lyubov
 Anatolyevna. This article is
 distributed under the terms of
 the [Creative Commons
 Attribution License](#), which
 permits unrestricted use and
 redistribution provided that the
 original author and source are
 credited.

Selection and Peer-review under
 the responsibility of the XXIII
 International Conference
 Conference Committee.

1. Introduction

Various languages texts corpuses began to emerge along with the possibility of digitalizing texts and marking in them a part-line affiliation, removing homonymy, etc. The corpus is an electronically embodied hypothetical array of speech activity, the existence

 **OPEN ACCESS**

of which was mentioned by Ferdinand de Saussure [10], speaking about the difference language and speech. Subsequently, the theory of speech activity was developed in the works of German linguists [6].

Hypothetically, the speech activity of a certain language is representable as a virtual constantly replenished array of texts in which everything written and said in this language is stored. The presence of such an array would allow us to quickly studying the development trends of the language, especially its evolution [7]. Also, such an array of texts would allow us correctly perceiving and tracking the functionality of certain language elements.

The appearance of various languages text corpus has led to the implementation of this hypothetical paradigm in practice [1]. Of course, the corpus does not cover all 100% of texts in a particular language, but even the texts contained in them allow us to draw some statistical conclusions and track certain trends.

2. Methods and Methodology

When mastering the possibilities of corpus linguistics, several important factors should be taken into account. Firstly, the corpus does not reflect absolute statistics, but the model of the world displayed in texts – in other words, if a word is widely represented in the corpus of texts, this does not mean that a given phenomenon or object is as widespread in real life. However, this means that this word is widely represented in fiction and journalism, which indicates the importance of the corresponding concept in public consciousness.

Secondly, corpus linguistics is “late” in relation to the realities of real life: for the needs of the corpus, texts need digitization and markup [2]. To implement it, it takes time, and today, in the conditions of a constant extensive growth of texts, it turns out that the realities displayed by the corpus are 2-3 years away from the present. This, however, allows you seeing a fairly reliable picture of the word use, if it has long been included in the language. The corpus may incorrectly reflect the picture of the use of neologisms and occasionalisms, however, while, for example, the lack of a word in the corpus is also significant and implies its low significance for linguistic realities.

Thirdly, when studying the history and statistics of the words use, one should take into account the comprehensiveness of corpus linguistics in relation to stylistic registers and functional styles of the language [3]. If you do not specify search parameters that take into account only, for example, newspaper texts, by default, a search in the case shows all the texts digitized in it, including, for example, information from online forums

or private correspondence [9]. In such documents, the word may be spelled incorrectly or in the vernacular version.

3. Results and Discussion

Despite the factors mentioned above that limit the capabilities of the corpus, its data can be used for students to conduct small linguistic studies. Two versions of such studies are proposed, which do not imply a deep immersion in corpus linguistics, but at the same time allow students getting acquainted with the realities of both their native and the studied language.

The first version of such a study is a comparison of statistics on the use of a group of words united by meaning in order to find out which one is used more often and which one less often, and try to find an adequate explanation of the results. Thus, first-year students conducted a mini-study on the National corpus of the Russian language [8] in the following groups of words: “shoes” (words “sapog” (“high boot”), “botinok” (“boot”), “tuflya” (“shoe”)), “citrus fruits” (“limon” (“lemon”), “apel’sin” (“orange”), “mandarin” (“tangerine”), “greypfrut” (“grapefruit”)), “colors” (“krasnyi” (“red”), “siniy” (“blue”), “zelenyi” (“green”), “chernyi” (“black”), “belyi” (“white”), “zheltyi” (“yellow”)).

Students received the following results:

When analyzing the thematic group “shoes”, the student decided to look for the forms of the words “high boots”, “boots” and “shoes” in the plural, as these items are most often found in pairs. The study has revealed the following results:

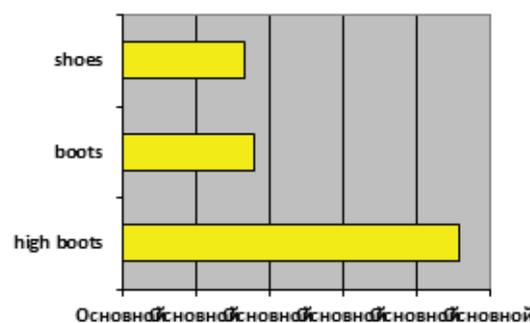


Figure 1: Thematic group “shoes” in Russian

3349 contexts in the word “shoes”, 3604 contexts in the word “boots” and 9181 contexts in the word “high boots” were determined. Students were asked to explain why, to their opinion, there is such a difference between the word “high boots” and the words “boots” and “shoes”. The following responses were received:

1) high boots, unlike boots and shoes, are part of the military uniform, and can be mentioned in the context of the description of the war, soldiers, etc.;

2) boots and shoes are gender-specific type of shoes: although both of them can be either male or female, nevertheless, by default, boots are more often called men boots, and shoes – women shoes. The high boots are unisex shoes;

3) high boots are the most ancient type of shoes, in comparison with boots and shoes, which leads to their wider mention in literature and scientific articles;

4) the word “high boot”, in contrast to the words “shoes” and “boot”, has a figurative meaning and is used in proverbs and comparisons (for example, “a boot” is a straight-forward and limited person, there is also a proverb “a bootmaker without high boots”, etc.).

When studying a group of words calling citrus fruits, students received the following results:

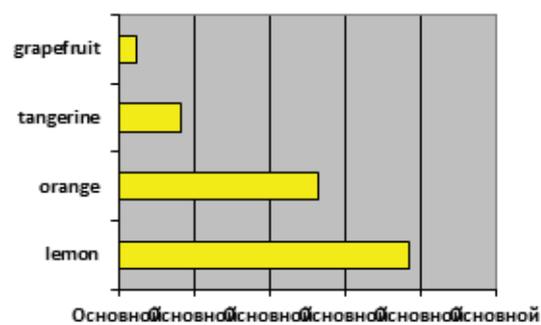


Figure 2: The names of citrus fruits in Russian

The corpus showed 770 contexts for the word “lemon”, 529 for the word “orange”, 165 for “tangerine”, 48 for “grapefruit”. Attempts to search for other fruit names — lime, pomelo, etc. — showed single occurrences, and it was decided not to consider. The discussion about the results has revealed the following versions:

lemon is often used in recipes: unlike other citrus fruits, it can be used with tea, its juice is included in many dishes;

lemons are more accessible than other types of fruits, and therefore they are more often mentioned in the corpus;

the word “lemon”, unlike other citrus fruit names, has figurative meanings: for example, in colloquial terms, “lemon” is called a million. Also, the word “lemon” is used for comparisons: “sour, like lemon”, “as if swallowed a lemon”, etc.

The color naming analysis shows the most diversification:

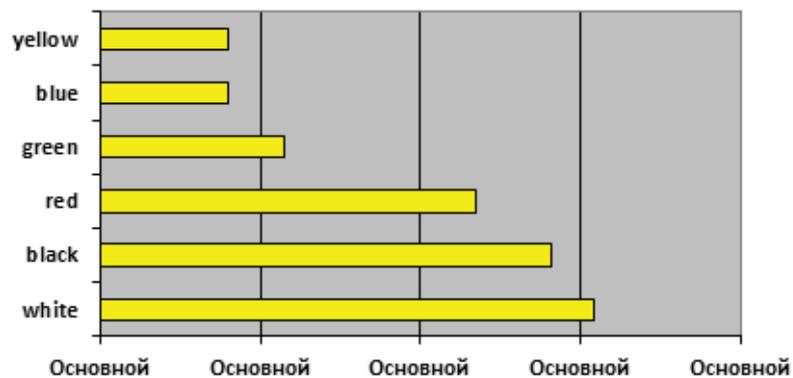


Figure 3: The color names in Russian

The word “white”, having the greatest number of occurrences, is reflected in the case 15 423 times, “black” – 14 078, “red” – 11748, “green” – 5 748, “blue” – 4 017, “yellow” - 3966.

Students were unanimous in discussing the statistics obtained: the difference in the number of uses of the “three leaders” – the words “white”, “black” and “red” – and other words is explained by a large number of figurative meanings and phraseological units with the indicated words.

If the teacher speaks a foreign language, students can be offered a comparison of the received data with the data of the foreign language corps. From the above results, it seems productive to make such a comparison on the last point, since the first two will not have the correct correspondence in English. The words “high boots” and “shoes” will correspond to the generalizing word “shoes”, and in the analysis of citrus fruits it will be difficult for students to separate the word “orange” as “fruit” from the color “orange”.

The following statistics were obtained from the names of the colors in English:

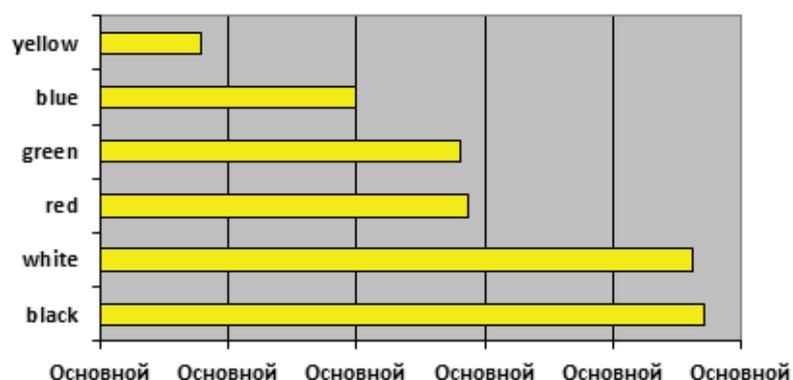


Figure 4: Color names in English

In absolute terms, the statistics are as follows: “black” – 23609, “white” – 23111, “red” – 14374, “green” – 14059, “blue” – 9941, “yellow” – 4 317. Students need to clarify that a larger number of entries in the British Corpus [4] does not mean that this word is used more often in English than in Russian: the fact is that the British Corpus is more extensive, and today it contains more contexts. The relative statistics are of some interest: in general terms, the frequency of use of certain words coincides with Russian (the word “black” is slightly ahead of the word “white”, and in the Russian corpus, on the contrary, the word “white” is slightly ahead of the word “black”, which caused students to laugh: it seems to indicate a more positive perception of the world by Russians, compared with native English speakers).

This option of mini-research allows students, on the one hand, to dive deeper into the study of direct and figurative meanings of words (basically, the difference between the number of occurrences of two words that are similar in meaning is explained by the presence / absence of figurative meanings), and on the other hand, to speculate about the history of words and their use in their native and foreign languages.

The second option of mini-research offered to students involves an appeal to a single meaningful word, the history of its use reflects various social and historical realities. As an example, we demonstrate the words “corruption”, “referendum” and “sport”:

When analyzing the results of such a mini-study, you should also pay attention to students that, by default, the corpus shows the results in the interval between 1800 and 2014 (if desired, the student can change these parameters). The word “corruption” is interesting because the first cases of its use are reflected in the corpus and a sharp rise is visible in the 2000s, due to a wide discussion of corruption in the media, mass publications with exposures, etc.

The word “referendum” demonstrates clear “peaks” of use associated with the referendum in the history of modern Russia.

The highest point of use occurs in 2014, when a referendum was held in Crimea, and it was often mentioned in the media.

It makes sense to demonstrate to students that not all words have a peak in use in the latest media and literature. To this end, the word “sport” was investigated, showing the following results:

The peak of the use of the word “sport” was in 1916, which, as students reasonably reasoned, was associated with the preparation for the 1916 Summer Olympics, and then their cancellation because of the First World War. This led to increased attention to the very concept of “sport”, which is confirmed by the presence in the corpus, such, for example, contexts for 1916:

This episode is not only curious in itself, but rather for characterizing the place that sport occupies in English life.

Yesterday, on the field of the Sokolniki Sports Club, the first performance of the Petrograd football club Sport took place against the Chukhlinko-Sheremetev club.

Comparison with the data of a foreign language seems difficult in this case, since students have little knowledge of the history of England and the United States and will not be able to quickly compare the data with historical facts. In addition, in the British and American National Corporuses [4, 5], due to the large volume of material, the historical corpus and the modern corpus are separated, which also does not allow us to demonstrate a comparison with the statistics shown by the Russian National Corpus.

4. Conclusion

Using the study of the native language and comparing it with foreign data from corpus linguistics in the classroom with students allows you to intensify the discussion of lexical meaning, semantization of a foreign word and the use of a comparative method of learning.

It seems more reasonable to use the following procedure for working with the above and similar statistics. Initially, students are invited to suggest what the relative distribution of the group of words will be in terms of frequency of use, and then demonstrate real statistics. It is recommended to do the same in the case of working with the corpora of the students themselves: first introduce the words themselves, listen to the audience versions, then show the data received.

The introduction of even small fragments of work with the body in junior undergraduate courses during the development of their native and foreign languages allows students to conduct independent studies, accessible for implementation even using a mobile phone, and in a new perspective, consider the words of both their native and foreign languages.

References

- [1] Biber, D., Conrad, S., Reppen, R. 1998. *Corpus linguistics: investigating language structure and use*. Cambridge University Press, Cambridge, 650 pp.
- [2] Biber, D., Reppen, R. 2015. *The Cambridge Handbook of English Corpus Linguistics* – Cambridge University Press, Cambridge, 398 pp.

- [3] Boulton, A. Corpora in language teaching and learning <https://www.cambridge.org/core/journals/language-teaching/article/corpora-in-language-teaching-and-learning/> [Accessed February 19 2020]
- [4] British National Corpus <https://www.english-corpora.org/bnc/> [Accessed February 10 2020]
Corpus of Historical American English <https://www.english-corpora.org/coha/> [Accessed February 15 2020]
Dormagen, H. 1977. *Theorie der Sprechfähigkeit*. Weinheim, Basel, 209 pp.
- [5] Kozlovskiy V.V., Braslavskiy R.G. Civilization mode of structuration of social inequalities in russian society https://www.isras.ru/files/File/ezhegodnik/2018/25_Ejegod_2018-25.06_Kozlovsky.pdf [Accessed 19 February 2020]
- [6] National Corpus of Russian Language <http://ruscorpora.ru/new/> [Accessed 16 February 2020]
- [7] O'Keefe, A., McCarthy, M. 2011. *The Routledge Handbook of Corpus Linguistics*. Routledge, London, 712 pp.
- [8] Saussure, F. de, 1995. *Cours de linguistique générale*, Payot, Paris, 455 pp.