A PROJECT WORK AS A WAY OF BRINGING CORPORAS TO SECONDARY SCHOOL

MARINA KOGAN1 – VICTOR ZAKHAROV2

1 Peter the Great St. Petersburg Polytechnic University
2 Saint Petersburg University


Abstract: Corpus linguistics is one of the most dynamic and rapidly developing areas of modern linguistics. It affects all areas of linguistics, including methodology of teaching foreign languages, translation and other linguistic disciplines. Corpus linguistics has had a direct impact on teaching foreign languages. However, in general, it remains a marginal method in teaching. Analysis of publications on the subject allows us to conclude that very few studies are long-term and aimed at working with schoolchildren. This article proposes a model for the development of sustainable interest among high school students in online corpora as sources of linguistic information, including the initiation stage in the form of project work in mini-groups to study well-known sayings with the consequent stage aiming at completing tasks supplementing the main textbook on a regular basis. The organization of project work addressing the corps of 11th grade students of the Natural Science Lyceum at Peter the Great St. Petersburg Polytechnic University is described. The paper outlines further research.

Key words: corpus linguistics, language pedagogy, longitudinal studies, method of projects/project work, proverbs, sayings

1. INTRODUCTION

Corpus linguistics (CL) is a relatively new direction in linguistics, which is engaged in the creation and use of corpora for solving various linguistic problems. Today’s corpora have become an integral part of linguistics and one of the methodological cornerstones used for research in vocabulary, grammar, discourse. Corpora are increasingly used as the basis for dictionaries, course books, and grammars (Chambers, 2019, p. 461). CL’s impact on English lexicography scholars compare with a “revolution” (Hanks, 2012). After the advent of corpora, all linguistic science including applied linguistics became different.

Its role in teaching foreign languages is also very important and since the early 1990s, enthusiasts have been talking about the revolution that CL approaches will bring to the field of teaching foreign languages. However, almost 30 years later, researchers are forced to admit that this revolution has not happened yet, with the only important exception of the “indirect application” of CL approaches in teaching through the use of
modern dictionaries and textbooks, which all are prepared on the basis of corpora. However, direct access to corpora by both foreign language (FL) teachers and students is so limited globally that Bolton and Cobb called it a marginal activity on the map of Language pedagogy, despite compelling evidence from a meta-analysis of 64 experimental papers on DDL efficiency and effectiveness in teaching English to different categories of students, with different levels of FL competence (Boulton – Cobb, 2017).

2. BACKGROUND. LITERATURE REVIEW

2.1 Problems hindering the widespread implementation of CL approaches in teaching foreign languages

Challenges hindering the widespread adoption of CL approaches in everyday teaching practice include:

1) insufficient ICT competence of FL teachers and insufficient training in the field of corpus linguodidactics;
2) an unusual format for presenting query results in the form of truncated concordance lines;
3) insufficient level of students’ FL competence to analyze authentic examples contained in the corpus;
4) lack of FL teachers’ readiness of preliminary preparing students to perform such tasks;
5) periodical changes of the resource interface and query syntax, which makes detailed manuals from previous years outdated (e.g. (Shaw, 2011);
6) lack of corpora created with students’ needs (Cobb, Boulton, 2015, p. 4) and pedagogical goals in mind (Braun, 2007);
7) gap between the results of corpus studies and their applicability in teaching foreign languages (Charles, 2007, Chambers, 2019);
8) adherence of language teachers to the priority of grammatical rules described in textbooks over language patterns fixed in corpora;
9) technical problems and problems associated with lack of time and financing (Chambers, 2019).

2.2 Analysis of longitudinal studies on using corpora by language learners

There are very few studies examining how deeply rooted the skill of using corpora is in students after the end of the experimental corpus course. To date, attention has focused primarily on student achievements and evaluation of corpus work immediately after a corpus course has been completed (Charles, 2014, p. 31).

Pérez-Paredes and Sánchez-Hernández (2018) found out that the use of corpora had limited or no impact on the writing practices of Spanish researchers two years after they had received in-service training on using corpora as a resource in writing scientific articles in their field.
The most consistent longitudinal research has been carried out by M. Charles. In one of her articles (Charles, 2014), she showed that out of 40 subjects 70% continued to use their created corpus a year later, of which 38% were active users who accessed the corpus every week and 32% were inactive users who accessed the corpus once a month on average. In her plenary talk at TaLC 2020 (Charles, 2020), M. Charles summarized the accumulated data for a longer research period from 2009 to 2017: she collected feedback from 221 participants one year after completing the course. During this period, the number of postgraduate students who has heard about corpora before the start of the course increased from 50 to 65%, but the number of those who use them has not changed (about 23%), with frequent users using corpora every week in their writing practice accounting for 10% of users. The results of the postponed survey one year after the end of the course show that the total number of users remains approximately stable at 62%, with the number of active users remaining at about 40%. These findings lead us to put forward the following hypothesis: the earlier students become familiar with corpora, the more chances there are that they will develop a sustainable interest in this type of linguistic resources and that they will be using them when necessary in their future studies of the foreign language. In the context of this hypothesis, two questions arise: what is the best students’ age and material/activity for introducing corpora to them?

2.3 Corpora as a resource for organizing students’ project activities

The documents regulating the goals and objectives of modern education in Russia pay great attention to the involvement of students in project activities. A special journal for school teachers of the foreign language, named “Inostrannye Jazyki v Shkole” [Foreign languages at school], regularly publishes articles on the theory and practice of using the project method in teaching foreign languages. However, as noted by Morozov – Urazayeva (2018), teachers very rarely turn to work with corpora although corpus linguistics provides a valuable tool for carrying out project activities in FL classes.

Here, one can mention the A. Boulton’s paper (2011) on MA students being familiarized with corpus linguistics through the project activities aiming at using corpora and CL approaches for conducting research in students’ fields of interest. This is another example of a long-term study, the results of which were presented at EuroCALL 2019 (Boulton, 2019) and TaLC 2020 conferences. In the reports from these gathering, it was noted that the students in question ‘gradually got involved in the game’ after overcoming initial difficulties. Obviously, the assignments requiring MA students to address the corpora described in this paper are not suitable for high school students. The direct transfer of corpus linguistic methodology onto FL instruction is not really appropriate in a secondary school context and it needs to be

1 Web-site of the journal Inostrannye Jazyki v Shkole: https://iyash.ru/
harmonized with theories of FL instruction, i.e. with didactics (Wicher, 2020, p. 32). This implies that the assigned corpus tasks should be designed in close relation to the main coursebook used by learners.

With the overall aim to have the students explore on-line corpora from as many different angles and in as many different ways as possible, we decided to organize the introductory phase of corpus work in a hands-on mode with the COCA\(^2\) and NOW\(^3\) corpora in the format of a project work.

### 2.4 The selection of the research problem

The research problem for 11\(^{\text{th}}\) grade students at the Natural Science Lyceum (NSL) in Petersburg should be designed in a way that by solving them, the students will become familiar with corpus search techniques, the analysis of the results, with the possibility of testing their own hypotheses etc. The tasks should be relevant to language level and interests of students, they should take into account their strengths, go in line with the subject (discipline) syllabus i.e. be compatible with the course book prescribed for teaching and learning English in NSL, as well as to allow the teacher to organize team work. We think that the task to conduct corpus search and analysis of proverbs and sayings meet all the above mentioned requirements. Such a task will allow students to ‘safely’ start an independent corpus search avoiding most difficulties and ‘dangers’ described in literature on the one hand and to get familiar with the most important features of modern big online linguistic corpora on the other.

Proverbs and sayings are included in most English textbooks, so high school students are familiar with at least the most common ones. Researchers and EFL teachers see a great potential in use of proverbs and sayings to activate grammatical constructions, enrich vocabulary, develop communicative competence, play a relaxation game, etc. (Pavlova, 2010).

The results of the experiment with the first-year linguistic students to study proverbs and idioms using corpus resources and tools described by I. Komarova and M. S. Kogan (2019) are very promising: A survey conducted at the end of the course showed that the first year students sincerely liked the hands-on corpus tasks and found this method of learning idioms effective. We believe that the analysis of proverbs and sayings and how they are used in the corpus is an interesting project assignment for high school students, which will learn them corpora skills. Besides, proverbs and sayings, like many other phraseological units, are fixed in the corpora not only in their standard form, but also in different modifications. Revealing this fact contributes to development of traits crucial for true researchers/scientists: the


\(^3\) Davies, M.: Corpus of News on the Web (NOW). Available online at: https://www.english-corpora.org/now (2016-).
ability to discover and interpret new unexpected facts. In the context of our hypothesis, we consider this task to be a good starting point for involving students into independent corpus work in the future.

3. THE ORGANIZATION OF AND TASKS FOR THE PROJECT WORK

The study was conducted at NSL in Autumn 2020. The participants were 16 students of the 11th grade (aged 16-17). They are very good at math, physics, and informatics and have diverse interests. They have different levels of command of English because they came from different schools. Their interest in English classes in NSL vary from very dedicated to rather superficial.

In NSL, Forward textbook is used at English classes as the main coursebook. Each Unit has a vocabulary section with a number of collocations and fixed expressions to learn. Some Units focus on idioms and present at least some proverbs and sayings, including To Err is Human..., which is used as the Heading of Unit 3 (Grade 10) (Verbitskaya, 2016). The number of exercises for training the proverbs and idioms is insufficient, which is typical of most English language textbooks (Cobb, 2019). Each unit includes a so-called Project ideas section. According to the syllabus there are two 45 minutes long classes a week in the autumn term.

The main part of the 3-week Project work was performed outside the classroom, with a 30 min introductory lecture and one 45 min class allocated for mini-teams reports about their findings at the end of the project. During the introductory presentation the teacher set the goals of the project and did the following:

- Gave a brief but carefully prepared talk about corpus linguistics;
- Introduced briefly COCA and NOW corpora;
- Specified the corpus-based task for project work for each mini-team;
- Supplemented students with detailed written instructions on conducting corpus and dictionary research and requirements for the final Report;
- Formed 6 mini-teams of two to three students to deal with a particular proverb/saying.

Then, each team worked independently outside the classroom for two weeks. Finally, each team presented their findings in front of the class answering their classmates and teacher’s questions.

Each mini-team was given one item from the following list of sayings’ parts for qualitative and quantitative analysis in the COCA and NOW corpora during the project work.

1. To err is human…
2. The grass is always greener…
3. Necessity is the mother of…
The students had to conduct quantitative research in the COCA in the mode *List* and both quantitative and qualitative research in the modes *Chart* and *KWIC*. From the *List* mode, they had to obtain the total frequency and number of occurrences of the proverbs/sayings in the canonic form and its modifications containing the given part. In the *Chart* mode, they had to visualize and comment the results by using the bar chart, with the aim of evaluating how the phrase has been used in different genres and time periods, of becoming familiar with a list of examples of the phrase (the given part of the saying) used in a particular genre (each member of the team could choose a genre independently). Finally, they had to make a conclusion about the canonic form of the saying and confirm the conclusion using Cambridge English Dictionary (https://dictionary.cambridge.org/).

In the *KWIC* mode with the Right end alignment function, the students had to study the proverb endings; calculate the number of instances of the standard/canonic and modified forms of the saying and provide examples of modified forms. In case of recurrent modifications, the students were asked to specify these modifications. Also, they had to provide examples of the saying endings which they found unusual, striking, or original. Another task drew students’ attention to a sentence type: they had to notice if the sentence containing the saying or part of it is affirmative, negative, interrogative, an exclamation or an unfinished thought (marked with a series of 3 dots at the end of each sentence.)

In the *KWIC* mode with the Left end alignment function, the students had to calculate the number of instances of the saying at the beginning of a sentence; as a part of a sentence (the first word starts with a small letter); as a quotation used in the inverted commas. Finally, students had to learn the frequency of the canonic form of their saying in the COCA corpus.

The students had then to become familiar with the NOW corpus. The NOW corpus-related tasks were included into the project work because

1. We wished to provide a first-hand experience of work in more than one corpus to the students;
2. In this way, the students could see the influence the corpus size had on their search results;
3. The NOW corpus enabled us to familiarize students with some functions that are difficult to access in COCA, for example, with arriving at the source text following a hyperlink, a randomized sampling function that allows one to get 100 or 200 examples of the saying use in random, rather than in reverse chronological order. The students had finally to submit a Report on the results of the research.
4. RESULTS AND DISCUSSION

The participants showed a sincere interest in the CL introduction lecture. They asked some technical questions on corpus markup which is due to their general high level of computer science and programming knowledge. At the end of the lecture, there was a lively discussion on the issue of how often native English speakers use proverbs in their everyday speech. The discussion allowed the teacher to emphasize that this is a problem which needs exploration and the corpus research can help answer this question.

At the final lesson, all the students noted that they spent a lot of time on the task of multidimensional analysis of the saying part in the two corpora, which on average amounted to 5-6 hours per group, or about 2 hours per person. There were no complaints about the difficulties associated with the analysis of the concordance lines, which is in line with our expectations, as in the COCA there were fewer than 100 examples of the use of 5 out of 6 sayings (see Table 1).

<table>
<thead>
<tr>
<th>Saying’s part</th>
<th>COCA</th>
<th>NOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>To err is human</td>
<td>87</td>
<td>737</td>
</tr>
<tr>
<td>The grass is always greener</td>
<td>79</td>
<td>359</td>
</tr>
<tr>
<td>Necessity is the mother of who laughs last</td>
<td>78</td>
<td>1043</td>
</tr>
<tr>
<td>Early to bed and early to rise</td>
<td>26</td>
<td>115</td>
</tr>
<tr>
<td>Birds of a feather</td>
<td>230</td>
<td>1678</td>
</tr>
</tbody>
</table>

Tab. 1. Search results for parts of proverbs/sayings in COCA and NOW

The selection of a randomized sample of 100 examples in the NOW corpus also allowed students to analyse them quite comfortably. The KWIC view mode with a given phrase in the centre, and the closest words coloured in different colours, with the total number of no more than 200 examples can be called a user-friendly interface, because it allows users to perform a convenient qualitative and quantitative analysis of the proverb use contexts both in the canonical form and in its various modifications.

The students conscientiously analysed the concordance lines and also described the trends in the use of their proverb over time, using the results obtained in the Chart view. They admitted the usefulness of the corpora for linguistic research, but complained that it was tiresome to analyse concordance lines. They noticed that although the modifications of sayings are multiple, the number of really unusual, striking examples of sayings’ endings is very small.

The students enjoyed the opportunity of changing different modes / formats of visualization in Chart view mode (frequency by year (Fig. 1), all subsections at once (Fig. 2), frequency by country (Fig. 3).
As an additional assignment, the students had to get familiarized with the extended usage context of their favourite modification of the saying searching in the COCA corpus. They were also asked to go to an external site in order to see the full text of the article searching in the NOW corpus. In the Reports, they noted that this was a good way to find out exactly how the unusual modification of the saying they liked is used.
While doing the project work, students faced some technical problems. When asked to elaborate what went wrong, they admitted that they often preferred a trial and error method of getting familiar with a new resource/software to following the instructions (some problems could have been avoided if the instructions had been followed). Another reported technical problem was loosing connection with the website when accessing it from mobile devices/smartphones. This might be attributed to the fact that the mobile version of the COCA website is less stable, because students who experienced it could not cope with it. Those students were unwilling to continue to use the corpus until the connection becomes more robust.

5. CONCLUSION

We believe that:

1. The selected material for the corpus analysis and the proposed format of work (mini-team project work) are effective for introducing high school students to modern online corpora.

2. This task contains elements of both true research and a computer game (the latter due to the unpredictability of answers and multiple searches).

3. We hope that our subjects have acquired basic skills of conducting the corpus research thanks to repeated use of the resource during the project work.

4. Some problems, typical for novice users of the COCA/NOW corpora, result from the complex nature of this resource.

5. Thus, despite some controversy in learners’ responses to online corpora, we think that the suggested way of familiarizing learners with this linguistic resource (by means of a project work performed in mini-teams with hands-on the corpora as the initiation stage) is the correct first step on developing sustainable interest in corpora in high school students.

6. Further research is required to clarify the following issues:

1) how to incorporate corpus-based tasks into the learning process, with the aim of in-depth studying language phenomena, that is, as naturally as possible and in accordance with the main textbook content;

2) how to provide a support during students’ individual and independent work with corpora while doing the assigned corpus-based tasks;

3) how to measure the impact of corpus-based work on students’ language and technical skills development;

4) how to evaluate students’ interest in and readiness/willingness to use this linguistic resource in the future.

Bibliography


CHAMBERS, Angela: Towards the corpus revolution? Bridging the research–practice gap. In: Language Teaching, 2019, Vol. 52, No 4, pp. 460–475. DOI:10.1017/S0261444819000089


SHAW, Erin Margaret: Teaching Vocabulary through Data-driven Learning. Brigham Young University 2011, 130 p. Available at: https://scholarsarchive.byu.edu/etd/3024/.
