



USING CORPORA IN THE LANGUAGE TEACHING AND LEARNING

Salomova Gavhar Abdinazarovna
Teacher, Uzbekistan

Umarova Rushana Rustamovna
Master's Degree of Pedagogical Institute of Termez State University

Annotation

This article describes corpus linguistics this method represents a digestive approach to deriving a set of abstract rules by which a natural language is governed or else relates to another language. Originally done by hand, corpora are now largely derived by an automated process.

Keywords: corpora, linguistics, pedagogic, method, approach.

Introduction

Corpus linguistics is the study of language as expressed in samples (corpora) of "real world" text. Corpus linguistics adherents believe that reliable language analysis best occurs on field-collected samples, in natural contexts and with minimal experimental interference. Within corpus linguistics there are divergent views as to the value of corpus annotation, from John Sinclair advocating minimal annotation and allowing texts to 'speak for themselves', to others, such as the Survey of English Usage team based in University College London advocating annotation as a path to greater linguistic understanding and rigor[7].

The principles of corpus linguistics have been around for almost a century. Lexicographers, or dictionary makers, have been collecting examples of language in use to help accurately define words since at least the late 19th century. Before computers, these examples of language were essentially collected on small slips of paper and organized in pigeon holes. The advent of computers led to the creation of what we consider to be modern-day corpora. The first computer-based corpus, the Brown corpus, was created in 1961 and comprised about 1 million words. Today, generalized corpora are hundreds of millions of words in size, and corpus linguistics is making outstanding contributions to the fields of second language research and teaching [6].

Relevance of The Topic

So what exactly is corpus linguistics? Corpus linguistics approaches the study of





language in use through corpora (singular: corpus). A corpus is a large, principled collection of naturally occurring examples of language stored electronically. In short, corpus linguistics serves to answer two fundamental research questions:

1. What particular patterns are associated with lexical or grammatical features?
2. How do these patterns differ within varieties and registers?

This is the idea that forms the backbone of corpus linguistics.

It's important to not only understand what corpus linguistics is, but also what corpus linguistics is not. Corpus linguistics is not:

- Able to provide negative evidence
- Able to explain why
- Able to provide all possible language at one time.

There is now a wide range of fully corpus-based reference works (such as dictionaries and grammars) available to learners and teachers, and a number of dedicated researchers and teachers have made concrete suggestions on how concordances and corpus-derived exercises could be used in the language teaching classroom, thus significantly “enriching the learning environment” [6. p-51]. Indicative of the popularity of pedagogical corpora use and the need for research in this area is the considerable number of books and edited collections some of which are the result of the successful “Teaching and Language Corpora” (TaLC) conference series that have recently been published on the topic of this article or which bear a close relationship to it[1].

In this part I wish to examine the relationship between corpus linguistics (CL) and language teaching (LT) and provide an overview of the most important pedagogical applications of corpora. As Figure.1 aims to illustrate, this relationship is a dynamic one in which the two fields greatly influence each other. While LT profits from the resources, methods, and insights provided by CL, it also provides important impulses that are taken up in corpus linguistic research. The requirements of LT hence have an impact on research projects in CL and on the development of suitable resources and tools.

The present part will investigate what influence CL has had on LT so far, and in what ways corpora have been used to improve pedagogical practice. It will also discuss further possible effects of CL on LT and of LT on CL, and highlight some future tasks for researchers and practitioners in the field[3].

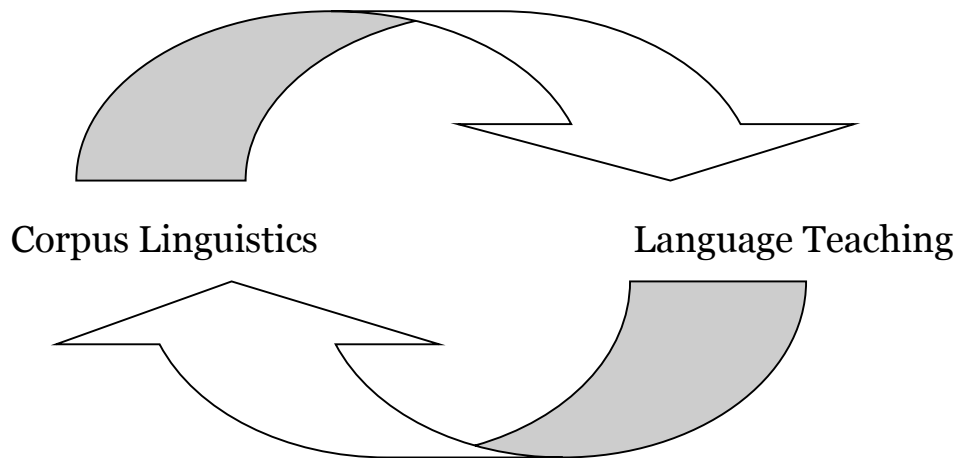
The relationship between corpus linguistics (CL) and language teaching (LT)

Since the mid 1990s, dictionaries based on corpora (collections of naturally occurring texts) have been widespread in English as a second language (ESL) classrooms. These dictionaries, based on large collections of natural language, not only provide learners





with information about word meanings, but also provide important information about word use resources, methods, insights needs-driven impulses



As a natural extension of using dictionaries based on corpora, teachers have become increasingly interested in using information from corpora to inform and create language-learning materials.

Corpora can provide a rich source of authentic material, and, therefore, examples of the language students will encounter outside the language classroom. Corpora can also provide students with many examples of the target feature (e.g. a vocabulary item or grammatical structure) in a concentrated manner, to help them better understand the feature and its contexts and contexts of use.

Before discussing some ways to bring corpora into a classroom, a word about corpora and corpus linguistics is in order.

The information from these studies, along with the rich descriptions of English provided in the Longman Grammar of Spoken and Written Language and the Cambridge Grammar of English, helps to provide a picture of the language that students will encounter, and therefore can be of great value in helping teachers plan what to teach and as a resource for developing teaching materials [12].

This chapter will discuss three ways for teachers to provide learners with hands-on corpus activities. First, teachers can bring in material from corpus searches and have students work with the teacher-prepared material. Secondly, teachers can use some of the-online corpora that are available. This section will focus on four available corpora that are very user-friendly (COCA, Time, MICASE and MICUSP). Thirdly, teachers can bring in existing corpora or create specialized corpora for their class (e.g. a corpus from readings or from student papers) and have students interact with the corpora.

These three ways are described in the sections that follow. Each has certain advantages and, of course, the ideas can be used in combination. For example, a



teacher might bring in some prepared concordance lines to introduce new vocabulary, and then later have students search an online corpus to see more examples of the words in context, in order to provide students with greater exposure to the different senses of the target word. This type of exposure to language can help learners get a better idea of the patterns of use and the words that co-occur with the new vocabulary.

Using Teacher-Prepared Corpus Material

In many places classes may not have easy access to computers, or may not be able to access the computer lab during class time. Teachers can still use corpus activities without having computers available for students. Instead of the students interacting with the corpus, the teacher will explore the corpus and bring the results into the classroom in the form of teacher-prepared material. For example, teachers can bring in word frequency lists or concordance lines that feature target vocabulary.

There are several advantages of teacher-prepared corpus material for learners. One major advantage is that teachers can control the material. Since the teachers search the corpus for the students, and then bring in those results, teachers can make sure that the vocabulary load is not too great, and that the students are exposed to the target form in a way that is meaningful and relevant for the students. This is a definite advantage in beginning courses where vocabulary load is an issue. In a lower level class the teacher might decide to delete the second-to-last and last lines of the examples of the verb concordance in Figure 1, since these contain difficult discipline-specific vocabulary (e.g. foot pounds, erg, /, N). Removing these lines does not impact the authenticity of the material; rather, it helps provide lower level students with meaningful and non-distracting input. Teacher-prepared concordance lines allow teachers to check that the content is appropriate for their learners. Prepared concordance material is also an ideal way to introduce students to reading concordance lines, something that can be distracting or confusing at first since sentences are often not complete[11].

As corpora and corpus tools become more available, and as teachers become better trained and more comfortable with using corpus resources, the ways in which corpora will be used for language learning will continue to expand. One aspect that will not change is the need to match learner goals and teaching resources and to use appropriate resources to accomplish teaching and learning goals. Corpora are one more tool toward that goal[12].

Corpora have a central role to play in our understanding of language. Over the last three decades we have seen corpus-based approaches take off in many areas of linguistics. They are valuable for language learning and teaching, as has been shown





in relation to the preparation of learners' dictionaries and teaching materials. Some language teachers have used them directly with students, but while there have been some successes, 'corpora in the classroom' have not taken off as corpora in other areas of linguistics have. Most attempts to use corpora in the classroom have been through showing learners concordances. The problem with this is that most concordances are too difficult for most language learners - they are scared off. However corpora can be used in the classroom in a number of other ways that are not based around (or do not look like) concordances. In this paper, after a little history, we present two of them[13].

A corpus is a collection of texts. We call it a corpus when we use it for linguistic or literary research. An approach to linguistics based on a corpus has blossomed since the advent of the computer, for three reasons:

A computer can be used for finding, counting and displaying all instances of a word (or phrase or other pattern). Before the computer, there were vast amounts of finding and counting to be done by hand before you had the data for the research question.

As more and more people do more and more of their writing on computers, texts have started to be available electronically, making corpus collection viable on a scale not previously imaginable. The costs of corpus creation have fallen dramatically.

Advanced concordancers allow a range of further features like looking for the search term in a certain context, or a certain text type, and allow sorting and sampling of concordance lines and statistical summaries of contexts. Also, tools like part-of-speech taggers have been developed for a number of languages[14]. If a corpus has been processed by these tools, we can make linguistic searches for, for example, kind as an adjective, or treat followed by noun phrase and then adverb (as in "she treated him well") Lexical competence recently has been identified to be the most significant predictor to general language ability ; however, it is also identified by most learners to be one of the biggest challenges of language learning. Fortunately, with the advent of technology, a new view of language learning and teaching has emerged; attempts to integrate computers as tools in language classrooms and facilitate the learning have been made[9].

This paper suggests language corpora can enhance the quality of vocabulary teaching and learning in second or foreign language classrooms. By presenting benefits of language corpora to the pedagogy, it is hoped that this paper can be helpful to both teachers and learners who are struggling to search for an efficient way of teaching and learning vocabulary[10].





What are Corpora?

Corpora, plural term of a 'corpus', refer to electronic authentic language databases that can be available via internet or as software installed in desktops. Language corpora can be either collections of written or spoken texts; for example, collections of written texts can be extract from newspapers, business letters, popular fictions, books, or magazines, published or unpublished school essays and etc. Collections of spoken texts can be any recorded formal or informal conversations, radio shows, weather broadcasts or even business meetings and etc.

What Can Corpora Do?

Language corpora can be used by anyone who is engaged in language learning, teaching, or research; language learners or even native speakers may find it useful to assist academic writing or lexical knowledge; teachers can utilize the authentic collections of data as classroom materials for ESL, EFL, or EAP (English for Academic Purpose) learners; language researchers or linguists often use corpora as language sources to analyse certain aspects of a certain language. Usually users of corpora use the searching tool, the concordance, to look for vast number of authentic language contexts analysed from corpora ; this feature provides users not only better quality of examples but more exposures to an unfamiliar word.

Examples of Corpora

Most software based corpora need to be purchased; however, there are still lots of free online resources, available from internet, for teachers and learners[8].

The web-concordance is one of the language projects presented by the Virtual Language Centre in Hong Kong. Users can search for language samples from various corpora such like students' academic writings, Time Magazines, the Bible, business and economy and etc[5].

In the years since COBUILD, all ELT dictionaries have come to be corpus-based. As they have vast global markets and can make large sums of money for their publishers, competition has been fierce. There has been a drive for better corpora, better tools, and a fuller understanding of how to use them. Textbook authors have also taken on lessons from corpora, and many textbook series are now 'corpus-based' or 'corpus-informed'.

Learners' dictionaries are designed to help learners understand and use words and phrases. A corpus is another resource to help with the same task. How do they relate to each other?





They are both records of the language. The corpus is a sample of the language in the raw. The dictionary is a highly condensed version of roughly the same material. The relation between the two is easy to see when we consider how modern corpus-based dictionaries are prepared[16]. One of the main inputs, at leading dictionary publishers including Collins, Macmillan and Oxford University Press, is word sketches: one-page corpus-based summaries of a word's grammatical and collocational behaviour. Is this more corpus-like or dictionary-like? It is automatically-produced output from the corpus, making it corpus-like, but it is a condensed summary of what was found there, making it dictionary-like. On a continuum from corpus to dictionary, it is somewhere in the middle[4].

Most learners do not want to be corpus linguists, and concordances are unfamiliar and difficult objects. But dictionaries are familiar from an early age, sometimes even loved. Learners will not be put off if they are expected to look items up in a new kind of dictionary. This suggests a strategy for bringing corpora into the classroom: disguise them as dictionaries.

Dictionary-users often find the examples are the most useful part of a dictionary entry. Moreover, where dictionaries are electronic rather than on paper, the traditional space limitation on examples disappears: there is room for lots of examples. This is an area where the corpus can help: they are nothing but examples. However they are not selected or edited examples[17]. Choosing examples for a dictionary is an advanced lexicographical skill: they should be short, use familiar words, without irrelevant grammatical complexity, and they should give a typical example of the word in use and provide a context which helps the learner understand what it means.

Conclusion

In conclusion while we cannot yet program computers to do the task anything like as well as people, we can perform some parts of it automatically. We can rule out sentences which are too long, or too short, or which contain obscure words, or which have many words capitalised or lots of numbers or square brackets or other characters which are rare in the kind of simple, straightforward sentences we are looking for.

The machinery has been embedded into the Sketch Engine, and concordance lines can now be sorted according to GDEX score, so the 'best' ones are the ones that the user sees as their search hits.





References

1. Biber, D., S. Conrad and R. Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press
2. Real Grammar. Harlow: Pearson Longman. Coxhead, A. 2000. 'A New Academic Word List'. *Tesol Quarterly*, 34(2): 213-38
3. Stig Johansson "Contrastive Analysis and Learner Language: A Corpus-Based Approach University of Oslo 2008
4. Abdinazarovna, S. G. (2020). The Reflection Features of Abbreviations and Acronyms of the English, Russian and Uzbek Languages. *Тил, Таълим, Таржима* Халқаро Журнали, 2(1).
5. Salomova, G. (2014). The Reflection Features of Abbreviations and Acronyms of the English Language. In *The Second International Conference On Development of Pedagogical Science in Eurasia* (Pp. 174-176).
6. Salomova, G. A., & Khurramov, R. S. (2021). Abbreviation in English, Uzbek, Russian Languages Different Structural Types. *Web of Scientist: International Scientific Research Journal*, 2(12), 646-654.
7. Саломова, Г. А., & Хуррамов, Р. С. (2021). Структура, Семантика И Информационный Потенциал Аббревиатур. *Таълим Ва Ривожланиш Таҳлили Онлайн Илмий Журнали*, 1(6), 179-186.
8. Саломова, Г. А., Болтаева, М., & Холмуминов, А. (2020). Употребления Форм Глагола И Случаев В Современном Английском Языке. *Интернаука*, (1-2), 23-24.
9. Саломова, Г. А., Дустбоева, М., Кулниезова, Ш., & Гулбаева, Г. (2018). Неличные Формы Английского Глагола. *Интернаука*, (42-2), 8-10.
10. Саломова, Г. А., Жумаева, Н. А., & Хамроева, Д. (2020). Теоретические Основы Обучения Лексике Английского В Системе Образования. In *Педагогика И Психология В Современном Мире: Теоретические И Практические Исследования* (Pp. 63-67).
11. Саломова, Г. А., Маматмуродова, Т., Усарова, Н., & Чориева, Д. (2018). *Педагогические Науки. Наука Хxi Века*, 15.
12. Саломова, Г. А., Маматмуродова, Т., Усарова, Н., & Чориева, Д. *Современные Методы Обучения Английскому Языку: Игротехника Как Метод Обучения Английскому Языку На Начальном Этапе*.
13. Саломова, Г. А., Надирова, Х., Ашурова, Д., & Худойбердиева, С. (2019). Информационные Технологии На Уроках Английского Языка. In *Молодой Исследователь: Вызовы И Перспективы* (Pp. 29-34).





14. Саломова, Г. А., Файзиева, Ф. А., Рузикулова, М. О., & Паянова, М. К. (2019). Использование Икт И Интернет-Ресурсов На Уроке Английского Языка. In Молодой Исследователь: Вызовы И Перспективы (Рр. 36-40).
15. Саломова, Г., Абдуллаева, М., Хафизова, И., & Ахмедова, М. (2019). Современные Технологии Обучения Английскому Языку В Средней Школе. Интернаука, (7-1), 74-76.
16. Саломова, Г., Жумаева, Н., & Хамраева, Д. (2018). Методические Принципы Современных Методик Обучения Английскому Языку. Наука Ххi Века, (4).
17. Саломова, Г., Турсунов, Х., Бобоева, М., & Пайгамова, М. (2019). Словообразования В Английском Языке. Интернаука, (44-1), 54-55.

