TEACHING ENGLISH BASED ON CORPUS TECHNOLOGIES

Shayusupova Nargiza Bakhtiyorovna

Branch of Astrakhan state technical university in Tashkent region ,
republic of Uzbekistan. Assistant teacher

Abstract

The article considers corpus linguistics technology from the linguodidactic perspective. It focuses on a set of corpora, which might be used in English as a Foreign Language (EFL) teaching: the British National Corpus, the Corpus of Contemporary American English, the Times Magazine Corpus, the Corpus of Historical American English. The paper presents a definition of the instrumental competence as a coherent cluster of knowledge, skills and abilities necessary for the students to use corpus linguistics technology in foreign language learning. Students are supposed to learn how to work with some basic corpus tools: LIST, KWIC, CHART, COMPARE, COLLOCATE. The author shares personal experience of how to use corpus linguistics technology in foreign language teaching, presents a set of corpus-based exercises and describes the procedure of the learners` work with  the basic corpus tools. Using corpus linguistics  technology  in  EFL teaching develops students` critical attitude to English language and facilitates deeper understanding of some linguistic phenomena. Corpus technology has recently presented great potential for plenty of linguistic studies, which fall into some groups: (a) studies on corpus design and methodology (Iruskieta, Diaz de Ilarraza, & Lersundi, 2015; Weisser, 2016), (b) quantitative studies of English lexical and grammatical features (Schmidtke & Kuperman, 2016; López-Couso, 2015), (c) corpus-based studies on linguistic changes (lexical, syntactical and stylistic) (Koplenig, 2018; Perek, 2018), (d) corpus-based discourse analysis (Chernyavskaya, 2017, 2018) . Some papers have a proper look at the tools of corpus linguistics, for example, concordancer or KWIC (key words in context) and how to use them to solve linguistic issues. For example, Seretan and Wehrli (2013, p. 158) present "an enhanced type of concordancer that integrates syntactic information on sentence structure as well as statistical information on word co-occurrence". As we can see above, most of papers are devoted to using corpus technology to solve linguistic issues. However, we would like to look at corpus linguistics from another angle and focus on linguodidactic potential of corpus technology and its tools. The advantages of corpus technology for language learning and second language

learning are widely presented in the literature (Ackerley, 2017; Kogan, 2016; Leńko-Szymańska, 2017; Pavlovskaya & Gorina, 2017; Rodríguez-Fuentes, 2015; Tarnaeva & Osipova, 2016). The researchers suggest to use corpus technology for the following purposes: a) as an empirical component of lectures, student assignments and projects; b) to determine the meaning of words and identify differences in usage between synonymous lexical items; c) to study lexical collocations; d) to focus on linguistic evidence that either supports or contradicts the prescriptive grammar rules (Derybina, 2012; Kokoreva, 2013; Kuzminykh & Khoroshilova, 2017; Sosnina, 2017; Sysoyev & Evstigneev, 2014). The authors see the methodological value of linguistic corpora for some reasons. Firstly, they provide information about the real state of language, its historical, geographical and social variation. Secondly, they illustrate speech registers and genre diversity. Thirdly, they expand understanding of the functions of some language units. Fourthly, they allow to identify the most frequent language phenomena

and expand the students` vocabulary (Tarnaeva & Osipova, 2016).

2.   Problem Statement

"Today's challenging economic situation means that it is no longer sufficient for a new graduate to have knowledge of an academic subject; increasingly it is necessary for students to gain those skills which will enhance their prospects of employment" (Fallows & Steven, 2000, p.75). Taking this fact into consideration universities "have established an initiative to ensure that each of its students engages with these skills and has embedded this within the academic curriculum for all disciplines" (ibid). The importance of humanitarian subjects for successful employment is in particular highlight. Krepkaia & Mamleeva (2015) conclude that the strong points of humanitarian subjects should be taken into

consideration as emotional intelligence, communicative and social competencies developed by them help students to integrate into their professional areas effectively. Almazova & Rubtsova emphasize that liberal arts education is "a key foundation for a personality's complete development and humanitarian culture, universal humanistic values, personal purport and moral guidelines." (Almazova & Rubtsova, 2016, p.18). In the effort to develop some skills that might be helpful for the specialists of the future we offer to use corpus technology in teaching one of the university disciplines – English as a Foreign Language (EFL). EFL is a discipline which is studied by students of all specializations. That is why we can speak about

developing universal employability skills, which are necessary for professionals in all spheres. As far as we are concerned, universal employability skills might include the following abilities: independent learning, the retrieval and handling of information (texts, databases, diagrams, schemes, charts), analytical and critical thinking, problem-solving, ability to draw a conclusion etc. In this article we have a look at corpus linguistics technology, the use of which might contribute to the formation of the universal employability skills.

3. Research Questions

- Which corpora should be used in EFL teaching?
- What competence should be formed to use corpus linguistics technology in EFL studying?
- What knowledge, skills and abilities must students acquire to use corpus technology?
- Which exercises might help students to acquire the appropriate competence?

4. Purpose of the Study

The article is written to prove linguodidactic potential of corpus technology in teaching EFL. In this survey we want to share our personal experience how to use corpus linguistics technology in EFL teaching for university students.

5. Research Methods

Analytic-synthetic methods are used in the search and analysis of literature on the topic; experimental research method is used to identify the linguodidactic potential of corpus technology; modelling method is applied in describing the instrumental competence.

6. Findings

6.1. Linguistic corpora in EFL teaching

In the process of the corpus-based teaching EFL we recommend to use open-access corpora, i.e. freely-available online, or requiring only registration. There is a group of linguistic corpora, created by Mark Davies, Professor of Brigham Young University. This group comprises the British National Corpus (BNC), the Corpus of Contemporary American English (COCA), the Time Magazine Corpus (TMC), the Corpus of Historical American English (COHA). The British National Corpus is a monolingual corpus representing British English of the late 20th century. The corpus includes a collection of samples of both written and spoken British English from a

variety of genres, sources and subject areas. The total number of the collection is about 100 million words (BNC, 2018). The Corpus of Contemporary American English is the largest and the only freely-available corpus of American English. It contains 220 thousand texts, 520 million words of the period from 1990 to 2017. The corpus illustrates a wide variety genres: SPOKEN, represented by recordings of more than 150 television and radio programs (All Things Considered (NPR), Newshour (PBS), Good Morning America (ABC), Today Show (NBC), 60 Minutes (CBS), Hannity and Colmes (Fox), Jerry Springer, etc); FICTION includes short stories, novels, and plays from literary magazines for adults and children, film scripts; POPULAR MAGAZINES (Time, Men's Health, Good Housekeeping, Cosmopolitan, Fortune, Christian Century, Sports Illustrated, etc.), NEWSPAPERS (USA Today, New York Times, the Atlanta Journal Constitution, San Francisco Chronicl etc.); ACADEMIC JOURNALS divided into classes according to the classification of the library of Congress (class B – philosophy, psychology, religion, D – world history, K – education, T – technology, etc.) (COCA, 2018).

The Time Magazine Corpus consists of more than 275,000 articles from the TIME magazine archive and contains more than 100 million words of texts written from 1923 to 2006 (TMC, 2018). The Corpus of Historical American English is a historical corpus of 400 million words, including 100,000 texts from 1810 to 2009. It gives the possibility to trace the changes of the language over decades. The corpus contains texts from the digital library Making of America (MoA), including 267 monographs and over 100,000 articles of the magazines of the 19th century on education, psychology, American history, sociology, religion, science and technology, it also includes texts of different genres from the archives of universal digital library Project Gutenberg, Project Gutenberg (1810-1930) and from the website. All the linguistic corpora, have a similar structure. So, having acquired the skills to work with one corpus, students will be able to use the others. Before using corpus linguistics technology in teaching EFL it is essential to form an instrumental competence.

6.2. Instrumental competence: a definition and a content

Instrumental competence is a coherent cluster of knowledge, skills and abilities necessary for students to use corpus linguistics technology in foreign language classes to solve linguistic problems.

Knowledge comprises:

▪   knowledge of the structure and content of the linguistic corpora of Brigham Young University

(BYU): the British National Corpus, the Corpus of Contemporary American English (COCA),

the Time Magazine Corpus (TMC), the Corpus of Historical American English (COHA);

▪   knowledge of the specific and distinctive features of the BYU linguistic corpora;

▪   knowledge of the possibilities of information search in linguistic corpora;

▪   knowledge of the possibilities of using corpus linguistics technology for the analysis of

grammatical and lexical-grammatical phenomena;

 Skills comprise:

▪   the skill to make a query in the search bar correctly;

▪   the skill to generate a concordance;

▪   the skill to use a "wide context" tool;

▪   the skill to generate frequency charts by genre and year;

▪   the skill to make a list of frequency collocates to a word, a phrase.

Abilities comprise:

▪   the ability to work effectively with the LIST tool, which allows to find the most frequent

collocates to the phrase and make lists of the context use of the desired unit;

▪   ability to work effectively with the CHART tool, which allows to make diagrams of the

frequent use of the studied grammatical or lexical phenomena by genres and years;

▪   ability to work effectively with the KWIC tool, which allows to generate concordance lists;

▪   ability to work effectively with the COMPARE tool, which makes it possible to compare the

frequency and stability of the use of two words in relation to the third;

▪   ability to work effectively with the COLLOCATE tool, which allows to find the most frequent

phrases with the studied word.

6.3.   A set of exercises aimed at the formation of the instrumental competence

Proceedings of International Educators Conference

Hosted online from Rome, Italy.
Date: 25ᵗʰ November, 2022
ISSN: XXXX-XXXX   **Website:** econferenceseries.com

6.3.1. The exercise on the formation of the skill to make a query in the search bar correctly

What phrase can NOT be found on the following query in the search bar?

[beat].[v*] * [nn*]

A.   Beat the Yankees

B.   Beat about the bush

C.   Beaten a bill

D.   Beaten dawn prices

 The correct answer is B.

The part of the query [beat].[v*] means that we are looking for any forms of a verb beat. The part of the query * [nn*] means that before a noun only one word is possible. So, the phrase Beat about the bush

can not be found on the query, because between a verb and a noun there are two words instead of one.

6.3.2. The exercise on the formation of the ability to work with the LIST tool and the skill to generate a concordance. Refer to the Corpus of Contemporary American English (COCA), sub-corpus Finance and money. Use the LIST tool and generate the concordances with the prepositions despite, in spite of, and conjunctions although, while, however, nevertheless. Comment on the syntactical features of the prepositions and conjunctions.

References
1. Rana Prajesh SJB. "Education and the Use of Technology". Republic Week. 34-44. 59

2. Motteram Gary. "Introduction." Innovations in Learning Technologies for English Language. 112-123 pages.

3. Healey Deborah. TESOL "Technology Standards Framework5. Virginia: Teachers of English to Speakers of Other Language, Inc., 2008. 234-244 Pages.

4. Ying Ren, Warschauer Mark, Lind Sonja and Jennewine Louise. Technology and English.

5. Language Teaching in Brazil. Letras & Letras. Uberlandia. 34-46 pages.