

## Main Features of Corpus Linguistics in Vocabulary Studies

*Davlatalieva Zarina Asqarali qizi*

*A Teacher of Namangan Institute of Engineering and Technology*

### ABSTRACT

*Is to study various features of corpus linguistics, to highlight the most comprehensive existing classifications of corpora, description of the corpus based technologies and to work out the ways of rendering them in the process of translation.*

**KEYWORDS:** *moral, birds, Simurg, ghazal, rubai, Hudhud, Alisher Navoi, philosophical, story, description, spiritual, praise, nature.*

Corpus linguistics is considered to be an approach to the study of language rather than a branch of linguistics, which focuses on the analysis of real samples of language use. Corpus linguistics was born with John Sinclair and the Cobuild project at the University of Birmingham (UK)<sup>1</sup>. From its emergence in the 1960s, the popularity of this approach has grown to the extent that it has an impact in language teaching. As a matter of fact, the influence of corpus research can be shyly felt in syllabus design, teaching materials (dictionaries and books) and classroom activities. With regard to classroom tasks, a very small number of teachers (mainly university teachers) opt for the use of corpora in their classrooms and most EFL students claim not to be acquainted with these tools. Thus, the explicit use of corpora is not as widespread for the presentation and practice of the vocabulary of a foreign language as other types of classroom tasks, such as, fill in the gaps exercises and matching tasks despite of its advantages<sup>2</sup>. After examining some of the most well-known corpus tools which can be used for vocabulary teaching and referring to how teachers may take advantage of these language databases in EFL settings, this paper presents four corpus-based tasks which show different ways of exploiting corpora for vocabulary teaching and learning. The ultimate goal of this paper is to discuss the advantages and disadvantages of corpus work in the acquisition and learning of English vocabulary as well as in the development of some of the key competences mentioned in the Common European Framework of Reference for Languages. As will be illustrated in this paper, an analysis of corpus-based data promotes the development of some basic skills for lifelong learning by encouraging students to be active learners in the classroom and to apply their critical thinking to the study of vocabulary. In this sense, corpora contribute to a shift in language teaching by giving special emphasis to two key aspects: autonomous learning and the study of real language use (which favours the ability to communicate messages in a foreign language and the skill to participate and succeed in any type of social encounter). In the same way, it may foster the introduction of new technologies in the classroom (digital competence), especially when students are asked to work directly with the corresponding computer software. The following sections will deal with all these issues more in detail. The revival of the interest in vocabulary teaching in the last 20 years has been ascribed to two factors: accessibility of computerized databases of words (language corpora), and development of word centred or lexical approach in vocabulary instruction. For many years learning lexis was secondary to grammar learning in language acquisition due to the assumption held by some

<sup>1</sup> DeCarrico, J. 2001. Vocabulary learning and teaching. In M. Celce-Murcia (ED.). Teaching English as a Second or Foreign Language (3rd ed). Boston, MA: Heinle & Heinle, pp.

<sup>2</sup> Carroll, J. B., Davies, P., and Richman, B. 1971. The American Heritage Word Frequency Book. Boston MA: Houghton Mifflin.

researchers that once students learnt grammatical structures, lexical items to fill the slots in syntactic frames could be learnt later according to the students' needs. In the era of auto lingualism, which emphasized learning grammatical and phonological structures, vocabulary learning was left to take care of itself<sup>3</sup>.

Corpus Linguistics is a methodology within the field of linguistics that has been developing rapidly since the year 1964 when the first computerized corpus, The Brown Corpus<sup>1</sup> Collecting large amounts of text in order to analyze linguistic phenomena was not a new concept when corpus linguistics arrived as a methodology. As Meyer points out in a recent article, early dictionaries were based on a large body of published works and millions of citation slips of naturally occurring language. Furthermore, concordance lines (i.e., a word displayed within a surrounding context) as a format for displaying , was completed. Corpus linguists are mainly interested in descriptive or functional interpretations of language and study linguistic phenomena through the empirical analysis of large computerized databases of language called corpora. A corpus is "a large and principled collection of natural texts", which is compiled so that it is representative of the language in general, a dialect, or other subset of the language. Corpora may contain language based on written texts, transcribed speech, or both. These texts are stored electronically, and then analyzed using computer software programs called concordance generators, concordancers, or, generically, concordancing software (every instance of a word in a text or collection of texts has been around for centuries, as Tribble and Jones explain: In its original sense a concordance is a reference book containing all the words used in a particular text or in the works of a particular author (except, usually, the very common grammatical words such as articles and prepositions), together with a list of contexts in which each word occurs.<sup>4</sup> Books like this have been in use since the Middle Ages, especially in Biblical Scholarship. The earliest known complete concordance of the Latin Bible was compiled by the Benedictine Hugo de San Charo in the thirteenth century. Hugo, it is said, was assisted by no fewer than 500 monks.

Corpus-based methods and activities can help. This brings us to the discussion of how corpora are having a direct effect on the language classroom. Essentially, there are two ways to directly engage second language learners in corpus work in the classroom:

- 1) they can be given direct access to a corpus and concordancing program on a computer; or
- 2) they can be given print-outs containing the raw data, or concordance output, from a corpus.

To start with, a corpus has been defined as 'a collection of naturally occurring language texts, chosen to characterize a state or variety of a language.'. Large and not so large computerised collections of texts and samples of texts put together according to systematic principles have found a permanent place in the field of linguistics for a variety of purposes. Corpus studies are successfully integrated into language research today in practically every discipline and the growth of corpus based research and analysis in practically every discipline in linguistics is immediately observable.

---

<sup>3</sup> Foster, P. 2001. Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*, M. Bygate, P. Skehan, and M. Swain (eds), 75– 93. Harlow: Longman.

<sup>4</sup> Johns, T. (1991). 'From printout to handout: grammar and vocabulary teaching in the context of data driven learning.' In: T. Johns & P. King (eds), *Classroom Concordancing*, Special issue of *ELR Journal* 4. University of Birmingham: Centre for English Language Studies, pp. 27-45.

**References**

1. Abuzahra, N., Farrah M., & Zalloum, S. (2016). Using corpus linguistics in language classroom from a constructivist point of view. *Arab World English Journal AWEJ Special Issue on CALL*, 3, 229-245
2. Alkhezzi, F., Al-Dousari, W. (2016). The impact of mobile learning on ESP learners' performance. *The Journal of Educators Online*, 13(2), 73-101.
3. Alqahtani, M., & Mohammad, H. (2015). Corpus linguistics impact on students' performance and satisfaction. *TOJET: The Turkish Online Journal of Educational Technology*, 14(4), 102-112.
4. Basal, A., Yilmaz, S., Tanriverdi, A., & Sari, L. (2016). Effectiveness of applications in vocabulary teaching. *Contemporary Educational Technology*, 7(1), 47-59.
5. Carroll, J. B., Davies, P., and Richman, B. 1971. *The American Heritage Word Frequency Book*. Boston MA: Houghton Mifflin.
6. Cui, G., & Wang, S. (2008). Adopting corpora in EFL teaching and learning. *The University of Southern Mississippi*, 1(1), 69-80.