www.wordlyknowledge.uz Volume: 1, Issue: 2, 2023

CORPUS LINGUISTICS

Jabborova Gulhayo Faxriddinovna

University of innovation and socio-economics

Department of pedagogy and psychology

ANNOTATION: The article is based on the skill of creating a corpus in modern linguistics and it has been described as having a distinct aspect. The emergence of Corpus Linguistics be, its functions in language development are analyzed.

KEY WORDS: linguistics, Electronic Library, corpus linguistics, text.

CONCLUSION

Corpus linguistics encompasses the compilation and analysis of collections of spoken and written texts as the source of evidence for describing the nature, structure, and use of languages. This work typically brings a quantitative dimension to the description of languages by including information on the probability with which linguistic items or processes occur in particular contexts. Corpora vary greatly in size and design but most are nowadays in electronic form with purpose-built computer software to support analysis. Present-day corpus linguistics has grown out of a long tradition of using texts as the empirical basis for linguistic description, studying all levels of language, including phonology, lexis, grammar, and discourse. Corpora are often annotated to show grammatical classes and functions. Software to analyze grammatical structures or to identify collocations by means of concordancing has revolutionized text analysis. In focusing on the company which particular words tend to keep and on the ways in which language users habitually express themselves, corpus linguistics has thrown new light on how languages vary systematically in different historical, regional, and sociolinguistic contexts, genres, and registers. Probabilistic descriptions of languages can complement other methodologies used by linguists and have implications for work in a number of fields in addition to linguistic description. These include natural language processing, language education, and cognitive linguistics.

Corpus linguistics and studies of language variation show a broad overlap in objectives. The primary aim of corpus linguistics is the accurate qualitative and quantitative description of language use, based on representative samples of authentic usage. An essential part of this description is information about the distribution and frequency of different forms and functions under different linguistic and extralinguistic conditions (i.e., variation). Thus the corpus-based Longman grammar of spoken and written English (Biber et al., 1999) incorporates quantitative comparisons between American and British English, and between the genres of academic writing, reportage, fiction, and conversation. Similarly, the central task for any study of language variation or change is to demonstrate some difference in language usage (whether in language forms, in the functions of those forms, or in the relative frequencies or distributions of those forms or functions) under different conditions of use (distinguished by such features as time, place, purpose, style, or attributes of the participants). Though different types of data can be used to establish the existence of language variation (including social dialect surveys, questionnaires eliciting particular forms or expressions of particular functions, or surveys of attitudes towards usage, which can be helpful in interpreting any differences observed), the most direct and most generally useful type of evidence comes from quantitative comparison of samples of authentic language usage representing each of the conditions of use to be contrasted. Hence much of the study of linguistic variation is equivalent to comparative corpus analysis. The overlap is almost total for study of historical language change, which quite routinely depends on description and quantification of language features in some

www.wordlyknowledge.uz Volume: 1, Issue: 2, 2023

reference version of a particular text, or set of texts, representing different historical periods of a language – again, equivalent to a corpus.

Despite this considerable mutual interest in facts of distribution, however, there are also some important differences in emphasis between these enterprises. Social dialectology is most centrally concerned with adequately representing a language community, through achieving a carefully balanced mixture of speakers from different social categories. Priority is placed on achieving controlled comparisons between these social groups; thus social dialect surveys generally attempt to elicit language of a desired type, such that all informants perform matched tasks (which, however, may or may not be equally natural for every informant). By contrast, most corpora attempt to exhaustively represent a particular language variety, through including a balanced mixture of many different text types. Each text type is, in turn, usually represented by samples from a large number of different sources, allowing robust generalizations to be made about specific genres as well as about the target variety as a whole. The main collection priority for this purpose is to include real or authentic language events. In practice, some of the more 'private' speech events may have to be specially arranged for collection purposes, but even in such cases, participants are as much as possible recorded in familiar social groups and settings, without constraints on how the event develops. This practice has the result that different social groups are represented in corpora in their most typical linguistic activities, giving the most 'natural' data possible. However, to the extent that different social groups have different interests, activities, and ranges of everyday linguistic experience, it is effectively impossible for corpus data to achieve exact matching of all social groups by task or situation. Accordingly, the use of corpora has, to date, been extremely productive for variation linked to variety, situation, or linguistic function, but less so for social variation.

This article reviews three areas at the interface between corpus linguistics and translation, namely, corpus-based approaches to (1) translation studies, (2) translator education, and (3) translation practice. With reference to (1), it surveys approaches, corpus typologies, and findings concerning translation strategies, norms/laws, and universals. Moving on to (2), it provides examples of possible uses of corpora (parallel, comparable, learner) in the translation classroom and points out that building a corpus may in itself be a valid exercise in text analysis and documentation. Finally, (3), it considers the role (still rather marginal, but growing) currently played by corpus tools in the translation professions.

Corpus linguistics is a relatively new discipline, and a fast-changing one. As computer resources, particularly web-based ones, develop, sophisticated corpus investigations come within the reach of the ordinary translator, language learner, or linguist. Our understanding of the ways that types of language might vary from one another, and our appreciation of the ways that words pattern in language, have been immeasurably improved by corpus studies. Even more significant, perhaps, is the development of new theories of language that take corpus research as their starting point.

Corpus linguistics has been around a long time in the form of concordances used by lexicologists, lexicographers, and language syllabus designers. Its recent revival is linked to implementation of the computer in building corpora and for convenient data access. This has brought about the expansion of existing databases and the creation of new ones. A primary goal of corpus studies is identification of the linguistic characteristics and patterns associated with language use (both typical and atypical) in different contexts, for example, genres, settings, and audiences. The focus is empirical analysis, whereas the interpretation can be either qualitative or quantitative. Also new about contemporary corpus linguistics is the addition of such approaches as discourse analysis, intercultural communication, and conversational analysis to the applied linguists' tools and techniques.

www.wordlyknowledge.uz Volume: 1, Issue: 2, 2023

REFERENCES:

1.Croft, W., & Cruse, D. A. (2004). Cognitive linguistics. Cambridge University Press.

2.Gries, S. T. (2016). Quantitative corpus linguistics with R: A practical introduction (2nd ed.). Routledge.

3.Hunston, S., & Francis, G. (2000). Pattern grammar: A corpus-driven approach to the lexical grammar of English. John Benjamins Publishing.

4.Stefanowitsch, A. (2019). Corpus linguistics: A guide to the methodology. Language Science Press. http://langsci-press.org/catalog/book/148

5.Tognini-Bonelli, E. (2001). Corpus linguistics at work. John Benjamins.