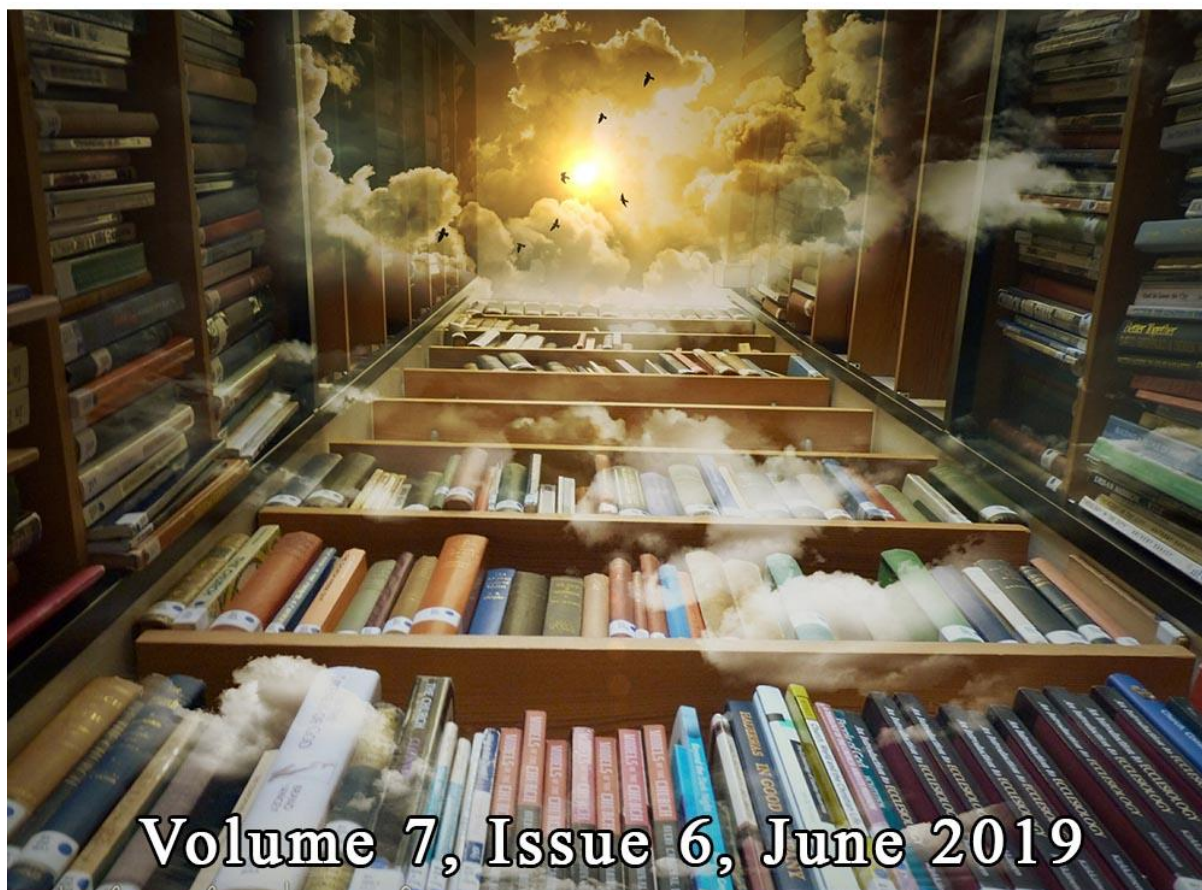


ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER

ISSN-2321-7065

IJELLH

**International Journal of English Language,
Literature in Humanities**



Volume 7, Issue 6, June 2019

www.ijellh.com

Amita Rawlley Thaman

Assist Prof, Dept of English

SCD Govt College

Civil Lines, Ludhiana, Punjab, India

amitarawlley@gmail.com

Viability of Corpora in Language Learning And Teaching

Abstract

The paper aims at capturing the importance and viability of corpora in language teaching. These large databases provide real-time exposure to the learners and researchers. In the contemporary era of knowledge deluge, corpora form an important means of quick and accurate information gain and immensely aid in teaching, learning and research work. They extend valuable insights to the teachers and help them in imparting effective teaching in their classrooms. In other words, corpora have ushered in a major revolution in as much as real-time and authentic knowledge gain is concerned. Large volumes of theoretical knowledge along with practical illustrations can be tapped from the corpora. Corpora analyse and lay bare the dynamics of a language. All nuances of language variation viz. dialect, accent, register and style are now freely available for access. Authentic study materials can be developed with the help of these. Corpora are the sweet gift of the modern world of digitisation and information boom. We are in fact already reaping benefits from them without knowing.

Keywords: Corpus, types of corpora, use in research, classroom pedagogy, reference publishing, syllabus design, materials developer, tools for teaching, testing.

Coming from the Latin word *body*, corpus refers to a large collection of texts. It is a body of written or spoken material which is digitally recorded in the form of typed, audio or visual forms. It often forms the basis of a qualitative or quantitative linguistic analysis. The plural of corpus is *corpora*. Natural or real world texts are known as corpora. A corpus is always principled and representative. The design matrix of a corpus ideally captures all the essential variables so that it is representative of a particular field vis-à-vis *age, gender, location, years of experience, native or non-native associations* etc. The corpus of a postgraduate English classroom in an urban area will be totally different from that of a similar classroom in a rural area. Similarly, the corpus of adult learners in a language classroom would be different from that of child learners. Corpus data are ideal for non-native speakers of the language.

The first systematically organised computer corpus was the Brown Corpus. A corpus may be small- consisting of 50000 words, or large enough to incorporate millions of words. Structured corpora of English as well as other languages are now readily available for access. Downloadable full text data is now available for COCA (Corpus of Contemporary American English), COHA (Corpus of Historical American English), GloWbe (Corpus of Global Web-Based English), NOW (Corpus of News on the Web), Wikipedia and many more. Corpora facilitate research with the highest degree of accuracy. These not only help the teachers in preparing effective lesson plans, but also enable the study of second language learning. Corpora provide valuable insights to all knowledge seekers. These prove to be precious resources for language teaching as they capture the essential flux of language development in specific time frames.

Basically, corpora provide real-time descriptions of language covering all aspects of language, be they *phonetic, lexical, morphosyntactic, semantic or pragmatic*. Thus they are of enormous use in compiling dictionaries and grammars. The lexicographers and

grammarians use the multimillion word corpora to examine word frequency, patterning and semantics. Major grammars of English are now corpus informed.

Types of Corpora

General-purpose Corpora do not belong to a specific field. They represent language in its broadest sense. These can be written (newspapers and magazines) or spoken (governmental proceedings, conventions and public meetings) eg. *The British National Corpus*

Specialised Corpora are designed with specific research goals in mind. They comprise of register-specific descriptions eg. *Bioscope Corpus*.

Pedagogic Corpora are adopted by the teacher for language teaching. They consist of various types of material for effective teaching.

Learner Corpora aim at representing language as produced by learners and include spoken and written language samples produced by non-native speakers. They help identify areas prone to committing errors.

Parallel Corpora are corpora aligned side by side for the purpose of comparison eg. *English-Norwegian Parallels Corpus*.

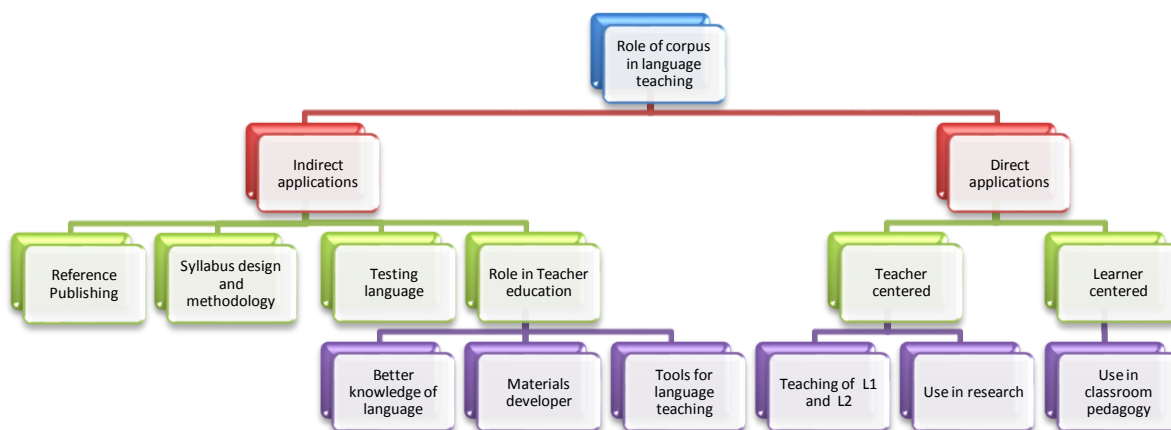
Comparable Corpora consist of similar texts in two or more languages/language varieties eg. *International Corpus of English (ICE)*.

Monitor Corpora grow regularly and are helpful in monitoring the language trends eg. *Corpus of Contemporary American English (COCA)*.

Historical or Diachronic Corpora look at changes across a lifetime and include texts from different periods of time eg. *Helsinki Corpus of English Texts*.

Role of Corpus in Language Teaching

The corpora play an extensive role in various areas connected to language teaching. There are a number of direct and indirect applications of corpora.



DIRECT APPLICATIONS

1. Teacher centered

Teaching of L1 and L2

The availability of corpora is of immense use to the teachers. As corpora offer real-time description and information of the changes language is going through, it comes in handy for the teacher in chalking out effective lesson plans for the students. The teacher gets clues on the learning process of the students as well as the common difficulties they come across while learning certain concepts. Many times structured and illustrative teaching plans are available for certain topics.

Use in Research

The corpora are extremely valuable for the teachers in pursuing their research work. Firstly, vast amount of informative data is available for ready reference. Needless to mention here

that the corpora being digital, are easily updated after regular intervals. Thus they provide the latest information to the teachers which is imperative for their research work.

Learner Centered

Use in Classroom Pedagogy

In the contemporary era of globalisation and digitisation, vast amount of knowledge is available for the learners willing to learn. The learners can access the corpora and can endeavour to improve their learning experience. Actual lectures by the authors of books are available. There are corpora of learner interviews, words and sentence patterns. The learners can sift and choose information as per their requirements. *Longman Learners' Corpus (LLC)* is a 10 million word corpus containing samples of written English produced by students from over 70 different countries.

Indirect Applications

Reference Publishing

Corpus data is widely used in reference publishing for a couple of decades now. Volumes of machine readable authentic data are available through corpora for grammarians and lexicographers. This data not only captures the contemporary flux in language but also provides the frequency of words and sentence patterns used by different people speaking the same language in different parts of the world. Apart from this, the exact usage of materials helps in accurately placing the denotations and connotations. Subtle changes in meaning and usage of a lexical item can be tracked through monitor corpus. *Collins Cobuild English Dictionary* published in 1987 was fully corpus-based. By the end of 20th century, new editions of major dictionaries were partially corpus informed. Such dictionaries were a more enriching experience for the learners as they carried valuable tips for them based on an

anticipation of possible errors. Similarly corpus informed grammars area fully updated versions of language use. The *Longman Grammar of Spoken and Written English* is entirely corpus-based. The newest combinations of words in language use can be depicted in illustrations and are extremely helpful for the learners since these are based on actual usage. *Cambridge Grammar of English* by Carter and McCarthy (2006) relies on the *Cambridge English Corpus* and the *Cambridge Learner Corpus* with the unique feature of error warnings.

Syllabus Design and Methodology

The design of language teaching syllabus is largely supported by the existing corpora. The items used by speakers native or non-native for real-life communication form the basis such syllabi. Commonly occurring lexical items, word patterns and sentence structures can be used for carving out an up-to-date draft of syllabus which could help the subsequent language learners in overcoming common communication hiccups.

Testing Language

Testing language is an important aspect of teaching and learning process. Valid and reliable test items can be successfully created through the corpora of learners and teachers. Corpora informed construction of tasks and test items reflect authenticity in the domain of target language. Levels of language proficiency can be identified by analysing the learner and expert language user corpora. Plausible distractions for multiple choice questions can be mulled over with the help of available corpora. The domain of language testing can thus draw heavily from corpus linguistics.

Role in Teacher Education

(a) Better knowledge of language

Recent developments in the field of corpora are of immense advantage for those in the teaching fraternity. Vast descriptions of language are now available for reference. Spoken texts of native and non-native speakers lend several insights to the teacher to help understand the stance of language users.

(b) Materials developer

Teachers can now develop endless material drawing from the available corpora. Thus they can plan their individual lessons in order to teach effectively. But for this they should undergo rigorous practice/training on how to use corpora for teaching. Suitable corpus-based tasks and exercises can be developed to impart real-time contextual exposure to learners.

(c) Tools for language teaching

Information from the language corpora greatly help in improving pedagogical practices in the classroom. They provide insights into possible uses of words and phrases in variable contexts. This helps makes teaching in closer correspondence to real-time use of target language. A study of learner corpora gives us valuable information about the difficulties faced by learners and the common errors committed by them. *Concordancing* is a core tool in corpus linguistics to find occurrence of words and phrases at the click of a finger. However, a fully corpus-driven approach to teaching is too challenging and also entails a lot of time-input. Selection, description and sequencing of corpus-informed data can be presented in reference tools i.e. monolingual learners' dictionaries and pedagogical grammars.

Conclusion

Corpora are treasure troves of latest language descriptions and can be immensely valuable for all concerned with teaching, learning and research of language. They are the sweet gift of the modern world of digitisation and information boom. In other words, corpora have ushered in a major revolution in as much as real-time or authentic knowledge gain is concerned. Large volumes of theoretical knowledge along with practical illustrations can be tapped from the corpora. But they have to be handled with care in order to reap full benefit. Corpus-informed study or teaching does not come easy and entails a lot of structured planning and careful selection of materials for learning or presentation. Handling the variability of language amidst the vast amount of data is a Herculean task.

References

- Aijmer, Karin. ed. *Corpora and Language Teaching*. 1984. pp.1-10.
- Casas-Pedrosa, Antonio-Vincente & Fernandez-Dominguez, Jesus & Alcaraz-Sintes, Alejandro. "Introduction: the use of corpora for language teaching and learning". *Research in Corpus Linguistics*. 2013. pp.1-15.
- Cushing, Sara T. "Corpus linguistics in language testing research". Sage Journals. Editorial. 2017.
- Granger, Syliane. "The contribution of learner corpora to reference and instructional materials design". *The Cambridge Handbook of Learner Corpus Research*. 2015. pp.486-510.
- <https://britishcouncil.org/voices-magazine/corpora-english-language-teaching>
- <https://journals.openedition.org/apliut/5433>
- https://lexically.net/wordsmith/corpus_linguistics_links/Roemer
- <https://www.sketchengine.eu/user-guide/user/manual/corpora/by-language/english-text-corpora/>
- <https://www.thoughtco.com/what-is-corpus-language-1689806>
- McCarthy, Michael & Evison, Jane. "Using Corpora in Language Teaching". Centre for Advanced Language Proficiency Education and Research. CALPER Digest 2004.
- Saeed, Naveen Al & Waly, Salma. "Corpus for Classrooms: Ideas for Material Design". *Proceedings of the 10th METU ELT Convention*.
- Samburskiy, Denis. "Corpus-informed Pedagogical Grammar of English: Pros and Cons". *Procedia – Social and Behavioral Sciences*. Vol 154. 28 Oct.2014.263-267.
- Sinclair, John McH. ed. *How to Use Corpora in Language Teaching*. 2004. pp.308
- Xiao, Richard. "How can corpora help in language pedagogy". *The Postgraduate Conference in Applied Linguistics, 2009 – lancaster.ac.uk*