Corpus-Based Studies – Some Perspectives

P. Aswini
Research Scholar, School of Social Sciences and Languages, VIT University Vellore, Tamil Nadu, India.

Dr. R. Srinivasan
Associate Professor, Department of English School of Social Sciences and Languages, VIT University Vellore, Tamil Nadu, India.

Abstract
Corpus Linguistics is one of the fastest emerging language teaching methods in modern applied linguistics. Several studies have been carried out to study the effectiveness of using Corpus Linguistics to comprehend the dynamics of language learning and to harness the benefits of learning a language through corpus based approach. This paper examines some of the significant studies on Corpus Linguistics carried out since the 1990s, and it attempts to outline the theoretical and experimental findings to help scholars, teachers and students using corpora.

Keywords: Corpus Linguistics, Corpora, Corpus-based studies, English Language Teaching.

Introduction
Corpus Linguistics is the study of language using naturally occurring language samples. It employs specialized software programmes on a computer to analyze various aspects of language. Hence, it helps obtain and analyse data quantitatively and qualitatively rather than relying on mere theoretical studies of language.

For a long time, collections of text in language have been compiled and studied. In the middle ages, scholars began making lists of all the words for a particular text with its immediate instances in all contexts—what today is called concordancing. In the pre-computer age, only works of special importance, such as the Vedas, Qur’an, Bible or the works of Shakespeare or classical Latin and Greek authors had concordances. Later, scholars counted the frequency of lexical items from single as well as collections of texts and produced lists of the most frequent words used in a particular text. Corpora were used in the areas of language acquisition, syntax, semantics, and comparative linguistics among others. Though the term Corpus Linguistics was not formally used, much of the work carried out was similar to corpus based research without the assistance of computer.

Modern Corpus Linguistics relies heavily on computer and the term corpus linguistics has almost become synonymous with the Computer Corpus Linguistics. The first modern computer readable corpus was the Brown Corpus of Standard American English which was developed in 1961. Following that, the London-Lund Corpus of spoken British English, British National Corpus (BNC), Bank of English (BoE), etc. were established. Today, Corpus Linguistics has become a separate branch of linguistics that is on par with other braches such as Socio linguistics, Psycho Linguistics and Neuro Linguistics.

In recent years, research on corpus linguistics has gained attention among applied linguists. Many scholars such as Biber, Johanson, Leech, McCarthy, Hunston, Francis and Conrad have contributed to the development of computerised Corpus Linguistics. Among them, John Sinclair is the most influential and prominent scholar of modern-day Corpus Linguistics. In paper [1] the author observes, “the corpus based approach is evolving, through theoretical elaboration and empirical realization, into a coherent, composite and rich paradigm that addresses a variety of issues pertaining to theory, description, and the practice of translation”. While corpus is profitable for language learning, some studies have attempted to use corpus driven materials to assess its impact in the classroom for second and foreign language teaching and learning. This paper reviews the existing empirical as well as theoretical corpus based studies to provide an insight on the subject to anyone who is willing to pursue this line of research further.

Theoretical and Empirical studies on Corpus-based approach
2010 – 2015: The most prolific period in Experimental Corpus Linguistics

In paper [2] the researcher demonstrated how to incorporate two in-house specialised corpora from the official website of the Liquor Control Board of Ontario into English for Specific Purposes (ESP) course for non-native speakers of English at University level. The authors attempted to assist learners to understand and explore how language features are used in their subject areas. This course, containing specialized vocabulary, was offered to first-year university students by the Applied English Department. The course aimed at acquainting the students with the linguistic features of the wine industry, with a focus on red wine and white wine. Lists of vocabulary were prepared based on corpora analysis, and as supplementary materials, in-house corpora were introduced to the learners. The results of the pre and post-test illustrated that students showed significant improvement in the acquisition of content and language. The authors opined that this study would be fruitful for future research to investigate the same linguistic features in the same area by using different wine review articles from newspapers and magazines.
To ascertain the effectiveness of corpus-based approach in teaching collocations of synonymous pairs, [3] conducted a study comparing the effectiveness of corpus-based approach with the effects of the traditional approach. Participants with equal competence on learning collocations were selected and were then divided into two groups, namely experimental and controlled groups. The materials contained different types of synonyms and their collocations along with the collocation exercises. The experimental group was taught using concordancing materials of BNC (British National Corpus) as the concordacing software. The control group, on the other hand, was taught collocations in a traditional way. The data were collected through pre-test and post-test writing. Findings revealed that a difference between the two approaches and the corpus-based approach had an effect on the comprehension and production of collocations of synonyms. The experiment concluded by insisting on the use of collocations in order to bring fluency in L2 writing among L2 learners. The reasons for this are based on the evidence that even advanced students have problems in using collocations.

To investigate the effectiveness of corpus-informed, contextualized and de-contextualized vocabulary instruction, [4] conducted a quasi-experimental study. In this study one of the experimental groups was exposed to corpus-informed vocabulary instruction and this group was informed explicitly about the rationale behind corpora and the use of concordances. In contrast, the other group was treated with decontextualised vocabulary instruction. The control group was instructed the target words in a meaningful and coherent context. After the eight week experimentation process, the results showed that the group treated through decontextualized learning activities scored statistically higher mean compared to the other groups. The group treated with corpus – informed instruction showed some progress, though statistically insignificant. The control group which was instructed through contextualized vocabulary activities, made the least progress. The authors conclude stating that corpus-informed vocabulary instruction will demand a certain level of technical skills. It can be achieved by a considerable level of training to the instructors as well as to the learners. However, it is necessary to consider certain factors like the learners proficiency levels, needs, ages, and the cultural background of the learners to use corpus in the classroom.

In paper [5] the author conducted a quasi-experimental research to explore the feasibility of a corpus by using the Corpus of Contemporary American English (COCA) to distinguish between close synonyms which have parallel meanings but cannot be replaced one for the other. Students were given 40-minute training to use COCA and were expected to make their judgement of choosing the appropriate English words from a collection of real language samples. The results of this study showed that the use of corpus was potential for determining word choice. From this study, [5] observed that corpora can be an effective tool for language learning as well as for the identification of the mistakes that students most often commit. Further, the researcher stated that it would not be advisable to use corpus to solve grammatical, lexical, and discourse related problems.

2000-2010: Emergence of Corpus-based studies in Corpus Linguistics

In paper [6] the authors through their study attempted to show the prominence gained by paper-based Data Driven Learning (DDL). They explored the concordance lines to discover frequent grammatical features and word derivations and to make generalizations from the samples to the language in general, even without the help of computers. The researchers built a series of computer-based and paper-based Data Driven Learning exercises which focused on the basics of grammar. Moreover, a four-step inductive and deductive approach was used. In the first part, the authors posited essential consideration in choosing the apt corpus for classroom DDL exercises with reference to its relevance, accessibility, and size. In the second part, they provided examples from computer-based and paper-based vocabulary and grammar exercises. The third part described the exercises that were used in the Nihon University Study. The final part highlighted the advantages and disadvantages of computer-based and paper-based corpus exercises. The final evaluation of their course displayed statistically positive results in most areas, and the students too gave a positive feedback about this course.

In paper [7] the author demonstrated the ongoing building of a speech corpus to be utilized by applied linguists, EFL, and ESL students. Their exposition was divided into three sections. In the first section, they substantiated the need for developments in the teaching of listening and pronunciation to EFL and ESL students. Further, it illustrated the need to exploit authentic language in teaching/learning process with the objective of advancing social inclusion. The second part established the characteristics of native speech flow which would cause difficulties for EFL/ESL students. The researchers provided guidelines to access the Dublin Institute of Technology’s slow-down technology to help students become aware of the features of native speech without tonal distortion. The final part illustrated a new speech corpus that was being developed at DIT and provided detailed information on the special features of this novel Speech Corpus which would meet the requirements of the teachers and learners and which was lacking in the traditional corpora.

In paper [8] the author shows the potential of corpus techniques with beginning-level students in English as a foreign language learning context by using a Japanese-English parallel corpus through his research project. The project investigated the following areas that were the responses of the students to concordance-based teaching activities, the learning outcomes of concordance-based teaching activities in various learning contexts, and the optimum features of a computer interface between the corpus and the users. By using appropriate vocabulary teaching material based on CALL, students were allowed to compare the lexical items and language patterns between two languages in order to gain new vocabulary as well as to comprehend patterns and usage in both languages. The evaluation of the learning effect was discerned through open-ended questionnaire based on learners’ impressions. The results revealed that this system was meaningful and useful. The author concludes by insisting to focus on reference skills which are essential for students in
the IT era to help them acquire necessary information about foreign languages even after finishing their course work. The author [9] in his article addressed what corpus linguistics is and how it can be applied in the field of teaching. The author has provided information on the advantages of corpus-based analyses to discover language patterns, and has investigated the usage of tag questions by male and female members and has also analysed the acquisition of irregular past participles by children. The author has pointed out some of the difficulties involved in executing Corpus Linguistics for the purpose of teaching. In addition, he has included a list of corpus links to support anyone who is willing to explore corpus linguistics.

1990’s – Setting the tone for Corpus-based studies
In paper [1] the author gives a collection of retrospective studies on corpus-based studies. The aim of this article was twofold. The first one attempted to outline a new field of research in translation studies. It contained the discussions regarding theoretical issues pertaining to the object of study, scope, and methodology of the corpus based approach. The second part, on the other hand, laid out that corpus-based approach was slowly developing by means of theoretical elaboration and empirical realisation of translation and translating. Moreover, it illustrated the problematic issues involved in the theory, description and practice of translation. Author [10] insights on the empirical studies and the author’s discussion on the important role of computerised corpora within the discipline as a whole are the major highlights of this work.

Conclusion
From these studies, it is obvious that corpus linguistics is a wonderful resource that can be exploited in the language classroom because it is highly relevant in this digital age. The articles cited above illustrate how research on Corpus Linguistics has evolved significantly since the late 1990s to emerge as one of the most fascinating methodologies for language analysis and language learning. While corpus has proved beneficial on several areas, there are a few challenges that users need to keep in mind before employing corpora in the classroom. The ability to use computers with ease and the skill to exploit the nuances of online corpora are desirable prerequisites for reaping the full benefits of Corpus Linguistics. In classroom contexts, students can get hands-on experience from the support and guidance of the teacher. They should then be able to employ the benefits so gained into real-life language learning and use. Materials derived from Corpus Linguistics are of great help in course design and content development for preparing teaching materials. Hence, a further scrutiny of the research on corpus based materials would be beneficial for language learners to explore language use in the real context. The above studies indicate that corpus-based approach is an appropriate and engaging means of focusing on a real, exploratory method to teach or learn a language in this age of information technology.

References