# CORPUS LINGUISTICS – RELEVANCE TO ENGLISH LANGUAGE TEACHING (ELT)

**Moin Uddin**
Ph.D. Research scholar
Department of English, Osmania University,
Hyderabad,Telangana ,India.

## Abstract

*In this paper an attempt has been made to demonstrate how the methods of corpus linguistics can profitably used for an effective teaching of English as a second language in the Indian context. Language studies fall into two broad areas, namely studies of structure and studies of use (pragmatics). A structural analysis of the English language will describe the grammatical similarities and differences across sentences. But an analysis of use goes beyond such a structural description and explains why a language should have multiple structures to convey the same information. Studies in corpus linguistics can help teachers, syllabus designers and textbook writers of English as a second language in India, to investigate preferences of the structure in different varieties, registers and styles, as well as in spoken and written modes of communication, and collocational behaviour of words, phrases, clauses and large structures like discourse or text. In this paper some of these issues are examined and illustrated in view of teaching/learning of English as SL at the undergraduate level in Indian universities.*

*Keywords: Corpus, Structural Analysis, ELT, SL, Pragmatics, Collocation*

# CORPUS LINGUISTICS – RELEVANCE TO ENGLISH LANGUAGE TEACHING (ELT)

-   **Moin Uddin**

**Introduction**

With the decline of the Audio-lingual Method and the Structural Approach to the teaching-learning of L1 and L2 like English, after 1960s, Communicative Language Teaching (CLT) became popular and wide-spread in the field of English Language Teaching (ELT). This was so because Dell Hymes challenged the Chomskyan notion of Linguistic Competence (1957, 1965) and proved, by providing supportive data, that Linguistic Competence is limited in scope, based only on the grammatical knowledge of the native speaker. To fill the gap, Hymes (1972) proposed the notion of Communicative Competence which according to Comale and Swain (1980) consists of grammatical (or linguistic) competence, sociolinguistic competence, discourse competence and strategic competence. In this way, the importance of contextual meaning of linguist structures has been highlighted in the syllabus design and preparation of teaching material and production of textbooks. Thus, CLT has become the most recommended approach in ELT.

At about the same period, the study of language use in context become the theme of Discourse Analysis and the production and interpretation of meaning of language in use has been studied in Pragmatics. If Sinclair and Colthard's contribution (1975) revealed the benefits of analysis of classroom discourse, Widowson (1973) argued convincingly that contextual meaning (which he called coherence or rhetorical value) is as important as the study of structure in language studies.

The philosophers of language like J.L. Austin (1962) Searle (1969), Grice (1975) developed methods of language use under the rubric of Pragmatics. There are also other significant studies in this area by scholars like Leech (1974), Levinson (1983), Sperber and Wilson (1986) and a host of others.

The point to be noted here is that while all the studies in Discourse Analysis and Pragmatics are admirable as theoretical contributions, they lack the support adequate data. In a few cases the examples of data given in these studies tended to be artificial and contrived (Simpson, 1997:151). For this reason, the approaches of CLT in ELT have not been able to

provide much naturalistic, data and satisfied themselves with the illustration of a few real situations like Teacher-student interaction, Doctor-patient interviews, At the airport, At the restaurant etc. This has been a serious limitation in the earlier methods and approaches also in relation to ELT.

The major argument of the present paper is that such a limitation of natural and authentic data can be overcome by practitioners of ELT, if they take the help of Corpus Linguistics. This is explained in the sections below.

### 1. Corpus Linguistics *

Corpus linguistics deals with the analysis of large amounts of language data and many contextual factors at the same time. Thus, corpus analysis is purely quantitative and it facilitates the study of language in use. The investigation in corpus linguistics considers issues of preference of structures in different varieties, registers, styles as also in spoken and written modes communication, collocational behaviour of words, phrases, clauses and larger structures like discourses or texts. Such an investigation can also throw light on social group and individual preferences, gender preferences, intertextual comparisons and language use in a variety of situations like talking to a family member, reading a newspaper, reading an academic article, writing a letter etc. Therefore, this type of analysis reveals typical patterns rather than making judgements of grammaticality.

-------------------------------------------------------------------------------------------------------------

*Here the discussion and illustration is mainly based upon Jan Aurts and William Meijs (1986), Karin Aljmer and Bengt Altenberg (1991), Jan Startvik (1991), Dougles Biber, Susan Conrad and Randi Rippen (1998).

### 2. Objectives:

Corpus analysis has two main goals as follows:

i.   Assessing the extent to which a pattern is found; and
ii.  Analysing the contextual factors that impact variability.

This is a complex process methodologically for the following reasons:

(a) Typical patterns cannot be a base on intuitive or anecdotal evidence.
(b) This type of analysis calls for a large amount of data on the language used.
(c) Multiple contextual factors have to be considered, for example language used in conversation, academic articles, news reportage, literature or fiction etc.

Because of these reasons, studies of language use in the past remained unfeasible or impossible.

### 3. Features of Corpus-Based Analysis

(i)     It is empirical and analyzes the actual patterns of use in natural texts.

(ii)    It utilizes a large and principled collection of natural texts known as 'corpus' as the basis of analysis.

(iii)   It makes an extensive use of computers for analysis, employing both automatic and interactive techniques.

(iv)    It depends on both quantitative and qualitative techniques (qualitative in terms of interpretation on the basis of quantitative analysis).

An extensive use of computers is made in this type of analysis because computer analysis of data is reliable and a larger data base can be developed and stored, it can be used interactively and it enables a human analyst to make linguistic judgements which are otherwise different. As done formerly by scholars, this type of analysis enables research to identify complex 'associative patterns' and/or 'collocations'.

In the traditional approaches, it is not possible to handle large corpora. Corpus analysis can supplement the results of traditional analysis and thus it is only complementary to traditional approaches and not a substitute. The results of corpus analysis can be useful for lexical and structural analysis, conversation or speech analysis, analysis of spoken and written registers, varieties or styles of language use, language acquisition and developmental stages of children different ways of writing academic or research articles, grammatical errors made by L2 learners in writing essays etc., comparing errors made by L1 and L2 learners, and information ordering.

Corpus-based analysis can be applied to empirical studies in several areas of linguistics like lexicography, grammar, dialect and register patterns in sociolinguistics, language acquisition and learning in psycholinguistics, in stylistics, designing effective teaching material and testing in educational linguistics. It is also an integral part of computational linguistics.

As Biber et al (1998) point out

> For instance, where traditional approaches might identify a group of synonymous words, corpus-based lexicographic research attempts to show how related words are used in different ways and are appropriate in different contexts. (p. 21)

In corpus-based lexicographic research, six major issues are addressed:

i)      Meanings  associated with a particular word;

ii)     Frequency of occurrence of a word in comparison with other related words;

iii)    Non-linguistic association patterns of words ( i.e., context in registers, historical periods and dialects) ;

iv)    Co-occurrences or collocations of a word and its distribution with other words;

v)    The distribution of meanings and uses of words; and

vi)    The distribution of synonyms differently in different uses or registers etc.

A similar analysis of grammatical pattern or structures can be made by using a large corpus.

## 4.  Illustration

Let us now look at some examples of how corpus-based analysis can throw new light on grammatical pattern and lexical items like synonyms.

**4.1.1 Grammatical pattern:**  In English that–verb complement clause and to-verb complements similar in structure and meaning are different in frequency of occurrences. Traditional structural analysis can also describe the difference between these clauses, but it cannot explain why English has such multiple structures and why they have different preferences in different styles and registers of English. Consider the following

E.g. I hope that I can go. (that –verb complement)

I hope I can go. (that deletion)

I hope to go. (To-verb complement)

As, Briber et. al ( *op. cit.* ) observe, in a corpus of 4 million words  in two registers i.e. conversation and academic prose, their analysis revealed that that-clauses occur very commonly in conversation while to-clauses are almost common in the two registers. Further, that-clause occurs more frequently because of its alternative structure i.e. extraposed construction. Look at the following pair of sentences:

e.g. That these identifications are circular is unimportant at the moment.

It is unimportant at the moment that these identifications are circular.

The second sentence is easier to process and to understand because of the time-gap available to the listener (p.77). The authors observe that this alternative extraposed construction is an important factor that makes users to prefer that-clause to to-clause between the two complement clauses. They further demonstrate by giving the relevant data in a tabular form, that ESL textbooks give examples and descriptions of subject that-clauses, but that the textbooks fail to explain the functions of that-clause. This gap can be filled by corpus-based analysis. Next, they demonstrate how different verbs control these clauses. For example, want and try can take only one type of complement clause (i.e. to-clause) while other verbs like think, say and know control both types of complement clause. The most common lexical association of that-clauses are their

typical use in reporting thoughts, feelings and sayings which is the most common in conversation rather than in Academic Prose. Similar analyses can be made about Reference, Tense and Voice.

### 4.1.2.  Synonymous: <u>Big</u>, <u>Large</u> and <u>Great</u>

Likewise Biber et al (1998) show the difference of meaning and occurrence of these synonymous adjectives in English, on the basis of their analysis of corpus consisting of data from Academic Prose and fiction. In a corpus of one million words, their analysis reveals that <u>big</u> and <u>great</u>   occur more frequently in fiction while <u>large</u> occurs with a greater frequency in Academic prose. As these authors farther observe:

<u>Big</u> is most often used to refer to physical size, white <u>large</u> is used more to refer to quantities and amount. <u>Great</u> is also used for amounts, especially in the collocation <u>great deal</u>, but this adjective has a wider range of meanings, ranging from intensity to size, to familial relationships (e.g. great –aunt ) (p.51)

<u>Great</u> also collocates, for example with <u>majority</u>, <u>variety</u>, <u>extent</u>, <u>part.</u> (*Ibid*)

### 4.1.3 Nominalizations

Biber et al also illustrate how corpus-based analysis can show the frequency of nouns. Their analysis shows that nominalizations like <u>movement,</u> development<u>,</u> <u>relation</u> and <u>equation</u> occur with a great frequency in Academic Prose because these words treat actions and processes as abstract objects separated from their human participants.

Similarly, of the nominalization endings like <u>–tion/-sion</u>, <u>-ment</u>, <u>-ness</u> and <u>–ty</u>, <u>-</u> <u>tion/sion</u> occurs 68 % in Academic Prose, 51% in Fiction and 56% in Conversation. They explain its highest frequency in Academic Prose being due to the fact that it denotes a generalized process or state.

### 5.  Tools of Analysis

For a corpus-based analysis of either grammatical structures or lexical items, programmes are available at a low cost. Most easily available of these are as follows

- I.     The London-Lund Corpus (500,000 words)
- II.    The Lancaster-Oslo/Bergen Corpus (LOB) (Individual  Registers, 12,000-160,000 words)
- III.   The British National Corpus (BNC) (4000,000 words of  Conversation register)
- IV.    The Longman-Lancaster Corpus (LLP) (5,000,000 words)

Some of these packages are also called 'concordancing' programmes.

## 6. Conclusion

Using these corpora, it is possible to select structures and lexical items like words and phrases appropriate to a given level of learning of English as a second language for the benefit of the learners in non-native countries like India. Such information is highly useful for syllabus design and textbooks production and preparation of teaching materials for different levels of learners of English in the field of ELT.

**Works Cited:**

Aarts, Jan and William Meijs (eds.) 1986. *Corpus Linguistics II*.  Amsterdam: Rodopi.

Aljmer, Karin and Bengt Altenberg (eds.) 1991. *English Corpus Linguistics.* London :Longman.

Austen, J.L.1962. *How to Do Things with Words*. Oxford: Clarendon.

Biber, Douglas, Susan Conrad and Randi Rippen. 1998. *Corpus Linguistics : Investigating Language Structure anf Use*. Cambridge;  Cambridge University press.

Canale, M and M. Swain. 1980. " *Theoretical Bases of communicative approaches to second language and testing*". Applied linguistics I. Pp.1-47

Chomsky, Noam 1965. *Aspects of the theory of syntax.* Cambridge MA :MIT Press.

Chomsky, Noam. 1957 *Syntactic Structures*. The Hague : Mouton.

Grice, H. P. 1975. " *Logic and conversation* ". In J. Cole and J. Morgan (eds) Syntax and semantics III. New York : Academic Press. PP :41-58

Halliday, M.A.K. and R. Hasan.1976. *Cohesion in English*. London: Longman.

Hymes , Dell . 1972. "*On communicative Competence"*. In Pride and Holmes (eds.) *Sociolinguists* pp.269-293

Leech, G.N. 1974. *Semantics: The Study of Meaning*. Harmondsworth : Penguin Books .1981, reprint.

Levinson, S. 1983. *Pragmatics.* Cambridge : Cambridge university press

Searle , J .R. 1969. *Speech acts*. Cambridge: Cambridge University Press.

Simpson , P. 1997. *Language through Literature*. London: Rout ledge.

Sinclair , J . McH , and M. Coulthard 1975.  *Towards an Analysis of Discourse*. Oxford: Oxford University Press.

Sperber, D. and D. Wilson. 1986. *Relevance : Communication and cognition.* Oxford : Blackwell.

Svartvik, Jan (ed.) 1992. *Directions in Corpus Linguistics: Proceedings of Nobel Symptoms 82* . Berlin: Mouton.

Widdowson, H. G. 1973. *An Applied Linguistic approach to discourse analysis*-Ph.D. Dissertation. University of Edinburgh (unpublished).