

# **LANGUAGE IN INDIA**

## **Strength for Today and Bright Hope for Tomorrow**

### **Volume 11 : 9 September 2011**

#### **ISSN 1930-2940**

**Managing Editor: M. S. Thirumalai, Ph.D.**

**Editors: B. Mallikarjun, Ph.D.**

**Sam Mohanlal, Ph.D.**

**B. A. Sharada, Ph.D.**

**A. R. Fatihi, Ph.D.**

**Lakhan Gusain, Ph.D.**

**Jennifer Marie Bayer, Ph.D.**

**S. M. Ravichandran, Ph.D.**

**G. Baskaran, Ph.D.**

**L. Ramamoorthy, Ph.D.**

## **Corpora: The Future of ELT in Pakistan**

**Rashid Mahmood, M.Phil.**

**Asim Mahmood, M.Phil.**

---

### **Abstract**

The advent of corpus linguistics has brought new dimensions in linguistic and language teaching theories. Traditional view of the internal structure of human language claimed that grammar provides empty skeletons of utterances, later filled with appropriate lexis in the course of discourse formation.

The new schematic approach introduces the notion of lexicogrammar and combines the previously separate fields of grammar and vocabulary. It sees utterance production “as exploiting ready-made memorized building blocks or ‘pre-fabs’, put together using simpler ‘jerrybuilding’ operations” (Aston 1995). It implies that the process of learning can be seen as approximating the observed patterns to form the schemata.

Corpus based research has an edge over intuition based research as it provides evidence from a large scale authentic data. In Pakistan, corpus based research can provide solutions to the problems like status of Pakistani English, confusing criteria for students’ errors, subjective judgments on differences from standard English, and material development on sound basis etc.

Very few researchers embarked on studying Pakistani English and none ventured studying it on the basis of corpus. In the absence of any research which may categorically establish standards of English in Pakistan, there are no parameters to differentiate errors from deviations.

Language in India [www.languageinindia.com](http://www.languageinindia.com)

11 : 9 September 2011

Rashid Mahmood, M.Phil. and Asim Mahmood, M.Phil.

Corpora: The Future of ELT in Pakistan

The corpus based research can open new vistas in ELT. It facilitates testing the hypotheses based on intuition. It can help establish Pakistani English a new variety. The results of such a research would revolutionize the practice of ELT. Language Policy would undergo a substantive change. This indigenized variety (just like British English, American English, Australian English, Indian English etc) will help syllabus designers, text book and grammar writers and lexicographers focus their attention on the areas where learners need more practice. Both spoken and written corpora are useful for researchers, teachers and students alike. Teachers can compile their own corpora based on the textbooks they are teaching. They can make frequency indexes of the vocabulary and can prepare concordances to disambiguate words' senses and to show the patterns of use.

## **1 Introduction**

A corpus is defined as “a collection of naturally occurring language text, chosen to characterize a state or variety of a language” (Sinclair, 1991). Corpora help us empirically analyse the actual patterns of use. Varieties of computer software are available to extract certain linguistic information from the corpus. The corpus studies allow both quantitative techniques and qualitative interpretation.

Corpus linguistics is not a branch or discipline of linguistics, rather a methodology or a tool. According to Meyer (2002), “corpus linguistics is more a way of doing linguistics than a separate paradigm within linguistics”. Corpus resources have grown a lot in the past ten years. British National Corpus (BNC), American National Corpus (ANC) and International Corpus of English (ICE) are huge repositories of English corpora. S.V. Shastri and his colleagues compiled first Indian English corpus, Kolhapur Corpus of Indian English, in 1998. It contains one million words of Indian English. The authors of this paper have compiled a corpus of three million words of written Pakistani English. Very soon the corpus will be available for pedagogical purposes.

The utility of corpora for ELT can not be overemphasized. The need of corpus studies is being felt by the discerned linguists and teachers. This area of research is in its infancy in Pakistan. The researchers including Talaat, Rehman, Baumgardner were either engaged in item analysis or in the analysis of process of the text to verify their subjective judgment. The scarcity of research on Pakistani English leads to the borrowing of results obtained from researches on other regional varieties of English. Such results usually do not exactly correspond to the variety of English used in Pakistan.

## **2 Why Do We Need a Corpus of Pakistani English?**

Data taken from any other variety of English can not substitute the corpus of Pakistani English due to significant differences among them. These differences occur at all levels of linguistic analysis especially at lexical and syntactic level. The following table shows some lexical variations in Pakistani English pointed out by Baumgardner and Taalat (1993). The frequency of these lexical items and their equivalents (which they have mentioned) in BNC has been

Language in India [www.languageinindia.com](http://www.languageinindia.com)

11 : 9 September 2011

Rashid Mahmood, M.Phil. and Asim Mahmood, M.Phil.

Corpora: The Future of ELT in Pakistan

studied from Mark Davies, Brigham Young University Website (2007) by the present researchers to verify their subjective judgment.

**TABLE 1**

## **Some variations pointed out by Baumgardner and Talaat (1993)**

<b>Words or phrases in PE</b>	<b>Equiv. In BE</b>	<b>Equiv. In AE</b>	<b>Freq. in BNC</b>	<b>Freq. of equiv. in BNC</b>
Ear-tops	Ear-studs	Ear-studs	1	1
Undertrial	An accused of a crime	An accused of a crime	3	2
Mudguard	wing	fender	5	20
Boots	Shoes	Shoes	2433	3506
Bogies	Railway carriage	Railway carriage	131	60
Footpath	Pavement	Sidewalk	414	1263
Conveyance	Transport	Transport	60	7011

The first column of this table shows the lexical items of Pakistani English henceforth PE pointed out by Baumgardner and Talaat. The second column gives their equivalents in British English henceforth BE and the third column gives their equivalents in American English henceforth AE (equivalents are also given by the previous researchers). The fourth column provides the frequency of PE items in BNC and the last column gives the frequency of equivalents in BNC.

The comparison provides interesting results. The frequency of ‘ear-tops’ (PE) and ‘ear-studs’ (BE) is same i.e. 1 in BNC. ‘Undertrial’ (PE) appeared 3 times while ‘an accused of a crime’ (BE) appeared 2 times in BNC. More results can be seen from Table 1.

**TABLE 2**

## Some variations pointed out by Baumgardner and Talaat (1993)

Words or phrases in PE	Equivalents in BE	Words or phrases in PE	Equivalents in BE
Bearer	A waiter in a club	Monthly	Protection money
Bed tea	Morning tea	Pen-down strike	A strike in which staff report to the office but do not work
Cash memo	Receipt	Side-hero	Supporting actor
Cent per cent	One hundred per cent	Stepney	Spare wheel
Hotel	An eatery	Tube light	Strip light
Incharge	A person in charge	Tease	Harass
Medical store / hall	Pharmacy	Seat	Place / position

Table 2 shows the items pointed out by Baumgardner and Talaat, which are common in PE but did not appear in BNC. ‘Hotel’ in the meaning of an eatery or restaurant, ‘Tease’ in the meaning of ‘Harass’ and ‘Seat’ in the meaning of ‘Place or Position’ have not been used in BNC.

These are some of the lexical trends manifested in Pakistani English. Can we label them deviations or they are mere errors? Any difference needs consistency to become a deviation. Such consistency can be verified only by a large corpus.

### 3 How Can a Corpus of Pakistani English Improve the Situation?

Hudson (1980) describes the procedure of establishing a standard variety under four steps: selection, codification, elaboration of function and acceptance. Regarding selection of a variety, learners of English in Pakistan face problems as textbooks follow British English but most of the exposure of English through international media is of American English. The influence of the both along with sociocultural differences results in differences from them. The in-depth corpus based analysis of Pakistani English can maintain the difference between errors

Language in India [www.languageinindia.com](http://www.languageinindia.com)

11 : 9 September 2011

Rashid Mahmood, M.Phil. and Asim Mahmood, M.Phil.

Corpora: The Future of ELT in Pakistan

and deviations on the basis of frequency of occurrence. Hudson (1980) comments on codification and acceptance “Some agency .... must have written dictionaries and grammars books to ‘fix’ the variety, so that every one agrees on what is correct. Once codification has taken place, it becomes necessary for any ambitious citizen to learn the correct forms and not to use in writing any incorrect forms..... the variety has to be accepted by the relevant population as the variety of the community.” The most important process i.e. codification can be carried out best by getting insight from the corpus studies. Codification draws clear line between errors and deviations. Systematic deviations will be considered norms of this variety and efforts will be made to eradicate the errors by remedial measures.

#### **4 Uses of Corpora for ELT**

Establishing the norms of Pakistani English as an independent variety will bring about a significant change in the language policy. Americans are proud of using American English in their own way, why the Pakistanis should be ashamed of not sounding like the Britishers. The change in language policy will be reflected in curriculum, syllabuses, text books and grammars. New trends and teaching materials are bound to appear on the basis of the norms of the newly established variety.

Syllabus designers can benefit a lot from the insight provided by the corpus studies. Corpus based research can help point out the frequent errors and then the syllabus designers can focus the linguistic areas where learners need more practice. Corpora are rich resources for the text book writers as well. A corpus presents authentic material actually used by the users of the language. The text book writers will benefit from the corpus analysis in two ways: first, they can introduce new vocabulary and syntactic complexities gradually (by indexing vocabulary and preparing concordances from corpora), secondly, they can introduce speaking activities to encourage natural responses. McCarthy (2004) writes, “If we look at hundreds of uses of must in the Spoken Corpus we find that, on average, only 5 percent of all its uses are connected with obligation (e.g. you must have a visa to enter the United States). Another 5 percent are in expressions such as I must admit and I must say. But the overwhelming majority of uses of must are in “predictive” statements such as that must have been nice, you must be hungry, etc. Statistics of this kind helps the text book writers set priorities in grammar teaching and find the most natural contexts for teaching grammar”.

In Pakistan, usually text books are the only teaching material with the teachers. It is quite difficult for them to find authentic examples to explain certain usage of language. If they are trained to use corpora, they can find hundreds of authentic examples. They can use concordances as a source of relevant and up to date linguistic information, which can be found more quickly than in a dictionary and accompanied by numerous examples of use. Teachers can create learner corpora from students’ writings or transcribed utterances and use it as a diagnostic tool for error analysis.

The students can be trained to browse freely through the corpus and by inventing their own queries for the software like AntConc or WordSmith etc. They may infer some reoccurring patterns or check some of their earlier presumptions concerning grammar or

Language in India [www.languageinindia.com](http://www.languageinindia.com)

11 : 9 September 2011

Rashid Mahmood, M.Phil. and Asim Mahmood, M.Phil.

Corpora: The Future of ELT in Pakistan

usage. Students may be asked to analyse their mistakes with the use of AntConc or WordSmith or to work with the corpus to solve linguistic problems.

Lexicographers can find out how words have actually been used and the frequency of occurrence can determine the entry of words in learner dictionaries. Grammarians can improve their descriptions to fit the facts of actual use of construction. Their descriptions can better match the needs of the learners. Linguists specifically researchers on language learning can see how often the learners with a particular L1 get something wrong. Stylisticians can see foregrounding easily by finding out lexical and grammatical patterns in the particular piece of literature. The results can be compared with the other works of the same author or other authors. The writings of the same author or different authors can be compared on all the levels of linguistic analysis by using sophisticated software. The recommendations of the present researchers to cater to the future needs of ELT regarding Corpus Based Research are as follows:

- Workshops on Corpus Linguistics
- Establishment of Corpus Linguistics Research Centre (CLRC)
- Collection of Data up to 100 Million words of Pakistani English on the pattern of BNC so that the results could be compared with BNC
- Lexicographical activities including Dictionary of Pakistani English and Dictionary of Collocation  
(Entries based on frequency of usage and their sense in PE)
- Grammar Of Pakistani English (Clearly defining the differences between PE and other varieties of English)

=====

### References

Aston, Guy. (1995). Corpora in language pedagogy: Matching Theory and Practice. In G. Cook and B. Seidlhofer (eds.) *Principle and Practice in Applied Linguistics*. Oxford University Press.

Baumgardner R. (1993). *The English Language in Pakistan*. Karachi, Pakistan: Oxford University Press.

Davies, M. (2011). Corpus Linguistics, BYU. Retrieved from <http://davies-linguistics.byu.edu/personal/> accessed on 16-07-2011

Hudson , R .A. (1998). *Sociolinguistics* . Cambridge, Cambridge University Press.

McCarthy, M. J. (2004). *Touchstone: From Corpus to Coursebook*. Cambridge, Cambridge University Press.

Meyer, C. F. (2002). *English Corpus Linguistics*. Cambridge University Press

Language in India [www.languageinindia.com](http://www.languageinindia.com)

11 : 9 September 2011

Rashid Mahmood, M.Phil. and Asim Mahmood, M.Phil.

Corpora: The Future of ELT in Pakistan

Sinclair, John (1991). *Corpus, Concordance, Collocation*. Oxford. Oxford University Press:

Talaat, M (1993). Lexical Variation in Pakistani English. In R, Baumgardner, (ed.) *The English Language in Pakistan*. Karachi, Pakistan. Oxford University Press.

=====

Rashid Mahmood, M.Phil.  
Assistant Professor  
Government College University  
Faisalabad  
Pakistan  
[ch.raashidmahmood@gmail.com](mailto:ch.raashidmahmood@gmail.com)

Asim Mahmood, M.Phil.  
Assistant Professor  
Government College University  
Faisalabad  
Pakistan  
[masimrai@gmail.com](mailto:masimrai@gmail.com)

Language in India [www.languageinindia.com](http://www.languageinindia.com)

11 : 9 September 2011

Rashid Mahmood, M.Phil. and Asim Mahmood, M.Phil.

Corpora: The Future of ELT in Pakistan