

Achieving Academic Proficiency Standards in Higher Education through Corpus-Based Language Teaching

Vijayakumar C

Assistant Professor, Dept. of English, Deanship of PYP, Imam Abdulrahman Bin Faisal University, Dammam, KSA

Email: vschinta@iau.edu.sa

ABSTRACT

Schools and colleges in India are now facing the challenge of preparing students for higher studies in countries like the USA and the UK, where English is the primary mode of study. It is argued that a large number of ESL and EFL learners fail to meet the expected academic proficiency standards of B2 of the CEFR due to their poor prior exposure to English. Mushrooming of institutions offering short-term training in English for TOEFL and IELTS sets the tone of the problem. However, corpus based approach, when embedded into our curriculum practices, can address a range of teaching issues, and help our students cope with the needs of higher education, in India as well as abroad. While corpus-based language teaching (CBLT) is not new to India, there are reasons why it has not received as much support as other approaches did. In this paper, I discuss the popular notions that withheld CBLT from its growth and present how corpora—general or specific—can assist ELT practitioners achieve high quality standards in higher education.

Keywords: *higher education, corpus based language teaching, curriculum design*

The importance of CALL technology in ELT has been strongly felt in the last twenty five years (Fotos & Browne, 2004). Higher education programs/courses offered by reputed universities in the developed countries through MOOCs and other online platforms have extensively utilized various computer-driven applications and modalities to organize their courses. While some CALL applications such as webcasts, emails, and mobile-based applications have provided course developers with appropriate

channels to organize and present the content, computer-based corpus tools have conveniently clubbed both the modality and language aspects of CALL (Sinclair, 2004). In fact, independent large-scale native speaker corpora such as Corpus of Contemporary American English (COCA) (Davies, 2009) and British National Corpus (BNC) and specific-purpose home-grown academic corpora such as Cambridge English Corpus (CEC) and MICASE have made a significant impact on the language

teaching practices in the developed countries such as the UK and the USA (Davies, 2009). Increasingly, the researchers and language teaching practitioners are depending on corpus evidence to successfully achieve their teaching and testing objectives (Gavioli & Aston, 2001; Sinclair, 2004).

While this widespread use of corpus in LT is overwhelmingly optimistic about its potential in the Western world, countries like India are skeptical about embracing corpus tools in formal language teaching contexts. In this paper I discuss how corpus inclusion can promote high standards in higher education.

Corpora in Language Teaching and Testing

Reliance on *evidence* is not new to language studies. In fact, a probe into the history of dictionary making reveals the importance real-life language instances in describing language, especially vocabulary and grammar (Hanks, 2012). Recent developments in computer technology have remarkably facilitated the process of organizing and presenting idiosyncratic patterns in real-life instances, both syntagmatically and paradigmatically, in effective ways (Hanks, 2012). The ability to process large size corpus—a collection and organization of texts—for specific patterns of language use and the inherent patterning of language use showed the world a colorful picture that was hitherto unknown (Sinclair, 2004).

Large scale applications of corpus linguistics

in language pedagogy eventualized when linguists attempted to analyze corpora to describe language in use. Lexicographers collected and analyzed large amounts of real-life data for form, meaning, and use aspects. Subsequently, corpus analysis procedures influenced the research practices of other allied fields of language studies such as ESP and EAP that looked for discourse specific language features (Biber, Johansson, Leech, Conrad, & Finegan, 1999).

The effects of corpora are mainly seen on three areas of ELT: learner dictionary making, classroom teaching, and learner corpora for SLA, ESP and EAP (Mukherjee, 2006). Although corpus studies have brought in radical reforms in dictionary making in terms of defining features, examples, collocations, pattern presentation among others (Hanks, 2012), their proposed classroom methods such as Data Driven Learning (DDL) (Johns, 1991) have not received as much support as the CLT and other methods of language teaching, especially in the third-world and developing countries. The following are some of the practicality related reasons for the disbelief among many ELT practitioners.

a. Corpus building and use are expensive.

Building an authentic database for teaching purposes requires the practitioners to collect a manageable size of a database, usually of a few hundred thousand instances. In developing countries the resources to build such large-scale databases are scarcely available. While building a database of

written texts is relatively cheaper, transcribing or coding the data into analyzable formats is both time consuming and expensive.

b. Corpus analysis requires specialist knowledge.

Data driven learning is a research process where the learner/teacher is required to deduce his answer by carefully analyzing raw instances. Often the instances make little sense to the observer because they are partial instances of real life experiences. Observers whose linguistic schemata are minimal could misrepresent their interpretations thereby forming wrong generalizations.

c. Corpus analysis is a time-consuming task.

In ESL countries learning English language is important; however, it is not mandatory that they use a native-speaker like English. Moreover, in formal or traditional classrooms teachers are under great pressure to cope with the institutional demands and the requirements of the ongoing and summative examinations. An intensive investigation of corpus for specific language aspects will consume a lot of time.

d. Corpus based language teaching needs access to computers.

Often corpus related demonstrations, in seminars or workshops, highlight the role of modern computers that can process texts in less than a few seconds. This explication has led teachers believe the inseparable role of computers within a language teaching

class. In fact, in many developing countries access to computers in the classroom is still a desire to be fulfilled.

e. Textbook based examinations

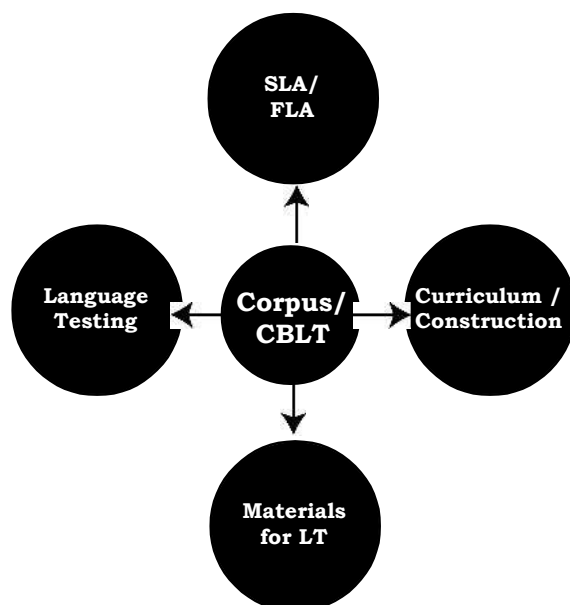
Most importantly, in developing countries the reforms introduced in ELT are mainly methodological. TBLT and CLT like approaches found their way into the language classrooms through the textbooks prescribed for study. Textbooks till-to-date are the primary sources of linguistic input for both teaching and testing practices. For instance, in countries like Saudi Arabia, many universities strictly follow pacing schedules in designing tests. Although innovative case studies that utilized recent research developments exist, classroom instruction is primarily textbook based.

The beliefs stated are pervasive and will influence the ELT practices some more time. To some extent, it is true that direct corpus use is an expensive intervention and requires a specialist's knowledge in building and organizing content. However, there are several ways we could utilize readymade and freely available corpus in achieving our objectives of language courses (Thurstun & Candlin, 1998; Davies, 2009). Indeed, in higher education institutions, online concordances are now fruitfully used to supplement and complement language instruction (Daskalovska, 2013), in many ways that facilitate the ongoing practices rather than interfere with them.

CBLT for Higher Education

Textbook writing, design of online platforms

for language study, design of large-scale evaluation procedures, study of learner language, discourse analysis are a few areas of ELT study that are currently relying on corpus evidence. Most importantly, in higher education contexts, where academic disciplines follow specific conventions of language use, the direct and indirect application of corpus tools and the effects of research findings is extensive (Gavioli & Aston, 2001). The following discusses how the current CBLT research can influence various stages of language curriculum implementation.



Learning Patterns and Language Teaching

A cross comparison between non-native learner language and large native speaker data enabled the SLA researchers identify the typical patterns and gaps in learner language use (MacWhinney, 2016). For instance, processing of a large learner corpus offered researchers (Gablasova, Brezina, & McEnery, 2017) with insights

about the development of collocation knowledge among the L1 and L2 users of English. The researchers found that the L2 users of English mostly preferred to use, for example, predictable and strongly associated words that frequently co-occur rather than low frequency words.

Similarly, in a longitudinal research study researchers attempted to form generalizations about L2 learners' language development patterns for tense and aspects by quantitatively analyzing a large database of learner language (Meunier & Litte, 2013). They concluded that advanced learners exhibited difficulty in using present progressive for planned events while they reported no issues with using the progressive to refer to 'ongoingness' of an activity (p: 72). The purpose of utilizing a corpus-informed investigation here, according to the researchers, was to inform the 'educational practices' of teaching tense with a high degree of certainty. Studies that relied on corpus methods in SLA are plenty. One of the main reasons to integrating corpus with SLA research is to provide reliable evidence that can inform pedagogy.

Institutional Registers and Curriculum Design

Since English ensures access to 'institutional registers' (Biber, 2006) and facilitate communication, the role of English is instrumental in higher education. The underlying assumption of academic English programs is that success of students on university level programs depends on their ability to handle discourse specific language features. Corpus based explorations could

help institutions define academic standards by identifying the idiosyncratic and frequent ‘linguistic characteristics’ (2006:2) of academic registers. These frequent ‘characteristic lexico-grammatical features’ are exploited by the discourse communities to achieve the intended communicative purposes (2006: 12). For example, a multi-dimensional analysis of the academic sub-corpus of Longman Grammar of Spoken and Written English (LGSWE) (Biber, Johansson, Leech, Conrad, & Finegan, 1999) identified the idiosyncratic features that are ‘much more common in academic prose’ (2006: 14). Some most commonly exploited features in academic prose, according to Biber *et al* (1999) include: non-finite relative clauses, nominalization, attributive adjectives, and derived verbs (especially formed with *re-* and *-ize*) (for a detailed list please read Biber, 2006). EAP programs across the world attempt to equip students either with the general linguistic aspects that are pervasive across all academic disciplines or with the register specific features that are specific to the disciplines the students will study. Corpus studies can help institutions identify and organize such register specific features for curriculum design.

When sub-skills such as identifying and analyzing are combined with a certain textual features such as verbs, adjectives, articles students can decode and produce texts appropriate for the contexts. For instance, for a stated objectives/sub-skills of identifying an author, the reader will attempt to decipher the linguistic clues such

as the use of personal pronouns and referential vocabulary. Similarly, to distinguish a fact from opinion, the learner needs to understand the use of verbs, signaling words such as *in my opinion*, *in fact*, *believes* among others. To analyze a text in terms of its cohesion and coherence the reader can be directed to notice transition expressions such as *firstly*, *subsequently*, *while*, *on the other hand* among others. In other words, the existing curriculum/syllabus can judiciously make use of corpus instances (longer or shorter) to provide ample practice for a specific aspect.

Authentic Instances in Dictionaries and Textbooks

Identifying and using language instances that do not pose a threat to a student’s understanding of the text is important. It is here the practitioners have to apply caution. The effect of ‘genuine’ or ‘real life’ or ‘natural English’ (Sinclair, 2004) instances on language teaching was strongly felt in the last 20 years. Series of textbooks such as *Touchstone* (McCarthy, McCarten, & Sandiford, Touchstone Series, 2005), free online corpus platforms such as MICASE and COCA (Davies, 2009), self-study materials (McCarthy & O’Dell, 2016), word-lists (Coxhead, 2000) are now available for use.

In the field of materials design two changes have been brought in: selection and identification of linguistic categories and design structure of materials. In identifying the authentic instances of specific linguistic categories researchers have used the

following criteria:

- a. Using of SLA research findings to identify learning objectives (Gablasova, Brezina, & McEney, 2017)
- b. Frequency of occurrence (Nation, 2018)
- c. Contextual relevance/ registers (Biber, Johansson, Leech, Conrad, & Finegan, 1999)
- d. Patterns of co-occurrence (Sinclair, 2004)
- e. Linguistic complexity of the text (choice of vocabulary, choice of grammatical structures and text length)
- f. Discourse specificity (reader specificity)

In designing materials they mainly use the strategy of reasoning: deducing the rule from multiple instances. Every target linguistic category to be taught is shown in multiple authentic instances, and the learners were expected to reason out the underlying rule(s) with the help of context, linguistic clues, repetition, and analysis. Subsequently, the learner can compare his/her answers with the textbook.

Language Assessment Practices

Free access to large databases and corpus analysis tools such as Range (Nation, 2018) simplified the process of compiling a corpus. Three types of corpora—reference, specialized, and learner—are now extensively built for various purposes (Cushing, 2017). Even though some corpora are primarily built to provide linguistic descriptions, a range of findings such as frequency studies, collocations, and context

specific grammar and vocabulary have significant effects on assessment.

While a reference corpus or a large learner corpus representing the target language or the target learner group can offer the curriculum designers and the test developers with the choice of linguistic aspects to be considered, a purposive specialized corpus identifies discourse specific typical features to be mastered and assessed. Moreover, the information obtained from corpus analysis is used to conceptualize “domain definition, construct definition, and the construction of tasks and test items that authentically reflect the target language use domain” (Cushing, 2017: 442). Statistical corpus based evidence is now helpful in determining the complexity of language use, distinguishing between B2/C1 levels of CEFR as the threshold proficiency for admission into higher education, and in ‘characterizing’ a specific level by providing ‘language specifications’ (Paquot, 2018).

If appropriate use of discourse specific linguistic features determines the success of students on academic programs, the institutions need to examine whether the students are capable of understanding and using them. Assessment of students’ performance needs to take into consideration the identification of learning goals, linguistic categories, content and appropriate scales and rubrics for assessment.

Constructs that insufficiently define their learning goals will not be able to provide

clear-cut guidelines for test development. Also, any random selection of linguistic categories that do not adequately represent the real life use of language will succumb to the dangers of being irrelevant. And fabricating hypothetical instances that are distantly related to the discourse in question skews content validity and test reliability. Finally, evaluation procedures that refer to vague interpretations of scales and rubrics will provide unreliable results. In other words, language teaching and testing is no way a straight forward process: Every decision presupposes a clear rationale with clearly stated end results. If the defined constructs do not represent the predefined goals; and if the presupposed objectives are not represented in the test-design practices, the end results of the course will be unreliable. Corpus based construct definitions, corpus-based syllabus design and corpus-based teaching and testing can help course developers and teachers address these concerns.

Conclusion

While promoting data-driven academic standards is a commendable choice, developing a far-removed standard of English that the students have a limited access to will affect the quality of educational intervention. However, corpus evidence when drawn especially from contexts of high relevance and when used with effective methodologies can revolutionize language teaching. Therefore, the ELT practitioners need to apply caution before embarking on a full-fledged use of corpus in language courses. In fact, novice

Journal of English Language Teaching LX/5, 2018

users can use corpus, initially, only in specific stages of language course development—in selecting instances to teach collocations, in identifying examples for specific grammatical categories and to devise comprehension related cloze items—and develop specific sub-skills that influence the overall learning process of the students.

Works Cited

- Oxford University Press. (2014). *Oxford Learner's Dictionary of Academic English*. Oxford: Oxford University Press.
- Biber, D. (2006). *University Language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins B.V.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Coxhead, A. (2000). A new academic worlist. *TESOL Quarterly* , 34 (2), 213-238.
- Cushing, S. T. (2017). Corpus linguistics in language testing research. *Language Testing* , 34 (4), 441-449.
- Daskalovska, N. (2013). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning* (September), 1-16.
- Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics* , 14 (2), 159-190.
- Fotos, S., & Browne, C. (2004). *New*

- perspectives on CALL for second language*
 Mahwah: New Jersey: Lawrence
 Erlbaum Associates, Inc.
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and Interpreting the evidence. *Language Learning* , 67 (S1), 155–179.
- Gavioli, L., & Aston, G. (2001). Enriching reality: language corpora in language pedagogy. *ELT Journal* , 55 (3), 238-246.
- Hanks, P. (2012). The corpus revolution in lexicography. *International Journal of Lexicography* , 25 (4), 398–436.
- Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal* , 4, 1-16.
- MacWhinney, B. (2016). A shared platform for studying second language acquisition. *Language Learning* , 67 (S1), 254-275.
- McCarthy, M., & O'Dell, F. (2016). *Academic Vocabulary in Use (2nd Edition)*. Cambridge: Cambridge University Press.
- McCarthy, M., McCarten, J., & Sandiford, H. (2005). *Touchstone Series*. Cambridge: Cambridge University Press.
- Meunier, F., & Litte, D. (2013). Tracking learners' progress: adopting a dual 'corpus cum experimental data' approach. *The Modern Language Journal* , 97 (S1), 61-76.
- Mishan, F. (2004). Authenticating corpora for language learning: a problem and its resolution. *ELT Journal* , 58 (3), 219-227.
- Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art—and beyond. *English Corpus Linguistics* , 3, 5-24.
- Nation, P. (2018). *School of linguistics and applied language studies* . Retrieved 6 29, 2018, from www.victoria.ac.nz: <https://www.victoria.ac.nz/lals/about/staff/paul-nation>
- Paquot, M. (2018). Phraseological competence: A missing component in university entrance language tests? Insights from a study of EFL learners' use of statistical collocations. *Language Assessment Quarterly* , 15 (1), 29-43.
- Sinclair, J. M. (2004). *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins Publishing Co.
- Thurstun, J., & Candlin, C. N. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes* , 17 (3), 267-280.