

Online Corpora for English Language Learning

P. Aswini

School of social sciences and Languages, VIT University, Vellore- 632014, Tamil Nadu, India

Dr.R. Srinivasan
Associate Professor

School of Social Sciences and Languages, VIT University, Vellore- 632014, Tamil Nadu, India

Abstract

The use of corpora and corpus- based materials has set a trend in language teaching and learning. This paper discusses how the corpora can be utilised for learning as suggested by various experts. Then it provides a framework about learner corpora and the information that is available to support teachers in understanding the mistakes that are frequently made by students. The paper also presents the list of websites for learner corpora. Finally, the issues involved in using corpora are discussed.

Keywords: online corpora, language learning.

Introduction

Language teaching and learning have been revolutionized by online corpora in computer assisted language learning. In recent years, several studies have been conducted to exploit online corpora materials, and to determine their impact on language teaching and learning. Corpora are collections of authentic data that are stored on the computer for performing language analysis. Corpora are available in various languages, and are accessible over the Internet. At present, some universities have their own corpora websites in their mother tongue as well as in English for the training of speech synthesis, recognition systems and phonetic training. There are different types of corpus such as general corpus, specialised corpus, native corpus, learner corpus. Each corpus varies in its purpose and size. Corpora have been utilized for teaching vocabulary, collocations, concordances, synonyms, phrases and idioms, and contrastive interlanguage analysis. Corpora provide an amalgam of authentic data which can be used and practiced in the classroom. It is an undeniable fact that corpora have contributed a lot to the EFL. Teachers and learners have access to online corpora where they can retrieve authentic examples of vocabulary, word usage and grammatical structure. Corpora offer a large collection of data for English teaching, which can be used and applied in the classroom, as well as by textbook authors and materials developers, or the teachers can develop worksheets and study materials based on corpora materials. (Simon Andrewes, 2012).

Exploiting corpora in the classroom

A corpus is an invaluable source for language learning in addition to its use for dictionary compilation and course book preparation. Using a corpus in the language classroom with some freely downloadable materials can be useful. Since anyone can obtain authentic materials through online corpora which assist them in learning the real language, this could be of immense help to teachers and learners alike. When students are exposed to the language of the native speakers, they learn to imitate and speak like the native speakers. In the classroom, the teacher can give authentic examples used by native speakers instead of giving examples from the text books. Learning the target language with the use of natural language will motivate students to learn a real language rather than an artificial one.

O’Keeffe, et al. (2007) argue, “Corpora can be utilised by instructors as tools for reflective practice and enhance their professional development”(p.220).As cited by Saeed&Waly,a study has proved that corpora enhance students’ language proficiency (Varley, 2008). In spite of its usefulness in the classroom, teachers are still wary about the potential of corpora in teaching complicated characteristics of language such as modals and idioms. Therefore, Neff et al. (2001), Simpson & Mendis (2003), Saeed& Waly ,paid attention on using corpus analysis to trace the problems that non- native speakers of English face in writing and studying instead of using corpora for formulating activities and tasks.

O’Keeffe (2007) points out that, in terms of teaching, numerous studies have shown us that the language presented in textbooks is frequently based on intuitions about how we use language, rather than providing actual evidence of use. While there are often pedagogical reasons for using scripted dialogues, their status as a vehicle for enhancing conversational skills has been challenged in recent years (Carter 1998; Burns 2001; Burns, Joyce and Gollin 2001; Thornbury and Slade 2006; O’Keeffe and McCarthy, Carter 2007).

Corpora are a standard tool for lexicographers to investigate word frequency, patterning and semantics in the compilation of dictionaries. Corpora have revolutionised dictionary making and it ultimately paves way for language teaching materials too (p.21). As cited by Daniel Krieger (2003), Barlow states that corpora can be

employed in three areas, namely syllabus design, materials development, and classroom activities.

Syllabus Design

As mentioned by Kringer (2003), the organization of syllabus depends on the teacher's cognizance of the class with reference to the requirement of students. Information about frequency and register could be conducive in planning course choices. Many online corpora provide data about frequency of words with the statistic information about how often a word is used in the target language and in different contexts. This helps language learners learn the most frequent vocabularies that are in use.

Materials Development

Kringer states by using a corpus, a material developer could generate exercises based on authentic examples which would help students identify features of language use themselves. Further, the material developers could carry out an investigation on the characteristics of language use, or exploit published researches on corpus as a reference guide for designing materials.

Classroom Activities

As stated by Kringer, here, students get hands-on experience by conducting language analyses using concordance program. The teacher can conduct a predetermined investigation which would lead to either predictable results, or it can help students do it on their own, leading to less predictable findings. This ultimately leads to data driven learning, which emphasises learner autonomy by motivating students to discover about language use.

Learner corpora

Learner corpus provides a novel type of data which helps researchers reflect on SLA (Second Language Acquisition) research. It studies the processes that take place in foreign/ second language acquisition, and FLT (Foreign Language Teaching) research, which aims to enhance the learning and teaching of foreign/ second languages (Granger,2002).

Learner corpora are collections of texts or samples of authentic texts produced by foreign or second language learners, stored on the computer. A well-known learner corpus is the International Corpus of Learner English. Sylvian Granger (2002) states, "learner corpora are situated within the non- native varieties of English, which can be broken down into English as an official language (EOL), English as a second Language (ESL), and English as a Foreign Language (EFL)".

Some learner corpora contain exam scripts in which all mistakes and errors made by students are highlighted. This information can be used by researchers and teachers to find which words or structures are difficult for English language learners. Further, it helps teachers to understand mistakes which are commonly made by learners in English language learning, and those which are a result of first language interference.

Lists of Learner corpora

The Cambridge Learner Corpus is the part of the Cambridge International Corpus (CIC) compiled by Cambridge University Press and Cambridge ESOL available at http://www.cambridge.org/elt/corpus/learner_corpus.htm

International Corpus of Learner English (ICLE) is the result of collaboration with a wide range of partner universities internationally, available at <http://cecl.fltr.ucl.ac.be/Cecl-Projects/icle.htm>

Longman Learners' corpus is owned and operated by Pearson Education Limited, available at <http://www.pearsonlongman.com/dictionaries/corpus/learners.html>

Thai English Learner Corpus (TELC) compiled by Assumption University, Thailand, available at <http://iele.au.edu/>

VOICE, an acronym for Vienna- Oxford International Corpus of English, compiled at Vienna University and supported by Oxford University Press. Available at <http://www.univie.ac.at/Anglistik/voice/>

MICUSP (Michigan Corpus of Upper- level Student Papers) launched by the University of Michigan, available at [http://micusp.elicorpora.info./](http://micusp.elicorpora.info/)

Issues involved in using corpora

Daniel Krieger (2003) observed the challenges involved in the application of a corpus in the pedagogical context. The first is the selection of corpus for teaching purposes for which any large corpus will help. Some online corpora are available free of cost or on CD ROM. The teacher has to choose the appropriate corpus according to the particular pedagogical context and representative of the target register. The chief concern is that the corpus must be chosen with the learning objectives of the class and it should serve the learning purpose. Another drawback is that a corpus may give more information about itself than about the language use.

As stated by Michael Post (2008), though the corpus contains samples of authentic language, it does not

give the whole picture of the language that is currently in use. Sample languages are analyzed to find patterns and these patterns almost reflect what have been gathered.

Saeed & Waly from their study found that students are not able to access some corpus software. They advocate that students should be acquainted with freely available online corpora. Another difficulty faced by students in searching a corpus is that students tend to become confused due to the massive amount of search results.

Conclusion

In spite of the prevalence of some issues in using corpora in the classroom, corpora have been recognized, and they are widely used in second language learning. Corpora offer an extremely valuable source of data that can be utilized in designing syllabi and preparing instructional materials for classroom use. Learner corpus proffers empirical data to identify the mistakes that are most commonly made by students. This information can be useful for teachers to identify the error patterns of language learners, and to assist them in language learning. Corpora support discovery learning, which in turn helps dispel fears, and enables student-centred language learning. The judicious use of Corpora in language teaching / learning can be an invaluable tool for language acquisition, particularly of great help in the areas of concordance, vocabulary acquisition, collocation, and data driven learning. It offers a scientific basis for language learning while taking cultural aspects into account.

Works Cited

- Andrews, S. (2012). English Teaching and Corpora. *Modern English Teacher*. 20:3. Accessed online at www.onlineMET.com
- Burns, A. (2001) 'Analysing spoken discourse: implications for TESOL' in Burns, A. And Coffin, C. (eds.) *Analysing English in a Global Context: A Reader*. London: Routledge, 123-148.
- Burns, A., Joyce, H., and Gollin, S. (2001) *'I See What You Mean': Using Spoken Discourse in the Classroom*. Sydney: National Centre for English Language Teaching and Research.
- Carter, R. A. (1998) 'Orders of reality: CANCODE, Communication and Culture', *ELT Journal* 52: 43- 56.
- Granger, S. (2002). "A Bird's Eye View of Learner Corpus Research", in Granger, S., Hung, J. & Petch-Tyson, S. (eds.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam & Philadelphia: Benjamins. Web.
<http://cl.indiana.edu/wiki/LearnerCorpora>
<http://iteslj.org/Articles/Krieger-Corpus.html>
<http://www.birmingham.ac.uk/.../corpuslinguistics/MPostCLessayPatternsOfManufacture.pdf>
- Kringer, Daniel. (2003). *Corpus Linguistics: What It Is and How It Can Be Applied to Teaching*. *The Internet TESL Journal*. Available at
- Neff, J., Martinez, F. & Rica, J. P. (2002). A Contrastive Study of qualification devices in native and non-native argumentative texts in English. Paper Presented at the annual meeting of the *American Association of Applied Linguistics*, February, 2001: St. Louis, MO.
- O' Keeffe, A., McCarthy, M. & Carter, R. (2007). *From Corpus to Classroom: language use and language teaching*. Cambridge: Cambridge university press.
- Post, M. (2008). *Patterns of Manufacture: A Corpus Linguistics Analysis of the methodology used to Disseminate Ideology within a presidential Speech for war*. Module 4 Assignment for Corpus Linguistics. Retrieved from
- Saeed, N. A., & Waly, S. *Corpus for Classrooms: Ideas for Material Design*. Proceedings of the 10th METU ELT Convention. (web)
- Thornbury, S. and Slade, D. (2006) *Conversation: From Description to Pedagogy*. Cambridge: Cambridge University Press.
- Varley, S. (2008). I'll just look that up in the concordance: Integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning*, 22(2), 133-152.

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:

<http://www.iiste.org>

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: <http://www.iiste.org/journals/> All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: <http://www.iiste.org/book/>

Academic conference: <http://www.iiste.org/conference/upcoming-conferences-call-for-paper/>

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digital Library , NewJour, Google Scholar

