Exploring fundamental engineering word list for engineering students: A literature review

Sarimah Shamsudin*, Noraini Husinb,c, Amerrudin Abd. Manand

Language Academy, Universiti Teknologi Malaysia, UTM International Campus, Jalan Semarak, 54100, Kuala Lumpur, Malaysia
Faculty of Education, Universiti Teknologi Malaysia, UTM, Skudai, 81310, Johor, Malaysia
Universiti Teknikal Malaysia Melaka, UTem, Durian Tunggal, 76100, Melaka, Malaysia

Abstract

Corpora have been proven to be useful in investigating language learning issues. It has also been influential in vocabulary learning. Through the years, data-driven vocabulary learning has been recognised as one of the effective means of learning specialised vocabulary. In the context of engineering students, they are required to read specialised textbooks which contain a substantial amount of technical vocabulary or terminology that are necessary in their field of study. Engineering dictionaries unfortunately may not carry all the terms found in engineering textbooks. There seems to be a need to develop a glossary of corpus-based engineering vocabulary or word list to meet the needs of these engineering students. This kind of word list may better assist them in their understanding of the engineering terms found in the textbooks. This paper reviews the literature which is related to corpora leading to the development of word lists and lexical syllabuses for courses meant for the teaching and learning of ESP at higher educational institutions. It then briefly discusses the relevance of developing a word list of engineering terms for beginner engineering students.

Keywords: vocabulary; fundamental; glossary; engineering terms; corpus; word list

1. Introduction

A corpus (plural form: corpora) is “a collection of texts, of the written or spoken word, which is stored and processed on computer for the purpose of linguistic research” (Renouf, 1987, p. 1). Corpora can provide language instructors and analysts with insights into typical features of their linguistic structures

* Corresponding author. Tel.: +603-2615-4479; fax: +603-2615-4212.
E-mail address: sarimah@ic.utm.my, ssarimah.kl@utm.my
and language use (Renouf, 1987). It is also becoming an increasingly useful tool as resources and models on which teaching-learning and training materials are based. Corpora can be very useful in helping learners succeed in language learning. With corpora, teachers’ and students’ roles have changed. Different from their traditional roles, teachers are now the managers of research and students are now true researchers who are in charge of their own learning activities using corpus data as their main source for language learning (Mudraya, 2006). This is called ‘data-driven learning’ (Johns, 1991).

Over the past two decades, corpora have been used widely in linguistic research and also in the teaching and learning of languages (Botley, McEnery & Wilson, 2000; Partington, 1998; Sinclair, 2004; Tribble & Jones, 1990; Thurnstun & Candlin, 1998; Wichmann, Fligelstone, McEnery, & Knowles, 1997). Many have begun to realise the need of having language corpora as the starting point to acquire details about the language to be learned. As Mudaraya (2006) states:

The availability of language corpora to language learners and teachers offers promising opportunities in learning a language, allowing learners to set up and carry out their own language analyses with the help of computer concordancing programs that are aimed at identifying collocations, or word partnerships, in which certain words co-occur in natural text with greater than random frequency. (p. 236)

Currently, there is a wide range of work based on corpus data for students and teachers. Besides that, contributions from researchers and teachers on language learning and teaching using corpus data have also been made which would eventually lead to the enrichment of language learning settings.

2. Corpora and ESP

McEnery and Wilson (2001), identify English for specific purposes (ESP) as a domain-specific area of language learning and teaching where:

Corpora can be used to provide many kinds of domain-specific material for language learning, including quantitative accounts of vocabulary and usage which address the specific needs of students in a particular domain more directly than those taken from more general language corpora. (p. 121)

In other professional-focused domains, a corpus varies. Some have very limited size, but one particular corpora known as monitor corpora will have continuous collections of texts, and the texts are persistently added until the corpora will be too huge to handle and normally will be efficiently removed (Sinclair, 1991, p. 25).

It has been argued that foreign language teachers normally create language learning materials that use simple language for their students which may results in the students experiencing problems in handling authentic examples that use much more complicated language (McEnery & Wilson, 1996). Corpus linguistics, which is a study of language as expressed in samples of “real world” texts (corpora) “can thus contribute to rendering learning a foreign language more effective since students will be faced with real language” (Cotter, 2006, p. 499). As Sinclair (1997) suggests:

It is probably easier to search through a file of relevant examples for what is required than to think up something that sounds natural. When ‘sounds natural’ is examined closely, it usually transpires that it is almost impossible to invent an adequate example; attempts made by language teachers, lexicographers and others to represent usage are often embarrassing and never reliable. (p. 31)

Sinclair’s argument may be a bit deep-seated, but in an ESP context, it is more challenging to produce real life examples since ESP teachers are seldom experts in their learners’ field of study. Furthermore, most of them are non-native speakers of English. Hence, that is why they may experience problems in producing examples from an unfamiliar field of knowledge in a second or foreign language. As for Chung and Nation (2003), they believe that it is important for ESP instructors to enable their students to identify technical terms and understand their meanings. As cited in Cotter (2006), Bogaards (1994) suggests:
Both learners and teachers will then have to rely on the materials used for the ESP course as the main source of new vocabulary. When using textbooks, vocabulary choices are already addressed by the authors, and at times, some do not take into account the specific needs of each group of learners. (p. 500)

Teachers can improve their learners’ proficiency in ESP by producing additional teaching and learning materials based on authentic texts. One of the ways to produce these materials is through the design of custom-made ESP corpora.

According to Pearson (1998), there are a number of factors that has to be taken into account when developing custom-made corpora. These factors include size, text type and origin, authorship, factuality, technicality, audience, intended outcome, setting and topic. Corpora that are tailored for ESP courses should be selected from the students’ specific field of study. The concerns of texts’ authorship and audience are crucial because these two factors may determine the technicality of texts. Concerns such as these have varying degrees of importance. The ESP teachers’ decisions must be made based on the needs of their learners as the usefulness of a corpus depends on what it is intended for (Kennedy, 1998, p. 70).

As corpus linguistics developed, determinations to construct larger English for Science and Technology (EST) corpora which include wider EST domains have been made. For instances, the “Jiaotong Daxue Computer Corpus of Texts in English for Science and Technology” (JDEST), which “comprises 2000 texts of about 500 words each – one million words in all” (Yang, 1985a, pp. 24–25), and the “Student Engineering English Corpus” (SEEC) of close to two million words built to “represent the engineering lexis encountered in English-language textbooks in basic engineering disciplines” (Moudraia, 2004, p. 139). Further, Kuo (2007) compiled a corpus of 1.84 million words consisting of 360 journal articles in ten scientific research fields with 420 word families and created the “Scientific Research Word List” (SRWL). For their part, Shamsudin et al. (2008) created a corpus on civil engineering materials from faculty-assigned texts on the topic.

Among all, the largest EST corpus created is the “Corpus of Professional English” (CPE) which is being developed by the Professional English Research Consortium (PERC), “an association of scholars, educators, publishers, test developers, and education providers committed to research in Professional English (PE) and the development of high-quality Professional English resources, products, and services to meet growing international demands” (http://www.perc21.org/cpe_project/index.html). Currently, it has about 28 million English words compiled from journal articles in the field of science and technology, such as physics, biology, chemistry, engineering, mathematics, and medicine. These journal articles were selected based on the data retrieved from the Journal Citation Reports in 2001 (consisting of about 5 700 journals) that provides “quantifiable statistical data for an objective and systematic approach to determining the relative importance of journals within their subject categories” and they were the top 20% impact factor in their respective field (http://scn.jkn21.com/~percinfo/eng_sub1.html#).

3. Corpus-Based Word Lists

There have been researches conducted which generate word lists from specialised or ESP corpora. Ward (2009) for example, created “Basic Engineering List” (BEL) consisting of 299 words for foundation engineering students from a corpus of engineering textbooks. Ward (2009) mentions:

BEL … represents a relatively easy target for learners whose high school education has not equipped them for the linguistic challenges they face in reading English language textbooks. The list is short and non-technical in nature but gives excellent coverage of a wide variety of engineering textbook material. By concentrating on word types rather than lemmas or families, it encourages learning not only of individual words but also of their lexico-grammatical environments. (p. 180)

Ward (2009, p. 179) further adds that because of its non-technical nature, BEL can be taught “by teachers unfamiliar with technical matters to all types of engineers in the early stages of their university careers, and that it will prove useful to them throughout their studies”.

Pacer: 1277
Various word lists, such as the “Academic Word List” (AWL) (Coxhead, 2000) have been developed for vocabulary learning and assist students in coping with many types of academic texts. Flood and West (1953) devised a “supplementary scientific and technical vocabulary” to West's (1953) “General Service List of English” words. It is a list of definitions of 425 words “expressing all ordinary scientific and technical subjects within the limits of non-specialized study” (West, 1953, p. 583). Knight and Bethune (1972, p. 505) later compiled a list of “science words students know” consisting of 277 words to give the teachers “evidence of the recognition of selected science words by students in grades 7, 8, 9, 10, 11, and 12”. Puangmali (1976) collected between 500-550 words from ten engineering texts (totalling 5,202 words) to compile a 1,148-different-word engineering English vocabulary list.

4. Application of Lexical Approach to Language Teaching

There are many factors that come into play when teachers decide what to teach in a language lesson. Among these are students’ textbooks, supplementary teachers’ handouts and notes and syllabuses we have custom-made for students. The important question that is central to teachers is whether they are teaching their students the most relevant linguistic items actually needed by the students. This is where corpus linguistics becomes important because it helps teachers in making decisions based on the real needs of their students.

According to Barlow (2002), the three areas in which corpus linguistics can be of assistance to teachers are:

- Syllabus design - the document that holds the language focus meant for the students based on their needs.
- Materials development - language teaching materials were often developed based on the teacher’s intuition. Corpora can now assist teachers in the development of teaching materials based on genuine examples which would then aid students in discovering features of actual language use.
- Classroom activities - one of the uses of a corpus is the analysis of linguistic data from concordancing programs. Through these programs language learners can discover and obtain valuable information about language use.

According to Krieger (2003), in corpus-based language studies, the teacher has a new role to play, i.e. that of a research manager as opposed to “an imparter of knowledge” which was most often the case previously. In this type of student-centred language learning, students will actually be involved in discovery learning in which they may learn and acquire authentic language used in real communication settings. Corpus and concordancing programs can encourage students to form generalisations and notice language behaviour which they have not seen before. Though there is no guarantee that students would produce tangible results from studying concordances, it would increase the awareness of students about real language use in real communication settings.

Arguably, the lexical approach to language teaching may carry syllabuses incorporating relatively smaller corpora than others. As Mudraya (2004) stated:

For language learning and teaching, smaller corpora can be more useful, as they are designed to represent the specific part of the language under investigation and are tailored to address the aspects of the language relevant to the needs of the learner. ... Furthermore, smaller corpora are more manageable, allowing easier and faster access to language data. (p. 66)

A number of smaller corpora which consist of technical texts developed for language learning have been mentioned in section 2 above.

All those corpora are normally grounded from selections of textbooks even though they actually differ in design and have different aims. The JDEST for instance was basically designed:
To monitor language teaching materials in order to learn how well the materials which have been developed for the learners of English are representing the authentic materials they are going to read in the future, and also possibly to provide some knowledge on the productivity of different multi-word term patterns. (Yang, 1986, p. 103)

Furthermore, JDEST would hopefully be used for the purpose of analysing the syntax and discourse of EST. (Yang, 1985b, p. 95).

The SEEC, however, had three main objectives: “(a) to establish a representative corpus of Student Engineering lexis; (b) to provide teachers and learners with a word list that could serve as the lexical syllabus foundation of English for Engineering; and (c) to explore the syntactical, morphological, lexical, and discursive features of Engineering English” (Moudraia, 2004, p. 142). BEL created by Ward (2009), is aimed at high school students who are not yet ready to read linguistically challenging textbooks in English. Although BEL contains a short and non-technical list of words, it covers a vast range of textbook materials in the field of engineering. It can therefore be assumed that basic engineering terminology can be taught to engineering students from various field of studies in the first year of their university programme, and that it may be of use for the rest of their studies. Despite their different purposes, all of the above-mentioned corpora have actually initiated the development of vocabulary lists and language teaching and learning syllabuses, especially for ESP and EST courses at the university level.

In the context of our intended research, one of the problems faced by Malaysian engineering students when reading engineering textbooks is the technical vocabulary in those textbooks. At tertiary level education, engineering students are required to read specialised textbooks which contain many specialised or technical vocabulary crucial to their area of the study. Engineering students may be demoralised when there are too many technical words in their specialised textbooks. They may have the tendency to jump to the engineering dictionaries when encountering new engineering terms. Engineering dictionaries unfortunately may not carry all the engineering terms found in engineering textbooks and looking up terms from them may be too time-consuming. A list of corpus-based engineering glossary may assist them to better understand engineering terms. To our knowledge, there is hardly any corpus-based engineering material developed in the context of Malaysia. Hence, there is a need to develop a corpus-based glossary of engineering terms suitable for Malaysian undergraduate engineering students.

5. Conclusion

The use of corpora in language teaching and research appears to be increasingly gaining ground today. The numerous examples of corpora created in the field of science and engineering discussed above are evidence of this. In ESP, corpora can also assist language teachers in tailoring language syllabuses, materials and classroom activities to the needs of their students. Yet another contribution of the language corpora is the construction of specialised word lists to help students in handling content area texts and vocabulary learning. One outcome of the construction of these word lists is the lexical approach to language learning. It was argued that engineering terms from word lists constructed can be taught to engineering students from various fields of studies in the first year of their university programme to help them in their current and future studies. In the context of our research, it is felt that engineering students should also be exposed to terms from engineering corpus. In view of the fact that non-native English speaking engineering students at the tertiary level in Malaysia are required to read specialised textbooks which contain a substantial amount of technical vocabulary that are necessary in their field of study, corpus-based engineering vocabulary or word list seems necessary. Engineering dictionaries unfortunately may not carry all the terms found in engineering textbooks. This kind of word list may better assist them in their understanding of the engineering terms found in their textbooks.
Acknowledgement

We would like to thank the Ministry of Higher Education of Malaysia and Universiti Teknologi Malaysia for funding this research under the Research University Grant Vote Number 04J50.

References

Barlow, M. (2002). Corpora, concordance, and language teaching. In Proceedings of the 2002 KAMALL International Conference. Seoul, Korea.
Bogaards, P. (1994). Le vocabulaire dans l’apprentissage des langues étrangères. Paris: Éditions Didier.
Botley, S. P., McEnery, A. M., & Wilson, A. (2000). (Eds.) Multilingual corpora in teaching and research. Amsterdam: Rodopi B. V.
Chung, T. M. & Nation, P. (2003). Technical vocabulary in specialised texts. Reading in a Foreign Language, 15(2), 103–116. Retrieved from http://nflrc.hawaii.edu/rfl/october2003/chung/chung.html on November 23rd, 2011.
Cotter, M. J. (2006). Teaching terms: A corpus-based approach to terminology in ESP classes. In Proceedings of the 5th International AELFE Conference. Retrieved from http://www.unizar.es/aelfe2006/ALEFEO6/3.%20terminology/69.pdf on December 2nd, 2011.
Coxhead, A. (2000). A new academic word list. TESOL Quarterly, 34(2), 213–238.
Flood, W. E. and West, M. (1953). Supplementary scientific and technical vocabulary. In M. West (Ed.), A general service list of English words with semantic frequencies and a supplementary word-list for the writing of popular science and technology (pp. 583–588). London: Longmans, Green.
Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning materials. ELR Journal, 4, 1–16.
Kennedy, G. (1998). An introduction to corpus linguistics. Essex: Pearson Education Limited.
Knight, D. W. & Bethune, P. (1972). Science words students know. Journal of Reading, 15(7), 504–506.
Krieger, D. (2003). Corpus linguistics: What it is and how it can be applied to teaching. The Internet TESL Journal, 9(3). Retrieved from http://iteslj.org/Articles/Krieger-Corpus.html on January 17th, 2012.
Kuo, Chih-Hua (2007). Constructing a scientific research word list. In JALT CALL Presentation Abstracts (p. 25). Waseda University.
McEnery, A., & Wilson, A. (1996). Corpus linguistics. Edinburgh: Edinburgh University Press.
McEnery, A., & Wilson, A. (2001). Corpus linguistics (2nd ed.). Edinburgh: Edinburgh University Press.
Moudraia, O. (2004). The student engineering English corpus. ICAME Journal, 28, 139–143.
Mudraya, O. V. (2004). Need for data-driven instruction of engineering English. IEEE Transactions on Professional Communication, 47(1), 65–70.
Mudaraya, O. (2006). Engineering English: A lexical frequency instructional model. English for Specific Purposes, 25(2), 235–256.
Partington, A. (1998). Patterns and meanings: Using corpora for English language research and teaching. Amsterdam: John Benjamins.
Pearson, J. (1998). Terms in context. Amsterdam: John Benjamins.
Puangmali, S. (1976). A study of engineering English vocabulary. RELC Journal, 1, 40–52.
Renouf, A. (1987). Corpus development. In J. M. Sinclair (Ed.), Looking up: An account of the COBUILD project in lexical computing (pp. 1-40). London: Collins COBUILD.
Shamsudin, S., Hamzah, M., Wan Mansor, W. F., Azahar, Z. R., & Ismail, M. (2008). Development of ESP courseware via the www: A computational linguistics analysis approach. In S. Shamsudin, & M.
Hamzah (Eds.), ESP in action: The Malaysian context (pp. 33–64). Skudai, Johor: Penerbit UTM Press.

Sinclair, J. (1991). Corpus, concordance, collocation. Oxford: Oxford University Press.

Sinclair, J. (1997). Corpus evidence in language description. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), Teaching and language corpora (pp. 27–39). Harlow: Longman.

Sinclair, J. M. (2004). How to use corpora in language teaching. Amsterdam: John Benjamins.

Thurstun, J., & Candlin, C. (1998). Concordancing and the teaching of the vocabulary of academic English. English for Specific Purposes, 17(3), 267–280.

Tribble, C., & Jones, G. (1990). Concordances in the classroom. London: Longman.

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. English for Specific Purposes, 28(3), 170–182.

West, M. (1953). (Ed.) A general service list of English words with semantic frequencies and a supplementary word-list for the writing of popular science and technology. London: Longmans, Green.

Wichmann, A., Fligelstone, S., McEnery, T., & Knowles, G. (Eds.) (1997). Teaching and language corpora. Harlow: Longman.

Yang, H. (1985a). The JDEST computer corpus of texts in English for science and technology. ICAME News, 9, 24–25.

Yang, H. (1985b). The use of computers in English teaching and research in China. In R. Quirk & H. G. Widdowson (Eds.), English in the world: teaching and learning the language and literature (pp. 86–100). Cambridge: Cambridge University Press.

Yang, H. (1986). A new technique for identifying scientific/technical terms and describing scientific texts. Literary and Linguistic Computing, 1(2), 93–103.