

Linguistic resources to facilitate science education

Uma Maheswari Viswanathan^{1,*},
A. Shahin Sultana¹ and Suhasini Shankar²

¹Department of English, B. S. Abdur Rahman Institute of Science and Technology, Chennai 600 048, India

²Sattva Training Academy, Grenville Terrace, Roxburgh Park, Victoria 3064, Australia

The present study emphasizes the need to mobilize linguistic resources to facilitate science education. In a multilingual country like India, when students opt for science education through English at any level – secondary, higher secondary or tertiary – they face the daunting task of mastering language as well as science. A study of nearly 1.5 million word corpus of science textbooks has revealed the level of lexical difficulty faced by the learners. Hence researchers have created corpus based word lists for each level on the basis of ‘keyness’ and ‘frequency’. These could be powerful pedagogic tools to facilitate science education.

Keywords: Corpus study, linguistic resources, keywords, science education.

AN all-out effort to popularize and promote science education by the government and various scientific bodies of India needs linguistic assistance to yield robust results. The perceived crisis in science education in secondary, higher secondary and tertiary levels is largely due to language crisis in science classrooms. The aim of the present study is to search for options to facilitate science education by explicitly considering the role of vocabulary, keywords analysis and target word lists. The use of corpus linguistics to create useful word lists has become crucial in English for academic purpose and English for special purpose, where word lists are created from a corpus that is closely related to the academic needs of learners. In India, the academic needs of learners are complex due to the absence of a single medium of instruction and uniformity in the boards of education. Though India has the second largest English-speaking group of 200,350,000 people¹ next only to USA, mastering the language used for academic or scientific discourse remains a formidable task for many.

Research in second-language (L2) acquisition has provided empirical evidence to show that vocabulary knowledge is key in dramatically improving L2 proficiency in reading^{2,3}, which is a prerequisite for comprehension. Hence a meaningful lexical analysis was done after building the corpus of science textbooks of secondary, higher secondary and first-year engineering courses. The analy-

sis helped find out the total number of words, keywords, frequency of keywords and the difficulty level of keywords. After harvesting the keywords, target word lists were compiled for each level. The keywords were further compared to find out if there exists a gradation among the three levels. Additionally, the higher secondary English textbooks were analysed to determine whether the English taught as a compulsory subject in schools contributed to the learning of science subjects.

Indian school education has three levels – primary, secondary and higher secondary. Ten-year study of general subjects in secondary level is followed by two years of elective or optional subjects in higher secondary. This is followed by college education. As far as science is considered, till class 10 it is general science which includes physics, chemistry, biology and environmental science. In higher secondary, four subjects and languages are prescribed. Regarding the medium of instruction, according to MHRD⁴ data, 25 languages are used to impart education till class 10 (excluding tribal languages) and 17 languages are used in classes 11 and 12. Besides, there are English-medium schools offering education through English at all levels. The convergence of students happens at different levels. Students either study in English-medium schools throughout, though the percentage is very less, or shift to English-medium schools from vernacular-medium schools for higher secondary education. This is because of the common entrance examination for admission to colleges and college education that is offered through English. There is also a significant number of students who study in vernacular-medium schools till class 12 and enter colleges. Whatever may be the medium of instruction, English is offered as a compulsory subject in schools. Keeping the migration pattern of students to English-medium instruction, the following textbooks were taken for study.

- Science textbooks (National Council of Educational Research and Training (NCERT)) of classes 8–10.
- Science textbooks (NCERT physics and chemistry) of classes 11 and 12.
- Engineering science (physics and chemistry) textbooks of first-year engineering colleges.
- English textbooks of classes 11 and 12.

The total number of words in the corpus was 1,515,054. This excluded the prefaces, formulae, equations, scientific symbols, numbers and appendices. The textbooks available in the NCERT website were downloaded, converted into txt. files and everything else except the words was removed. The initial analysis gave the details of the words counted as families, types and tokens. The manner in which words are counted for keyword analysis is important in vocabulary research. Tokens are individual words. Word repetition is not taken into account while counting the types. Word families include certain lemmas

*For correspondence. (e-mail: maheswaridgp@gmail.com)

Table 1. Word counts in the selected corpora

Academic level	Families	Types	Tokens	Keywords
Secondary	1660	10,331	398,303	670
Higher secondary (physics, chemistry)	6172	15,490	620,852	840
Higher secondary (english)	6002	11,094	111,927	195
Engineering science (physics and chemistry)	5766	17,023	384,928	1,027

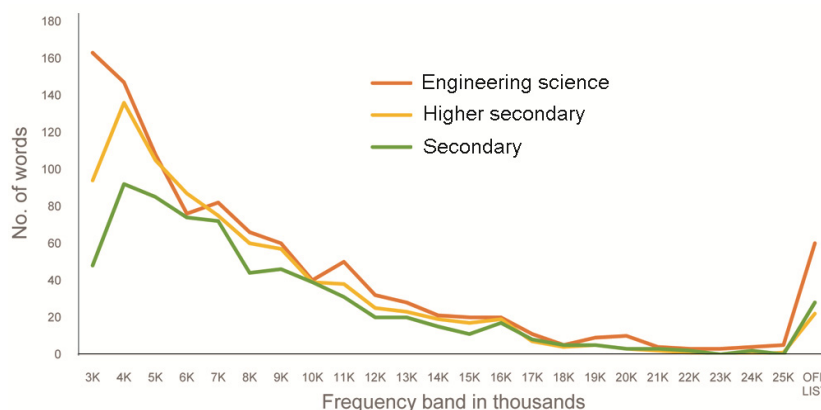


Figure 1. Frequency distribution of keywords in science textbooks of secondary, higher secondary and engineering science courses in India.

and derivatives grouped around a headword. For example, the words play, plays, played, playing, player, playful, playable, unplayable, foreplay and interplay form one family. The following sentence is taken from the science textbook of class 8:

‘This means that the members of a species can reproduce fertile offspring only with the members of their own species and not with members of other species.’

In the above, there are 27 word tokens, 20 types (the repeated words – members, of, the, with, species – are counted as single) and 20 families.

In keywords analysis, only word families are considered as it greatly reduces the number of words in the list. Moreover, it is expected that at the secondary level, students know the basic grammar of formation of adjectives, adverbs, plurals and use of affixes. Keywords are those words that have unusually high frequency in a chosen corpus in comparison to any reference corpus⁵. The British National Corpus – Corpus of Contemporary American English (BNC-COCA) available in lextutor v.8.3 was chosen as the reference corpus. Among the many criteria for choosing keywords, the concept of ‘keyness’ factor was adopted for this study. The software chooses keywords with a keyness factor of 25 and above. For example, the word ‘oscillate’ is found to have a ‘keyness’ factor of 42,987, which means that the word which has 214 occurrences in the selected corpus will have one natural occurrence in the 14 million word reference corpus, if both the corpora were of same size.

$$214 \text{ (word frequency)} / 69,695 \text{ (chosen text corpus)} \\ \times 14,000,000 \text{ (reference corpus).}$$

Table 1 gives details of word counts in terms of families, types, tokens and keywords for each level.

The keywords thus obtained are found to occur either in academic or science registry, and are used only in science classrooms. They are not used or explained in English classrooms as language teaching in Indian classrooms follows the communicative language teaching methodology. Consequently, the lexical difficulty leads to difficulty in understanding the concepts in science. It is common to find science teachers resorting to explanation in first language and the students resorting to rote learning. Rote learning helps in writing the answers, even though the students understand the concept. Science learning thus becomes a burden. The keywords were further analysed to find their places in the word frequency bands.

Words are generally classified as high, mid and low frequency. They are arranged in frequency bands of 1000. The words in 1K and 2K bands are the high-frequency words necessary for general communication. Academic words fall in mid-frequency bands (3K–9K), and technical and domain-specific words fall in low-frequency bands (>9K). The keywords in science textbooks chosen for the present study follow this pattern (Figure 1).

The keywords have relevance only with the corpus chosen. So the number of keywords has only statistical significance. Classification of words under frequency bands shows the difficulty level (Figure 1). The number

and type of words introduced in subsequent levels show the gradation. The number of keywords in secondary level is 670 and in higher secondary level it is 840. It does not mean that 170 words were newly introduced in the next level. Actually, 537 words were introduced in the higher secondary level because only 303 words were found to be common between the secondary and higher secondary keywords lists. Whereas for students who study in vernacular medium schools till class 10, all the 840 words are new. There is another dimension to the problem faced by the vernacular-medium students. There are words other than keywords, like names of the elements, chemical processes, etc. which are not familiar to them in English translation. For example, 'sulphuric acid' is கந்தக அமிலம் (kandaga amilam) in Tamil and copper is তামা (tama) in Bengali. While names such as ozone, aluminium, platinum are not translated in vernacular languages, names like oxygen, carbon dioxide, copper, mercury, etc. are translated. Thus, students are forced to find the meaning of many more words in the textbooks and master the registry within two years and write the exams that would determine their choice of profession. Lack of gradation in the vocabulary introduced in subsequent levels of education compounds the problem. Consequently, rote learning replaces cognitive learning, and science education loses its meaning and purpose.

The possibility of English language learnt in English classes helping the students in reading of core textbooks is also ruled out, as the 111,927 word English corpus and 620,852 word science corpus of higher secondary level share only 4949 words with each other. It is thus evident that communicative English has a limited role in developing proficiency in academic English. It is also a fact that learning is compartmentalized in schools, and teachers of science and English do not want to step into each other's area of instruction. It is probable that the English teachers do not know the concepts in science, and science teachers do not have training in language teaching. Ultimately, it becomes a huge burden for the students of science. Those who study in vernacular medium schools till class 12 face a similar problem in their first year in colleges⁶. While the situation of college students draws the concern of educationists, that of higher secondary students has been overlooked. The present study not only highlights their difficulties in science learning, but also attempts to ease their difficulties by providing them with the keywords harvested from the corpus of their textbooks in the form of word lists for each level.

The word lists thus created can be powerful pedagogic tools as they comprise target words necessary for understanding the textbooks. The corpus is static and the word lists once made based on recent developments in linguistics and sound parameters, can be used in all the Central Board of Secondary Education schools across India by either the English teachers or science teachers. It can be used in content language integrated learning, task-based

learning and classroom activities. It would be more fruitful if it is made part of the curriculum or used in material designing to promote integrated learning. Since language proficiency is paramount in promoting scientific temperament, awareness and interest in science subjects, we intend to share our resources with anyone in the academic community upon request.

1. *Ethnologue: Languages of the World*, 2017, 20th edn; [https://www.ethnologue.com/language/eng/\[Google Scholar\]](https://www.ethnologue.com/language/eng/[Google Scholar]) (accessed on 15 May 2018).
2. Hsueh-Chao, M. H. and Nation, P., Unknown vocabulary density and reading comprehension. *Read. Foreign Lang.*, 2000, **13**(1), 403–430.
3. Laufer, B. and Ravenhorst-Kalovski, G. C., Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Read. Foreign Lang.*, 2010, **22**(1), 15–30.
4. MHRD, Statistics of School Education, Ministry of Human Resource Development, Government of India, 2017; http://mhrd.gov.in/statist?field_statistics_category_tid=33 (accessed on 6 June 2018).
5. Scott, M., PC analysis of keywords – and key keywords. *System*, 1997, **25**(1), 1–13.
6. Viswanathan, U. M. and Sultana, A. S., An Indian ESWL: a pedagogic tool to improve the lexical competency of students in context of horizontal diversity. *Asian EFL J.*, 2018, **20**(4), 23–38.

Received 11 December 2018; accepted 12 October 2019

doi: 10.18520/cs/v118/i2/271-273

Design and development of a low-cost GNSS drifter for rip currents

Surisetty V. V. Arun Kumar^{1,*},
Rakesh Kumar Luhar², Rashmi Sharma¹ and
Raj Kumar¹

¹Earth, Ocean, Atmosphere, Planetary Sciences and Applications Area, and

²Mechanical Engineering Systems Area, Space Applications Centre (ISRO), Ahmedabad 380 015, India

Lagrangian drifters are analogues of particles that are relevant to flow-field characterization and therefore they represent realistic surface currents compared to Eulerian techniques. The use of global navigation satellite system (GNSS) in such drifters with Differential Global Positioning System mode at high frequency (5–10 Hz) sampling and post-processing kinematic results in position estimates with centimeter-level accuracy. In the complex nearshore zone, deploying

*For correspondence. (e-mail: arunkumar@sac.isro.gov.in)