

**Review research paper**

## **GENERAL ENGLISH AND MORE OR LESS TECHNICAL ESP ON THE EXAMPLE OF WORD LISTS**

Zorica Đurović, Milica Vuković-Stamatović

University of Montenegro

**Abstract.** *The renewed interest in vocabulary research, especially due to its meaning-carrying significance in technical Englishes, has cleared the way to the development of corpus linguistics methods based on real-life texts. As the learning/teaching of general English leads towards the learning/teaching of English for a Specific Purpose (in line with the learner's needs), so the creation of general English word lists imposes the creation of technical word lists, aiming to achieve a desired level of the target corpus comprehension. In this paper, we discuss the common grounds and specifics of various levels of general and technical English in terms of their lexical coverage of texts, specifically addressing the example of marine engineering technical manuals.*

**Key words:** *vocabulary, word list, corpus, English for Specific Purposes*

### 1. INTRODUCTION

For many reasons, primarily those of the economic nature, English is considered to be the *lingua franca* of the modern world, the learning of which imposes a growing challenge to both its learners and teachers. Learning and teaching English for a special purpose is thought to be even more challenging, since it relies upon the knowledge of general English, which needs to be upgraded and adjusted to meet the requirements of a specific professional area.

With the development of various scientific and technological fields, there are more and more areas of English for special purposes, branching out into various levels and sublevels. The Hutchinson-Waters's tree (Hutchinson and Waters: 1987) has significantly grown in the meantime, with dozens of the additional ESP branches and sub-branches. Fortunately, the development of the information and computing technologies has also brought us new tools and methodologies, such as corpus linguistics methods and software, which enable easier access to large quantities of real-life material and its multiple-level analysis. This type of analysis is primarily based on vocabulary, i.e. words as the meaning-carrying elements of a language and, at the same time, distinguishing components of its (professional) areas and subareas.

---

Submitted January 16<sup>th</sup>, 2021, accepted for publication March 3<sup>rd</sup>, 2021

*Corresponding author:* Zorica Đurović. University of Montenegro, 2 Cetinjski put, 81000 Podgorica, Montenegro | E-mail: zoricag@ucg.ac.me

## 2. THEORETICAL BACKGROUND

Following the main ideas and pioneering efforts of earlier times, not sooner than in the 60s of the previous century did corpus linguistics become especially recognized and productive with the appearance of huge computer-processible corpora (real-life texts). Those corpora may be comprised of one or more genres. In our case, we have chosen a very technical and (vocabulary-wise) demanding genre of marine engineering – their instruction books and technical manuals.

Furthermore, the technological development has also introduced some of the most modern methodologies to linguistics research, such as lexical profiling, which has flourished for several past decades. The software solutions developed for this very purpose enabled a detailed analysis of texts based on vocabulary frequency ranges, which in turn provided for the creation of specific word lists. The purpose of such specifically extracted vocabulary is to facilitate adequate reading comprehension of (professional) texts. The lists can be used either directly by learners or in the teaching process.

Naturally, the first word lists created were those of general English, aiming to provide the learners with the common vocabulary that occurs frequently across different texts (Brezina and Gablasova: 2013, Nation 2001). As the learning of professional language is based on learning the general one, the general English word lists are also essential for the development of the specialized ones.

### 2.1. General English word lists

Although there are numerous word lists of general English vocabulary, the most influential one to start with is West's General Service List (GSL), comprised of 2,000 word families (headwords with all their inflected and derived forms). The list was obtained from a 5-million word corpus as early as in 1953 and, for now, it still remains the most used word list in this kind of research.

A special contribution to the creation of general English word lists was given by Paul Nation, who derived as many as 25 word lists (available at <http://www.victoria.ac.nz/lals/about/staff/paul-nation>) from the British National Corpus (BNC, available at <http://www.comp.lancs.ac.uk/ucrel/bncfreq/flists.html>) and the Corpus of Contemporary American English (COCA, available at <http://corpus.byu.edu/coca>). This combined corpus comprises 450 million words, obtained from 90% written and 10% spoken texts. Each of the derived word lists contains 1,000 word families and they together provide a good base of common vocabulary, but are also useful for the creation of specific vocabularies. Nation also made four additional lists, which respectively include: the most frequent proper nouns, abbreviations, marginal words and transparent compounds (those written without a hyphen).

Although general service lists are expected to cover a great deal of a written text (78%-98% (Nation and Waring: 1997, Coxhead: 2000, Nation: 2013)), they are not sufficient for an adequate understanding of an academic text or a text related to some narrow professional field. In order to understand a written text, one would have to know 95% (Laufer: 1992) to 98% (Nation: 2000) of its words, or 95% of the words found in the target speech (Coxhead: 2018; Van Zeeland and Schmitt: 2013). For this reason, there are more and more lists of specialized vocabulary, developed in order to make this goal reachable with a most reasonable and efficient learning effort.

## 2.2. Academic word list

Once a learner masters the first 2,000-3,000 general (most frequent) English words, the learning should be oriented towards more specific areas, depending on the learner's needs and goals (Nation: 2000). The next recommended step would be the learning of a specialized vocabulary common to several professional areas, such as academic vocabulary, especially for non-native speakers intending to undergo studies and do research in English. Academic vocabulary would this way serve as a link between general and professional English.

The most influential and widely used academic word list is that of Averil Coxhead, which was created in 2000 and comprises 570 word families. It covers around 10% of academic texts of various scientific disciplines outside the first 2,000 general English words (Coxhead: 2000).

The Academic Word List (AWL) has been validated in a series of studies on text coverage (Hirsh and Nation: 1992; Hwang: 1989; Coxhead: 1998; Nation: 2000), which showed that the list of the most frequent 2,000 words and the AWL cover approximately 86.1% of academic texts, compared to 76.1% without the AWL, while with adding another 1,000 of the most frequent general words instead of the AWL raised this percentage 'only' 4.3% higher (Coxhead, 2000: 222).

Despite the fact that newer academic word lists have been developed since, the Coxhead's AWL is still the most frequently used one in the comparative research of the lexical profiling kind, employed most frequently together with West's GSL, which it was derived upon. For this reason, we use them together in this study as well.

## 2.3. ESP word lists

Specialized word lists can cover a wider scientific area or several professional fields (e.g. Kwary and Artha: 2017) or a very narrow professional subfields or crafts (e.g. Chung and Nation: 2003; Yang: 2015; Coxhead: 2018). Their coverage depends on how they were built – some of them were built 'from scratch', i.e. without pre-excluding any general words, whereas others have been derived by first eliminating the words belonging to the most general vocabulary and sometimes the academic vocabulary as well. The choice of the method of deriving them depends on learner's needs, i.e. on how well they know general (and academic) vocabulary.

In this paper, we test the coverage of two general English word lists (the GSL and Nation's lists), the AWL as a semi specialized word list, as well as two highly specialized, i.e. technical word lists – the engineering word lists derived by Hsu (2014) and Ward (2009), on the example of a corpus belonging to a very specialized area of engineering – marine engineering.

## 3. METHODOLOGY AND CORPUS

The method used for this type of linguistic analysis and research is called Lexical Frequency Profiling (Laufer and Nation: 1995), which provides the measurement of vocabulary levels in a certain type of text. The most updated and recommended software used for this purpose is AntWordProfiler 1.4.0w developed by Laurence Anthony (2014), as an upgraded version of the previously used RANGE programme (Nation and Heatley: 1994).

An additional programme – Familizer+Lemmatizer (Cobb: 2018) is used to expand the word lists available in the form of just the headwords into an all-member format, so that they can be used by the software for the lexical profiling of the text.

For our corpus, we used the texts of instruction books and manuals from a container ship, as one of the most frequent types of vessels in maritime transport of today. The texts were collected and selected with the assistance of the professionals from the area, so that they would include the most representative and contemporary manuals and instruction books for various onboard machinery (main engines, generators, separators, boilers, pumps and various other plants and devices) of significance for some of the most frequently employed ocean-going vessels for carrying containers. As such, they are essential for familiarizing marine engineers with the ship's engine room and machinery, its maintenance and repair. Therefore, their proper reading comprehension is of utmost importance to every competent and responsible engineer onboard. On the other hand, as a highly-technical genre of English for marine engineering purposes, it makes a perfect example for presenting the various types of vocabulary: general, academic, but also various levels of (more or less) technical vocabulary.

To prepare the corpus and convert it into plain text format, we used AntFileConverter (Anthony: 2017), after which we additionally 'cleaned' the text from tables, lists, references and conversion errors, in order to make it as 'clean', i.e. as accurate as possible. These efforts can often be very demanding and time-consuming, but are still very important for the precision of the results obtained (Nation: 2016). The resulting corpus used in this study counts 467,440 running words (tokens) from 22 technical electronic (mostly scanned) files.

#### 4. RESULTS

The first step in lexically profiling our marine engineering corpus was examining the amount of general English vocabulary and academic vocabulary in it, by calculating the coverages of West's GSL and Coxhead's AWL. These two word lists have been used (together) in many studies conducted to date, so we tested our corpus against them for comparative reasons.

Table 1 Coverage of GSL and AWL in container ship's instruction books and manuals

Word lists	Tokens	Coverage (%)
GSL	332,718	71.18
AWL	36,319	7.77
-	98,403	21.05
Total	467,440	100.00

As can be seen from Table 1 and as it could have been expected bearing in mind the extremely technical nature of the corpus, the general vocabulary (as per the GSL) amounts to 71.18% of the words used in the corpus, which is below the usual coverages of 78%-98% reported for various types of written texts (e.g. Nation and Waring, 1997). However, the coverage obtained here fits within the commonly reported range of 70%-71.9% reported for various academic texts (Coxhead: 2000). These results are understandable as the texts examined are meant to be used by educated and trained professionals such as marine engineering officers.

The AWL's coverage, however, is lower (7.77%) than the average of around 10% for most academic texts (Coxhead, 2000).

The results obtained also show that 21.05% of the words remain outside the first 2,000 English words and the list of academic words, i.e. every fifth word of our corpus would be unknown to a reader knowing just the frequent general and academic words, which would make the reading/understanding of the texts very difficult for them.

We also tested the corpus against Nation's word lists, as another (and more contemporary) source of general English vocabulary.

Table 2 Coverage of BNC/COCA lists in the corpus

BNC/COCA word lists	Coverage (%)
2,000 + proper n., abbrev., marginal words and t. compounds	77.26
3,000 + proper n., abbrev., marginal words and t. compounds	85.52
4,000 + proper n., abbrev., marginal words and t. compounds	88.28
5,000 + proper n., abbrev., marginal words and t. compounds	90.47
6,000 + proper n., abbrev., marginal words and t. compounds	91.29
7,000 + proper n., abbrev., marginal words and t. compounds	92.04
8,000 + proper n., abbrev., marginal words and t. compounds	92.65
9,000 + proper n., abbrev., marginal words and t. compounds	93.03
10,000 + proper n., abbrev., marginal words and t. compounds	93.40
11,000 + proper n., abbrev., marginal words and t. compounds	93.57
12,000 + proper n., abbrev., marginal words and t. compounds	93.89
25,000 + proper n., abbrev., marginal words and t. compounds	94.65

This step in the analysis confirms the lexical demand of highly technical ESP texts such as those belonging to our corpus. We can see that the coverage of 95% is not reached even with all the available 25,000 English words (25+4 BNC/COCA lists), not to mention the ideal threshold of 98%, reaching which would be a hardly attainable goal even for a native speaker.

These results clearly call for and justify the creation of specific, technical (ESP) word lists, which would reduce the percentage of unknown words to an acceptable minimum (not more than 5%). This has been additionally proved by analyzing the corpus against the available and applicable engineering word lists. Starting with basic English words, through academic vocabulary, such lists could be the next link towards more specialized lexical areas.

From all the available engineering word lists (Mudraya: 2006; Ward: 2009; Hsu: 2014; Yin et al.: 2013), the two most applicable to our present study were: Ward's word list of basic engineering English (BEEWL) and Hsu's engineering English word list (EEWL).

Taking into account that Ward's target group were students with a rather low proficiency in English, he did not exclude the first 2,000 words from his analysis. Thus, we followed his example for the purpose of comparing our findings with his.

Table 3 Coverage of BEEWL in the corpus of container ship's instruction books and manuals

Word list	Tokens	Coverage (%)
BEEWL (Ward)	64,786	13.86

The coverage of this list in our corpus is somewhat lower than that in Ward's corpus comprising 271,000 words from 25 university textbooks (16.4%). Still, it reaches a significant percentage (13.86%), which could recommend it for improving the vocabulary knowledge in marine engineering students, but only as an aid or a basis which needs to be upgraded with specific technical vocabulary for this special field, i.e. a more specific word list.

The other relevant word list for engineering would be the Hsu's list (2014), here used with the first 2,000 English words since it was derived that way (meaning that there are no overlaps between the two).

Table 4. Coverage of EEWL in the corpus of container ship's instruction books and manuals

Word list	Tokens	Coverage (%)
BNC/COCA 2,000	340,968	72.92
EEWL (Hsu)	41,108	10.08
Total	398,076	83.00

As we can see from Table 4, the coverage of Hsu's list in our corpus is 10.08%, which is significantly below its coverage in Hsu's corpus of textbooks from 20 engineering fields (14.3%), including marine engineering. This, again, does not come as a surprise, bearing in mind that textbooks are more narrative in nature than technical manuals and instruction books, and that the latter are used not by students (at least not on a regular basis) but by active professionals. In favour of this speaks one of Hsu's findings (2014) which points to the fact that marine engineering and biochemistry are, vocabulary-wise, the most demanding engineering areas.

## 5. CONCLUSION

Our intention in this paper was to show the relationship between general English and ESP on the example of word lists, by employing one of the most contemporary methods of lexical research – that of lexical profiling. Thanks to the modern software solutions used today in linguistic research, we were able to provide a detailed and accurate lexical profile of our target corpus – a carefully selected, highly technical material of great relevance in English for marine engineering purposes, specifically related to container ships as one of the most frequently vessels employed by the modern shipping industry.

This kind of material enabled us to calculate the presence of various levels of vocabulary, from general, through academic, to various areas and levels of engineering vocabulary, using the resources available for this kind of research based on the corpus linguistics methodology. The results point to the need for knowing the basic English vocabulary (e.g. first 2,000 (most frequent) English words), the advantage of learning the basic academic vocabulary (AWL), and how useful the existing general engineering word lists (Ward's BEEWL and Hsu's EEWL) are when applied to a specific engineering field (marine engineering for container ships). Furthermore, the results clearly call for the creation of a specific marine engineering word list, which one of the authors of this paper is already working on.

In addition to showing the practical usefulness and the need for specialized word lists for each ESP or any of its subfields, by the way they are created and the process of gradual learning/teaching itself, the results presented here point to a tight intertwining

and interdependence of all vocabulary levels of English, from the most basic to the most technical ones. Moreover, they also reflect the importance and advantage of sharing the results and findings among the researchers, to the benefit of all.

#### REFERENCES

- Anthony, L. AntWordProfiler, Build 1.4.1.0., Center for English Language Education in Science and Engineering, School of Engineering, Waseda University, Tokyo, Japan, 2014.
- Anthony, L. AntFileConverter (Version 1.2.1) [Computer Software], Tokyo, Japan, Waseda University, 2017. Available from <https://www.laurenceanthony.net/software>
- Brezina, V. and D. Gablasova. "Is there a core general vocabulary? Introducing the new general service list", *Applied Linguistics*, 36 (1), 2013: 1-22
- Cobb, T. "From corpus to CALL: The use of technology in teaching and learning formulaic language", in A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.) *Understanding Formulaic Language: A Second Language Acquisition Perspective* 2018: 192–211.
- Coxhead, A. "An Academic Word List", *Occasional Publication*, No. 18, LALS, Victoria University of Wellington, New Zealand, 1998.
- Coxhead, A. "A New Academic Word List", *TESOL* 34(2), 2000: 213–238.
- Coxhead, A. *Vocabulary and English for Specific Purposes research: Quantitative and qualitative perspectives*. London: Routledge, 2018.
- Hirsh, D. and I. S. P. Nation. "What vocabulary size is needed to read unsimplified texts for pleasure?", *Reading in a Foreign Language*, 1992: 689-696.
- Hsu, W. "Measuring the Vocabulary Load of Engineering Textbooks for EFL Undergraduates", *English for Specific Purposes* 33, 2014: 54–65.
- Hutchinson, T. And Waters, A. *English for Specific Purposes: A Learning-centred Approach*, Cambridge: Cambridge University Press, 1987.
- Hwang, K. *Reading newspapers for the improvement of vocabulary and reading skills*, Unpublished MA thesis, Victoria University of Wellington, 1989.
- Kwary, D. A. and A. F. Artha. "The Academic Article Word List of social science", *MEXTESOL Journal*, 41(4), 2017: 1-11.
- Laufer, B. "What percentage of text-lexis is essential for comprehension?" In *Special Language: From Humans Thinking to Thinking Machines*, Ed. Lauren. Ch. And M. Nordman, Multilingual Matters, 1992: 316–323.
- Laufer B. & Nation, P. "Vocabulary Size and Use: Lexical Richness in L2 Written Production", *Applied Linguistics*, 16(3), 1995: 307–322.
- Mudraya, O. "Engineering English: A lexical frequency instruction model", *English for Specific Purposes* 25, 2006: 235-256.
- Nation, I. S. P. and Heatley, A. *Range: A program for the analysis of vocabulary in texts (software)*, 1994, retrieved from <http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx>.
- Nation, I. S. P. and R. Waring 1997. "Vocabulary size, text coverage and word lists", *Vocabulary: description, acquisition and pedagogy* 14, 1997: 6-19.
- Nation, I.S.P. "Review of What's in a Word? Vocabulary Development in Multilingual Classrooms" by N. McWilliam. *Studies in Second Language Acquisition* 22, 1, 2000: 126-127.

- Nation, I. S. P. *Learning vocabulary in another language*, Cambridge: Cambridge University Press, 2001.
- Nation, I.S.P. "How Large a Vocabulary Is Needed for Reading and Listening?" *Canadian Modern Language Review*, 63 (1), 2006: 59–82.
- Nation, I. S. P. The BNC/COCA word family lists, 2012. <http://www.victoria.ac.nz/lals/about/staff/paul-nation> (accessed January 2019).
- Nation, I.S.P. *Learning Vocabulary in Another Language*, 2nd edition, Cambridge: Cambridge University Press, 2013.
- Nation, I. S. P. *Making and using word list for language learning and testing*, Amsterdam: John Benjamins, 2016.
- Ward, J. "A basic engineering English word list for less proficient foundation engineering undergraduates", *English for Specific Purposes*, 28(3), 2009: 170-182.
- Van Zeeland, H. and N. Schmitt. "Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension?", *Applied Linguistics*, 34(4), 2013: 457-479
- Waring, R. & Nation, I. S. P. "Vocabulary size, text coverage, and word lists", in *Vocabulary: description, acquisition and pedagogy*, N. Schmitt and M. McCarthy (eds.), Cambridge: Cambridge University Press, 1997.
- West, M. *A General Service List of English Words*. London: Longman, Green and Co, 1953.
- Yin. J. N. et al. "Development of the Engineering Technology Word List for Vocational Schools", *Malaysia International Journal of Education and Research*, vol. 1, no. 1, 2013: 43-59.