

SUPPLIER	Main engine and operating media	3	81,196
	Oil purifier	2	34,040
	Sea water desalinator	1	6,786
	Propulsor	1	5,313
	Safety instructions, environment	1	1,727
	Lubrication system	1	5,310
	Seal system	1	8,560
	Grease trap	1	6,622
	Generator sets	1	28,297
	Bilge water separator	1	8,271
	TOTAL	13	522,690
TOTAL	83	2,279,078	

"The painstaking process" (Nation 2016: 224) of the additional cleaning of material (in relation to tables, references, brands, typos and the like) was applied to the best of the author's abilities, considering the huge amount of the material, originally found mostly in scanned formats. The prepared Corpus of Ship Instruction Books (CSIB) was finalized with 1,769,821 running words (tokens). Bearing in mind that we have a very technical and discipline-specific genre in question, we may say that our corpus is of representative size and content, so as to guarantee the validity of results and conclusions produced.

4. Methodology

The initial method we used for our research is called Lexical Frequency Profiling (Laufer and Nation 1995). It provides authors and language teachers with the opportunity to analyse corpora in terms of vocabulary types and quantities and to test the coverage of available word lists, as well as to create new ones. The software that has recently been found most useful and convenient 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).

As a starting point, we used the results of previous research on the same corpus (Đurović, Vuković Stamatović and Vukičević 2021) in terms of coverage by the existing general service and engineering word lists, aiming to prove the lexical demand of the corpus, as well as the need to build a specific marine engineering word list (MEWL). For the latter purpose, the same programme (AntWordProfiler 1.4.0w) was used for several reasons. First, it has proved itself to be one of the most appropriate programmes for building specialized vocabulary lists (or glossaries or dictionaries) since, in contrast to some other programmes used for building vocabulary lists, it can exclude the most frequent general vocabulary from further analysis, which is estimated to be known by LSP learners. Another reason is the value of having comparable results and being able to make reference to relevant findings and word lists built upon the same methodology. Finally, the Anthony's software is readily

available, quite simple to use for any language and offered free of charge.

In setting the cut-off point in the frequency count, we were guided by the goal of reaching the desired level of reading comprehension (of ship instruction books), which anticipated a minimum of 95% of known vocabulary. With a list which was estimated to be of a size to represent an attainable task for language learners, taking into account the available time for learning, the reached level of 95% corpus coverage (with the most frequent GE words and the MEWL) would at the same time be its positive evaluation result (Dang and Webb 2016: 133).

For formatting the lists into headwords only, lemmas or expanding them into all-family-members form, we used the Familizer + Lemmatizer programme (Cobb 2018). For corpus preparation and converting it into "plain text" format, as required by the software, we used AntFileConverter (Anthony 2017).

Additionally, for detecting the concordances of a certain word within the corpus, the AntConc software (Anthony 2019) can be used. Owing to this kind of software programmes, each word from our list can be checked for concordances and n-grams to be included in a glossary or dictionary.

4.1 Previous research findings

In examining the target corpus, we first analysed its vocabulary load and types in terms of reaching the desired reading comprehension. Aiming to provide precise answers as to the quantity and types of vocabulary needed for the purpose and following the methodologies used by recognized authors in the field, these questions were the focus of our previous research (Đurović, Vuković Stamatović and Vukičević 2021). As such, we here briefly present the lexical profiling methodology and the results, since they clearly point to the need for a specialized marine engineering word list and provide solid justification for our current research.

In particular, in testing our corpus against the General Service List (West 1953) and the Academic Word List (Coxhead 2000), which are usually applied together in this kind of research, the cumulative coverage amounted to 79.46% (71.39% + 8.07%, respectively), which is lower than the average of 86.1% found for academic texts (Nation 2000: 27). Considering general English only, the coverage level of 71.39% is significantly below the usual coverage of 78–98%, as reported for various types of written texts (e.g. Nation and Waring 1997). This generally means that, with knowing only 2,000 first English words, even with the most common academic words, about every fifth word of the text remains unknown (20.54%), which would make both reading and understanding the text very difficult.

When calculating the total amount of general English words needed for adequate comprehension, using the more contemporary Nation's word lists extracted from the huge corpora of the British National Corpus and the Corpus of Contemporary American (BNC/COCA, available at <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists>), a coverage of 95%

was reached not sooner than with the 12,000 most frequent English words, whilst 98% coverage is not reachable even with all the available 25,000 English words, including the additional four lists of proper nouns, abbreviations, transparent compounds and marginal words. This means that a clear understanding of this kind of text is almost unattainable, even for a native speaker, without an engineering background.

In taking into consideration the available word list from adjacent fields and genres, the only two available lists which we found to be relevant and appropriate to this kind of research were Ward's Basic Engineering English Word List (BEEWL) (Ward 2009), and Hsu's Engineering English Word List (EEWL) (Hsu 2014). Both lists were extracted from undergraduate textbooks appropriate to various engineering fields. In addition, both of them showed lower coverage in our target corpus (13.53% compared to 16.4%, and 10.11% compared to 14.3%, respectively). These results are also understandable considering the very technical nature of our target corpus, when compared to that of textbooks, which are more narrative in nature.

Bearing all these findings in mind, we sought to create a specialized marine engineering word list, aiming to reach the adequate reading comprehension level for marine engineering instruction books. Taking into consideration that this kind of genre is abundant in both diagrams and tables, as well as supplemental explanations and abbreviations accompanying these materials, we opted to place the adequate reading comprehension threshold at 95%.

5. The Marine Engineering Word List

Carefully examining the best way to apply the methodology stated above in our research so as to be relevant, efficient and justified, we took several points into account. Firstly, our target learners are either university-level students or active seafarers undertaking professional training. Furthermore, bearing in mind the fact that English is learned in many countries from an early age, we deemed that it would be reasonable to expect them to be competent in reading and understanding at least 3,000 basic English words (BNC/COCA), especially considering that it refers to the receptive knowledge and not necessarily productive language skills. This follows certain recent trends in expanding the high-frequency list from 2,000 (West 1953; Nation 2001) to the first 3,000 English words (Schmitt and Schmitt 2014; Nation 2016). Additionally, adequate English proficiency is mostly required and tested by shipping companies during their recruitment procedures.

Finally, since in many countries, officers' training courses do not need to be organized through university courses, but can be organised through any certified training centre, and taking into account the fact that this type of text is more technical than academic (although the AWL list coverage of 8.07% is not to be overlooked), we decided to head directly to the marine engineering word

list (MEWL) for obtaining early specialization (Coxhead and Hirsch 2007) in English for Marine Engineering Purposes (EMEP).

6. Data analysis and results

In determining the frequency threshold for our list, we took into account the final objective of 95% coverage, the relevant lists to test against (and those to exclude from the count) and the reasonable list size. In the end, we opted for a frequency of at least 50, so the final list was formed of 337 headwords (see Addendum 1) with 73 transparent compounds (see Addendum 2). We did not apply the range and dispersion criteria here, since they are more relevant to huge corpora and extracting common vocabulary for various professional areas (e.g. Coxhead and Hirsch 2007) than one specific professional profile (e.g. Coxhead and Demecheleer 2018). In our research, the corpus comprises a variety of technical materials in terms of different machinery and ship's systems, all nearly equally important to the single occupation of marine engineers.

Aiming to obtain the most precise results possible, in analysing the corpus, we extracted the most frequent abbreviations and marginal words (the latter mostly typos and conversion errors), and added them to Nation's original lists, respectively. This selection process is never easy or perfect, since, for example, some abbreviations are recognizable (e.g. *cyl* for *cylinder*, *hfo* for *heavy fuel oil*), but sometimes it is difficult to distinguish between an abbreviation and a typo or conversion error. Fortunately, they were not numerous, owing to the carefully conducted cleaning process of the corpus, and, once added to one of the additional Nation's list, cumulatively, they do not affect the final results, apart from enabling the production of a "purer" word list.

In addition, in spite of our efforts to initially remove as many proper names as possible, so as to avoid the commerciality of the data, what remained was also added to the Nation's list of proper nouns. Taking into account that all the words considered above are expected to be easily understood and recognized from the context, thus not bearing a learning burden of significance, we excluded them from our further frequency analysis together with the most frequent GE words.

After having a close look at the Nation's list compounds, we deemed it too abundant for our target learners and significantly more difficult to learn compared to the most frequent general English words, thus we did not exclude it from our initial count. On the contrary, we decided to make a separate list of transparent compounds (Addendum 2) derived from the corpus itself. In that, we followed the recommendation of having a separate list of this type of vocabulary (Nation 2016: 70). Some of the items overlap with those from the Nation's list (e.g. *setpoint*, *standstill*) and the rest originate from the corpus (e.g. *sootblower*, *crankthrow*). Hyphenated forms are not frequent in this type of text, and they were initially eliminated by dividing their constituents into sepa-

rate words. In addition, those compounds that could not be easily understood according to their constituents, were left as part of the initial list, since they cannot be deemed transparent (e.g. *bulkhead*). Here we must note that, although presented separately, they should be taken as one with the initial list of words, since they are of equal significance to both vocabulary skills and course designs and materials.

The final results are presented in Table 2.

Table 2: Coverage of MEWL in the CMEIB

Word lists	Tokens	Coverage (%)
BNC/COCA 3,000 + proper nouns, abbreviations and marginal words	1,547,067	87.41%
Transparent compounds	12,806	0.72%
MEWL without compounds	130,994	7.41%
—	78,954	4.46%
Total	1,769,821	100%

As we can see, the coverage of the first 3,000 most frequent English words (with the additional three lists, as usually excluded first in the analysis) amounts to 87.41%. If we compare this to other findings, we can see that the general English coverage is lower in our target corpus than, for example, in business research articles, where 3,000 words with proper nouns cover as much as 90.84% of the material, in business textbooks, where the coverage is 94.15% (Hsu 2014: 251), in popular science books where 3,000 words with the additional 4 Nation's lists cover 92.65% (Vuković Stamatović 2020) or in various types of texts where 3,000 word coverage is of the order of 90–93% (Nation 2006). The coverage is lower even compared to the coverage of the first 2,000-word BNC/COCA lists in some types of texts, which ranges from about 79% (in various academic text) to 89% in school journals and novels (Nation 2006, 2016; Fraser 2007). This clearly points to the complexity of our target vocabulary in terms of its technicality. It also speaks loudly in favour of the need for a specialized word list to accommodate adequate coverage of the target genre.

For the purpose of illustration, we can use another option of the programme to see the corpus text coloured as per the word lists (Figure 1), where (in the extract presented), for example, pink stands for the first 3,000 BNC/COCA words, orange for MEWL, blue for the abbreviations, while the words outside the list are left black.

1.2.1 Oil and atomizing systems

The atomizing medium and oil are supplied from valves through flexible hoses to the burner lance.

The automatic shut-off valves, re-circulation valve, solenoid valves, and the rest of the

necessary burner mountings are all fixed on the wind-box. All shut-off valves are

standard ball valves, and the automatic valves are with electric/pneumatic actuators.

The arrangement of the components on the burner is made as simple as possible and

is therefore very user friendly regarding operation and maintenance.

The oil flow control valve is fitted together with a pneumatic converter/positioner. To

obtain a very accurate regulation of the oil flow, which is very important to have a large turn-down ratio, the valve is also fitted with an adjustable

Cv-adjuster. This Cv-adjuster makes it possible to adjust the pressure loss coefficient of the valve whereby the oil flow regulation can be optimised to

each single installation. The oil flow is measured by a screw-type flow meter.

Fig. 1: Level lists presented in colours

We can also present MEWL with compounds only using the same option, where green marks the MEWL words and transparent compounds are given in red, while the remaining are left black.

Proximity sensors per cylinder, i.e. one for each guideplane are placed on brackets typically mounted on the Exhaust side structure below both guide shoes.

The system monitors the variation in distance between sensor and guide shoe bottom surface, in bottom dead centre BDC. The measurements represent the relative wear in all three crank-train bearings, simultaneously.

. Seawater Cooling System

Fig. 2: MEWL and compounds' list presented in colours

This option could also be used for text glossing, which is often helpful in terms of reading technically demanding texts such as ours (Nation 2013). In this way, in addition to the glossaries frequently added to the textbooks, the learning materials can be compiled out of authentic texts (whether adjusted or not) with

detected and glossed technical terms, followed by a definition, translation, or similar.

The coverage of the "pure" marine engineering word list of 337 word families is 7.41%, i.e. 8.13% adding the list of 73 transparent compounds. The purity of the final list is attained by additional analysis of the list and detecting additional members to add to the Nation's original 3,000 families (e.g. *purifier*, *cleanable*, *abnormal*). Moreover, during the conversions, some technical words are not recognized, thus additional attention should be paid to "unclassified" words, such as e.g. *bunker* (*bunkering*), *alignment* (*misalignment*), and similar terms, which were subsequently added to the list.

In total, together with the first 3,000 GE words, proper nouns, abbreviations and marginal words, a level of 95.54% was reached, thus fulfilling our goal of attaining the adequate reading comprehension threshold. Taking into consideration that the desired level could be attained with not less than 12,000 general English words only, our final results perfectly fit the findings of Laufer and Ravenhorst-Kalovski (2010), by which the threshold of 95% is expected to be reached through the use of 4,000–5,000 word families.

The remaining share of the unknown vocabulary in our corpus would here form below 5% (4.46%), which means that the adequate reading comprehension threshold is comfortably reached. Adding the introductions, professional knowledge and abundance of schema and diagrams, the understanding of this kind of professional text should therefore no longer present any significant difficulties. However, an aid such as a glossary or dictionary for instruction books would certainly be welcomed by both (future) marine engineers and teachers of this challenging branch of ESP.

7. Glossaries, dictionaries and pedagogical implications

Having a ready-made list of the most frequent vocabulary extracted from marine engineering instruction books, the course material designers can then choose whether to use them as additional material, compile new textbooks and exercises with excerpts and examples taken from authentic instruction books, or else use the list in any other way they find suitable. Given the size of the list, it might form a convenient and achievable task for (a part of) university studies, where about 800 new words are considered a practical learning goal for each two years of study (Dang and Webb 2106: 174). Moreover, additional corpus linguistics methods can be introduced to students as well, and thus they could be required to find concordances, n-grams or full examples for the word use from the corpus itself.

In addition, the list would be a more than useful tool for English courses held at seafarer training centres, since the attendees would generally have some experience of using onboard instruction books. Furthermore, any marine engineer would be able to make good use of having such a concentrated list of specialized vocabulary at hand during their onboard service. The list, therefore,

could help in creating a monolingual or bilingual glossary, with or without pronunciation transcription.

Maritime English textbooks, like those of many other areas of ESPs, are often supplemented with a glossary, frequently a bilingual one (e.g. Spinčić and Pritchard 2009: 180-242) which covers the textbook syllabus. For this purpose, it would be sensible to use authentic texts, such as excerpts from various instruction books, and the word list might be divided into sub-lists or learning chunks in order to suit the organization of classes and exams. Frequently, textbooks cover more than one semester of studies or are printed in several volumes, which can be covered by a common glossary. Additionally, the list might also be added to existing glossaries to cover combined class material.

Considering stand-alone glossaries, which are frequently found online as useful links provided by various maritime organizations, institutions and companies, we first noticed the scarcity of specific marine engineering glossaries, in contrast to nautical or (general) maritime ones. The most of maritime glossaries are generally very simple in form. They most frequently offer just the target term or phrase with a simple context explanation (monolingual) or translation (bilingual), occasionally including additional details such as the phonetic transcription (sometimes offering both British and American variants) or phrases and collocations.

The extraction of the most frequent words is also applicable to a more comprehensive and demanding project, such as a marine engineering dictionary, whether that be creating a new one or adding to an existing one to make it more contemporary and comprehensive. Almost any dictionary has been built upon a corpus (Nation 2016: 176), and using the possibility of measuring the (lemmatized) vocabulary frequency range within it would certainly be an advantage in this regard (De Schryver and Nabirye 2018). Compared to word lists, which are intended to be learned for the purpose of unaided reading, the cut-off point for dictionaries can be moved to lower frequencies, so as to be more inclusive in terms of technical (low-frequency) vocabulary. A further advantage of the proposed methodology for building a technical glossary or dictionary is that, upon compiling a corpus of adequate composition and size, the building of the macrostructure would be solidly founded upon the frequency of the vocabulary beyond the most frequent GE words, the lists of which can be excluded from further analysis.

This kind of dictionary might also be supplemented by illustrations (e.g. Carić 2011) and/or examples from the corpus. This might prove especially useful, since using dictionaries to explore aspects of target words has been shown to be one of the most efficient ways of learning about vocabulary within the language awareness strand of learning strategies (Hirsch and Coxhead 2009; Macalister and Nation 2011). Here we must note that a technical glossary or dictionary would mostly not contain general English meanings and examples. For example, the word *pinion* would be here referred to as a type of gearwheel. However, when deemed necessary to include both, the notation GE (general

English) or ME (marine engineering) would follow (e.g. *average* (GE) and *average* (ME)). More considerations on polysemous words in marine engineering are given in Section 8. Limitations of the study.

The lists of most common abbreviations should be added to the dictionary, which might also include more relevant details, such as the marked part of speech and further examples, or perhaps illustrations. The words can be followed by the frequency mark, such as, for example, in Macmillan's Dictionary (2012), bearing in mind the specific corpus it was derived from.

For the purpose of generality and following the organization of various dictionaries, we offer a monolingual simplified example for the most frequent word (appearing 10,871 times) in the instruction books:

- (1) **valve** /vælv/ n. – a device that opens and closes to control the flow of a liquid or gas

Following the main term and definition, AntConc (Anthony 2019) (or another programme offering a similar function) provides an excellent opportunity to check the concordances of certain terms from the corpus. Noun phrases and certain collocations containing the target term usually follow the head definition.

- (2) **valve** /vælv/ n. –
by-pass valve –
check valve –
flow control valve –
motor-oil flow valve –
pressure valve –
return valve –
safety valve –
steam valve –
stop valve –
oil sample valve – , and so on.

To avoid headword repetition in multi-word units, it is sometimes replaced with a sign such as ~ (e.g. Rapovac 2002; Spinčić and Pritchard 2009).

- (3) **valve** /vælv/ n. –
by-pass ~ –
check ~ –
flow control ~ –, and so on.

With noun phrases where the term in question is not the head noun, these examples often come first, preceding those as in Example (2).

- (4) **valve** /vælv/ n. –
valve block –
valve body –

valve bottom –
valve bypass –
valve components –
valve control –

Again, these can be written as:

- (5) **valve** /vælv/ n. –
~ block –
~ body –
~ bottom –, and so on.

In addition to collocations, dictionaries often contain authentic examples of the use of the words within phrases and sentences, which can easily be found using the same programme. For example:

- (6) by-pass valve – ex. Adjust the oil pressure to a suitable level on the by-pass valve.

Comparing the means of presentation, word families are generally used for word lists for practical reasons, whilst in practice they are expanded, usually to include all family members when used for the programme analyses presented above. For dictionaries (and glossaries) it is more convenient and useful to use lemmas, so that different parts of speech can be presented as separate items (for example *actuate v.*, *actuation n.*, *actuator n.*). For this purpose, again Familizer + Lemmatizer (Cobb 2018) might be used. For example, the single word family *bear* comprises 12 family members (word forms), whilst, for instance, in a bilingual maritime dictionary (Rapovac 2002) we can find the meanings and examples of *bear*, *bearer* and *bearing*, in various collocations and contexts. Based on the proposed software being used, we should note here that the selection should rely upon the frequency of tokens converted directly into lemmas (and accompanied by compounds and phrases) in order to reduce the interference of the software in the selection process during the list expansion and reduction processes.

Further details on the organization of dictionaries vary according to their intended purposes and the choices made by the authors. Compared to the painstaking endeavour of finding relevant examples and the collection and processing of abundant material manually, such as, for example, was the case with the quotation slips used for the first Oxford English Dictionary, the idea of this paper was to present some of the available IT-supported methodologies of corpus linguistics, which make the process incomparably easier and faster, but also much more accurate, thus providing legitimacy to the use of this system.

8. Limitations of the study

Despite the best efforts of the author to apply the most accurate methodology

and make the best possible and reasonable decisions along the way, we accept that neither the methodology nor the results are ever perfect (Nation 2016: 182). For example, the conversion of large instruction books, some running to as many as over 800 pages of scanned material, resulted in many typos and conversion errors. This added to the words outside the lists, whilst a number of those should add to the frequencies within them.

The programme itself does not recognize different spelling options (e.g. *authorized* and *authorised*), and thus counts them as separate words with separate frequencies, where the frequency should in fact be cumulative. Moreover, multi-word units are always a point of issue in this kind of research (Nation 2016). For example, some of them can be written separately (e.g. *cam shaft*), and some as a single word (*camshaft*), which makes the statistics in terms of frequency not entirely precise. Additionally, careful attention should be paid to the words "unclassified" by the programmes, which are often the "missing" members of the general English word families or unrecognized technical words. In particular, general programmes such as Familizer + Lemmatizer do not recognize certain words from the marine engineering register, such as, e.g. *crosshead* or *bedplate*, and thus we had to retrieve many of them from the "unclassified" word types and manually add them to the lists used or produced by the programme.

When it comes to such a technical branch of ESP as marine engineering English is, additional attention should be paid to semantic issues. In particular, a certain portion of the words could be classified as belonging to Step 3 of a four-step rating scale for technical vocabulary (Chung and Nation 2003: 104), being referred to as polysemous or cryptotechnical words (Fraser 2009). This means that some of the words classified as belonging to the most frequent general English words often have a completely new and technical meaning in marine engineering contexts, either individually (e.g. *wear*, *draught*, *average*) or in collocations (e.g. *jacket water cooling system*, *guide shoe*). In the subsequent table, we present some examples from the first 3,000 GE words which have different meanings in marine engineering English, either individually, or in collocations. The definitions are somewhat shortened and simplified for practical reasons of presentation.

Table 3: Polysemous words from the first 3,000 GE (BNC/COCA) words

The first 1,000 (BNC/COCA, all family members)	General English	English for Marine Engineering Purposes
<i>actuate</i> v. <i>actuator</i> n.	(added to the word family <i>actual</i>)	start, move mover, starter
<i>rocker</i> n. <i>arm</i> n.	someone who plays or likes rock music one of the two long parts of our body ending with a hand	<i>rocker arm</i> — a rocking lever of the valve opening mechanism

<i>bearing n.</i>	the particular way in which someone stands, moves, or behaves	a machine element that constrains relative motion to only the desired motion
<i>excite v.</i>	make someone feel happy and enthusiastic	to start an electric motor
<i>film n.</i>	movie	a thin (oil) layer
<i>ground v.</i>	earth; touch the ground	touch seabed with the bottom of the ship
<i>mean adj.</i>	evil	average value
<i>mess n.</i>	disorder	<i>mess room</i> — ship's dining room
<i>room n.</i>	space	
<i>running n.</i>	moving quickly; controlling a business or organization	engine operation
<i>shoe n.</i>	something you wear on your feet	guide shoe — a piece of a crosshead guide
<i>skirt n.</i>	a piece of female clothing	cylindrical walls of a piston
The second 1,000 (BNC/COCA, all family members)	General English	English for Marine Engineering Purposes
<i>arrestor, arrester n.</i>	(added to the family <i>arrest</i>)	a device preventing the unwanted transmission of electricity
<i>average n.</i>	the mean	damage to the ship
<i>breather n.</i>	(added to the family <i>breathe</i>)	ventilation opening or (relief) valve
<i>guide n.</i>	an informative book or person guiding you	a device that directs the motion of a machinery part
<i>nut n.</i>	a type of food	a round hollow metal piece used with screws
<i>pin n.</i>	a small thin piece of metal with a sharp end	<i>gudgeon (piston) pin</i> — a round metal piece connecting the piston to the piston rod <i>crank pin</i> — a round metal piece for crankshaft connection
<i>stress n.</i>	a worried or nervous feeling	physical pressure on a material or vessel (hull)
<i>trunk n.</i>	the main part of a tree the braches grow out of	trunk (piston engine) — an engine without a crosshead
<i>jacket n.</i>	a piece of clothing	a protective or insulating cover around the cylinder
The third 1,000 (BNC/COCA, all family members)	General English	English for Marine Engineering Purposes
<i>draught n.</i>	cold air that blows into a room	the distance from the water line to the bottom of the vessel
<i>survey n.</i>	a set of questions designed to explore public opinion	a technical inspection
<i>surveyor n.</i>	a person asking a set of questions to explore public opinion	a person conducting technical inspections and examinations

Here again, we have to mention the conditionality of classification, since many of the technical terms are shared among various fields, in this case, especially with general and other branches of engineering. Overlaps are not a rarity in the produced word lists, which was the reason that some authors have decided to build common core vocabulary word lists, such as the Academic Word List (Coxhead 2000), the Science-Specific Word List (Coxhead and Hirsch 2007), and other similar lists.

When it comes to bilingual formats, things get even more complex, depending on the lexicographical features of the local language in relation to the maritime sector, which, most usually anticipates the colloquial use of numerous Anglicisms and words borrowed from Italian or Spanish, for example. For all the reasons mentioned above, this kind of analysis cannot be left to pure statistical processing; it is evident that human (expert) intervention of close account and expertise is in fact required every step of the way.

Considering the corpus itself, for university students, for example, it can be broadened by the use of textbooks and possibly scientific articles, although the latter can make a separate academic corpus to examine in terms of lexis. The reason for this could be the contents of the list we produced. We can easily spot words from outside the marine engineering field, but which are related to, for example, physics (e.g. *amplify*), medicine (e.g. *diaphragm*), electrical engineering (e.g. *coil*), and so on. Therefore, the list as is could be named more precisely as "the word list of ship instruction books", and new and additional marine engineering word lists might be developed from a wider or different corpus.

In order to overcome the possible shortcomings of the overall process, we invested our best efforts in following and taking into consideration various previous findings and recommendations. Among those, we tried to explain the application of the methodology and the decisions made along the way. This provides justification and makes the results clearer, but might also help other authors in building relevant specialized lists of other technical areas and, indeed, other ESP learners.

9. Conclusion

Analysing the most practical language needs of our target language learners — in this case future and active marine engineers following English for Marine Engineering Purposes courses, we embarked on the ambitious project of collecting, selecting and analysing their key corpus of marine engineering instruction books in terms of both vocabulary types and the overall lexical burden. Following certain previous research findings, we applied the lexical profiling methodology and some of the most updated software for the creation of a specialized marine engineering word list. The final list comprises 337 word families plus 73 compounds and cumulatively covers 8.13% of the target corpus. Together with the 3,000 most frequent English words (BNC/COCA) and an

additional and broadened lists of proper nouns, marginal words and abbreviations, we succeeded in reaching the desired coverage level of 95%, more precisely, 95.54%. This means that with a knowledge of the 3,000 most frequent English words and familiarization with a range of the most frequent proper names and abbreviations, with MEWL, less than 5 out of 100 words of this demanding type of technical material should remain unfamiliar to the reader, a percentage which should not significantly affect the comprehension of the text. Compared to the about 12,000 general English words initially needed to achieve a similar result, the amount of vocabulary required is incomparably reduced, and as such, the list we have produced should certainly provide for early specialization in English for Marine Engineering Purposes.

We have also put additional efforts into explaining the methodology and its application, and providing explanations and justifications for the decisions made during the (replicable) process and the final cut-off points, in order to make our own modest contribution to future research in the field and the enhancement of the available methodology. Moreover, we sought to provide practical suggestions for building (marine engineering) glossaries and dictionaries based upon a similar methodology, which will, hopefully, be a matter of interest for other ESP teachers and lexicographers, just as it is for the author of this work.

References

- Anthony, L.** 2014. *AntWordProfiler* (Version 1.4.1.0) [Computer Software]. Tokyo, Japan: Waseda University. Available at: <https://www.laurenceanthony.net/software>.
- Anthony, L.** 2017. *AntFileConverter* (Version 1.2.1) [Computer Software]. Tokyo, Japan: Waseda University. Available at: <https://www.laurenceanthony.net/software>.
- Anthony, L.** 2019. *AntConc* (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available at: <https://www.laurenceanthony.net/software>.
- Carić, T.** 2011. *Englesko crnogorski pomorski slikovni rječnik*. Tivat: Fakultet za mediteranske poslovne studije.
- Carver, R.P.** 1994. Percentage of Unknown Vocabulary Words in Text as a Function of the Relative Difficulty of the Text: Implications for Instruction. *Journal of Reading Behavior* 26(4): 413-437.
- Chung, T.M. and I.S.P. Nation.** 2003. Technical Vocabulary in Specialised Texts. *Reading in a Foreign Language* 15(2): 103-116.
- Cobb, T.** 2018. From Corpus to CALL: The Use of Technology in Teaching and Learning Formulaic Language. Siyanova-Chanturia, A. and A. Pellicer-Sanchez (Eds.). 2018. *Understanding Formulaic Language: A Second Language Acquisition Perspective*: 192-211. New York: Taylor & Francis.
- Coxhead, A.** 2000. A New Academic Word List. *TESOL Quarterly* 34(2): 213-238.
- Coxhead, A. and M. Demecheleer.** 2018. Investigating the Technical Vocabulary of Plumbing. *English for Specific Purposes* 51: 84-97.
- Coxhead, A. and D. Hirsch.** 2007. A Pilot Science-Specific Word List. *Revue Française de Linguistique Appliquée* 12(2): 65-78.

- Dang, T.N.Y. and S. Webb.** 2016. Making an Essential Word List for Beginners. Nation, I.S.P. (Ed.). 2016. *Making and Using Word Lists for Language Learning and Testing*: 153-167. Amsterdam: John Benjamins.
- De Schryver, G.-M. and M. Nabirye.** 2018. Corpus-driven Bantu Lexicography. Part 2: Lemmatization and Rulers for Lusoga. *Lexikos* 28: 79-111.
- Đurović, Z., Vuković Stamatović, M. and M. Vukičević.** 2021. How Much and What Kind of Vocabulary do Marine Engineers Need for Adequate Comprehension of Ship Instruction Books and Manuals? *Círculo de lingüística aplicada a la comunicación* 88 (in print).
- Folse, K.** 2004. *Vocabulary Myths: Applying Second Language Research to Classroom Teaching*. Ann Arbor, MI: University of Michigan Press ELT.
- Fraser, S.** 2007. *Providing ESP Learners with the Vocabulary they Need: Corpora and the Creation of Specialized Word Lists*. Retrieved from: https://ir.lib.hiroshima-u.ac.jp/files/public/2/24812/20141016150057516129/h-gaikokugokenkyu_10_127.pdf.
- Fraser, S.** 2009. Breaking Down the Divisions Between General, Academic, and Technical Vocabulary: The Establishment of a Single, Discipline-based Word List for ESP Learners. *Hiroshima Studies in Language and Language Education* 12: 151-167.
- Gao, J. and H. Liu.** 2020. Valency Dictionaries and Chinese Vocabulary Acquisition for Foreign Learners. *Lexikos* 30: 111-142. <http://lexikos.journals.ac.za>; <https://doi.org/10.5788/30-1-1548>.
- Hirsch, D. and A. Coxhead.** 2009. Ten Ways of Focussing on Science-specific Vocabulary in EAP Classrooms [online]. *English Australia Journal* 25(1): 5-16. Available at: <https://search.informit.org/doi/10.3316/informit.473250020621629>.
- Hsu, W.** 2014. Measuring the Vocabulary Load of Engineering Textbooks for EFL Undergraduates. *English for Specific Purposes* 33: 54-65.
- Hu, M. and I.S.P. Nation.** 2000. Unknown Vocabulary Density and Reading Comprehension. *Reading in a Foreign Language* 13(1): 403-430.
- Kurnia, N.** 2003. *Retention of Multi-word Strings and Meaning Derivation from L2 Reading*. Unpublished Doctoral Dissertation. Wellington: Victoria University of Wellington, New Zealand.
- Laufer, B.** 1989. What Percentage of Text-lexis is Essential for Comprehension? Lauren, C. and M. Nordman (Eds.). 1989. *Special Language: From Humans Thinking to Thinking Machines*: 316-323. Clevedon: Multilingual Matters.
- Laufer, B. and I.S.P. Nation.** 1995. Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics* 16(3): 307-322.
- Laufer, B. and G.C. Ravenhorst-Kalovski.** 2010. Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a Foreign Language* 22(1): 15-30.
- Macalister, J. and I.S.P. Nation.** 2011. *Case Studies in Language Curriculum Design: Concepts and Approaches in Action Around the World*. New York: Routledge.
- Maritime English*. 2015. Model Course 3.17. London: International Maritime Organization.
- Nation, I.S.P.** 2000. Review of *What's in a Word? Vocabulary Development in Multilingual Classrooms* by N. McWilliam. 2012. *Studies in Second Language Acquisition* 22(1): 126-127.
- Nation, I.S.P.** 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I.S.P.** 2006. How Large a Vocabulary is Needed for Reading and Listening? *Canadian Modern Language Review* 63(1): 59-82.

- Nation, I.S.P.** 2012. *The BNC/COCA Word Family Lists*. Available at: <http://www.victoria.ac.nz/lals/about/staff/paul-nation> (accessed January 2019).
- Nation, I.S.P.** 2013. *Learning Vocabulary in Another Language*. Second edition. Cambridge: Cambridge University Press.
- Nation, I.S.P.** 2016. *Making and Using Word Lists for Language Learning and Testing*. Amsterdam: John Benjamins.
- Nation, I.S.P. and A. Heatley.** 1994. *Range: A Program for the Analysis of Vocabulary in Texts (Software)*. Retrieved from: <http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx> (accessed January 2019).
- Nation, I.S.P. and R. Waring.** 1997. Vocabulary Size, Text Coverage, and Word Lists. Schmitt, N. and M. McCarthy (Eds.). 1997. *Vocabulary: Description, Acquisition and Pedagogy*: 6-19. Cambridge: Cambridge University Press.
- Rapovac, D.** 2002. *Pomorski rječnik, englesko-srpski*. Kotor: Fakultet za pomorstvo.
- Rundell, M.** 2012. *Macmillan English Dictionary for Advanced Learners*. 2012. Second edition. Oxford: Macmillan Education.
- Schmitt N. and D. Schmitt.** 2014. A Reassessment of Frequency and Vocabulary Size in L2 Vocabulary Teaching. *Language Teaching* 47(4): 484-503. doi:10.1017/S0261444812000018.
- Spincić, A. and B. Pritchard.** 2009. *An English Textbook for Marine Engineers II*. 4th edition. Rijeka: Pomorski fakultet.
- Twaddell, F.** 1973. Vocabulary Expansion in the TESOL Classroom. *TESOL Quarterly* 7(1): 61-78. doi:10.2307/3585510.
- Van Zeeland, H. and N. Schmitt.** 2013. Lexical Coverage in L1 and L2 Listening Comprehension: The Same or Different from Reading Comprehension? *Applied Linguistics* 34(4): 457-479.
- Vuković Stamatović, M.** 2020. Vocabulary Suitability of Popular Science Books for English for Science Classes: A Corpus-Based Study. Čubrović, Biljana (Ed.). 2020. *BELLS90 Proceedings: Belgrade English Language and Literature Studies. International Conference to Mark the 90th Anniversary of the English Department, Faculty of Philology, University of Belgrade. Volume 1*: 439-447. Belgrade: Faculty of Philology, University of Belgrade.
- 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. *A General Service List of English Words*. London: Longman, Green & Co.

Addendum 1: The headwords of the marine engineering word list (without transparent compounds)

aboard	bracket	crank	erosion
abrasive	brass	cylinder	evaporate
accessory	bronze	debris	ex
accord	buffer	decant	expire
acid	bulkhead	default	fasten
actuate	bunker	defect	fax
acute	cabin	deflect	fecal
adhesive	calibrate	deform	fixture
adjacent	calorific	detergent	flange
adsorb	cam	deteriorate	flap
aft	cartridge	deviate	flotation
align	casing	diagnosis	flue
alkaline	caterpillar	diagram	fluid
alternate	caution	dial	flush
aluminium	cavity	diameter	fore
ambience	centrifuge	diaphragm	foul
amplify	centripetal	diesel	friction
analog	certify	differential	furnace
annex	chlorine	digit	fuse
annular	clamp	din	gasket
anode	classification	dip	gauge
anti	clog	discard	generator
appendix	clutch	discrete	geometry
appliance	cock	disinfect	gland
arrow	coil	dismantle	globe
ash	collar	dispense	glove
assemble	combustion	displace	grease
astern	communicable	dissolve	grind
automate	compartment	distillate	groove
automobile	compatible	drip	gudgeon
auxiliary	comply	droop	halogen
axis	compress	dual	hammer
barge	con	duct	harness
barrel	condense	durable	hexagon
batch	cone	duration	hoist
bellow	configure	dynamic	hood
bilge	console	effluent	horizontal
blade	contaminate	eject	hose
blink	copper	electrode	hub
bolt	corrode	emulsion	hull
bonnet	crane	enclose	humid

hydraulic	nominal	rinse	sulphur
icon	nozzle	rod	sump
identical	offset	rotate	surge
idle	optimum	rubber	surveillance
ignite	opus	rudder	synthetic
impel	orifice	rust	tab
impulse	oxidation	saline	tag
incinerate	oxygen	sanitize	telescope
inert	parameter	satisfactory	terminal
inhibit	particle	sauer	thermal
insulate	paste	scavenge	thermostat
intact	permissible	scrape	thread
integral	pinion	scrubber	threshold
intercept	piston	seizure	throttle
interface	pliers	sensor	thrust
interval	plunge	serial	tiller
jacked	pneumatic	servo	tilt
kit	polyamide	shaft	tolerance
knob	polymer	silicon	torch
lance	potable	sketch	torque
lever	precaution	sleeve	torsion
linear	preface	sludge	toxic
linen	preliminary	socket	transducer
liner	prescribe	sodium	turbine
litre	preset	soiled	tween
loop	prolong	solenoid	usage
lube	propel	solvent	utensil
lubricate	propulsion	soot	vacuum
manifold	proximity	spa	valve
manoeuvre	pulley	span	vane
manometer	pulse	spanner	velocity
membrane	puncture	spark	ventilate
mesh	purge	specimen	verify
micro	quarantine	spindle	vertical
mineral	rack	spiral	vibrate
minimise	radial	splash	viscous
mist	ram	spool	volt
moisture	ramp	stack	warranty
molybdenum	recreation	static	weld
mount	refract	steer	wrench
multi	relay	stool	yoke
needle	residue	stud	
nipple	resilience	stuffed	
node	rim	suction	

Addendum 2: The headwords of marine engineering transparent compounds

backflow	eyebolts	outlet	shellfish
bedplate	feedback	overboard	shipbuilding
breakdown	flywheel	overflow	shutdown
bypass	freshwater	overhaul	sootblower
camshaft	gastroenteritis	overheating	standby
changeover	gastrointestinal	overlay	standstill
clockwise	handwashing	overload	thermometer
contouch	handwheel	override	throughput
copyright	inlet	overspeed	touchscreen
countdown	intake	overview	troubleshooting
crankcase	interlock	password	turbocharger
crankpin	intermediate	pipelines	upstream
crankshaft	keyboard	potentiometer	uptake
crankthrow	layout	screwdriver	upward
crosshead	login	seawater	warewashing
crosswise	millimeters	selfjector	whirlpool
datalogger	nameplate	servomotor	
dipstick	onboard	setpoint	
downstream	outbreak	setup	