http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

2. Theoretical background

2.1 English for Maritime Purposes

The term Maritime English embraces a range of distinctive registers utilised in onboard and international communications (Cole et al. 2007). As a specialised branch of English for Specific Purposes (ESP), its specific communicative features, genres and vocabulary used by the vast maritime discourse community have been explored by many Maritime English (ME) researchers. According to Čulić-Viskota and Kalebota (2013) Maritime English is characterized by its own particular jargon with specific grammatical forms. Bocanegra-Valle (2013) states that English for Navigation and English for Marine Engineering are among the five subcategories of ME, emphasizing that English for Marine Engineering stands out as the most technical of all the varieties. ME vocabulary covers the vital language of written documents dealing with ship construction, engineering, surveillance, maintenance and the operation of ships. The proper interpretation of such a specialised vocabulary requires extra-linguistic knowledge or, more simply, knowledge of the field (Borucinsky and Kegalj 2019). With this in mind, Đurović (2021) extracted and analysed a marine engineering frequency word list from the corpus of marine engineering instruction books and manuals. This article aims to do the same for the nautical database and present comparative results regarding the nautical vs. marine engineering corpora in terms of their respective lexis content.

2.2 Maritime dictionaries and corpus-based headword selection for technical (sub-)areas

The first lexicographic activities regarding seafaring, one of the oldest professions that have connected people all around the world, can be traced back to the thirteen century in the forms of nautical glossaries and sea-related nomenclature (cf. Pritchard 2013). *The Seaman's Dictionary* by Henry Mainwaring (1644) represents one of the oldest English dictionaries of Maritime English (ibid.). The first dictionaries were written for the sake of naming basic concepts regarding ship's construction and concepts related to the sea (ibid.). Over the centuries, maritime dictionaries and subject-specific glossaries appeared under a number of titles such as: a sailor's dictionary, a dictionary of sea terms, a nautical dictionary, a dictionary of seafaring, or a marine/maritime dictionary (ibid.). With the rise of maritime activities and maritime nations, maritime monolingual and bilingual dictionaries have expanded and embraced a variety of maritime-related concepts regarding technology, trade, law, ship's business, and so on. As was, and still is generally the case in specialized lexicography, a variety of methods have been applied to gathering and presenting specialized vocabulary and terms, with corpus linguistics approach being slower than in

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

general lexicography (Bowker 2010).

Corpus linguistics methods are not a novelty in lexicography, emerging first with the *Dictionary of the English Language* by Samuel Johnson from the 18th century, followed by the *American Webster Dictionary* in the nineteenth, and then the first *Oxford English Dictionary*, all the way up to modern paper and online dictionaries (Cambridge, MacMillan, Oxford, Longman and others). The first and most salient contributions of corpus tools in lexicography were made in the context of monolingual dictionaries such as the *Collins COBUILD English Language Dictionary* (Sinclair 1992; Abdelzaher 2022).

Today, this way of compiling corpora is considered conventional and very helpful for creating macrostructures of lexicographical materials. In the present study, we partly deal with the methodology that is most conveniently used and evaluated with numerous word lists tailored for LSP/ESP courses and vocabulary materials. However, bearing in mind that maritime communications are generally prone to restrictiveness and the practical needs of concise message transmissions, we consider the methodology generally applicable in building any lexicographical material, where anything from reference vocabulary lists to comprehensive dictionaries can be acknowledged as a lexicographic attempt to isolate a distinctive register, as Opitz (cf. Čulić-Viskota and Rummel 2022) advocates in terms of segmental dictionaries. Therefore, the aim is to bring the basic word-list building methodology to an extended, more compound and higher level of application in designing technical (segmental) lexicographical products.

In addition, corpus selection also provides for retrieving illustrative contextual examples and cross-referencing among sub-corpora. The metalanguage built this way is very convenient for technical dictionaries, given their dense and often complicated nominal groups, which can often be ambiguous when they are not placed in a wider context (Borucinsky and Kegalj 2019; Čulić-Viskota and Rummel 2022).

2.3 Frequency word lists and keywords

For practical reasons, especially for a quick and practical response to the professional needs of ESP learners, a new software-based methodology was developed in the second half of the previous century, providing the statistically-justified extraction of frequency-based headwords. The first frequency word lists, ranging from simple ones to composite lexicographical publications, were those created for General English (GE). They provide (English) language learners with access to the most frequent English vocabulary found in various types of texts. Following this psycholinguistic perspective in building a native-speaker mental lexicon (Abdelzaher 2022: 168), the extraction of the most frequent General English words has been applied to a variety of ESPs. Contemporary software solutions such as RANGE (Nation and Heatley 1994) and its upgraded version

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

AntWordProfiler (https://www.laurenceanthony.net/software/antwordprofiler/) enable the elimination of the most frequent GE vocabulary from further analysis, thus focusing directly on the most frequent technical vocabulary. In this way, more and more ESP word lists have been generated, offering the core technical vocabulary of specific ESP texts and genres.

Although today it is generally considered outdated, one of the most influential GE word lists in regard to the recorded lexical analyses is still West's General Service List (GSL), dating back to 1953. It comprises 2,000 word families[1] and was obtained from an English corpus of 5 million words. It is often used together with the Academic Word List (AWL) of Coxhead (2000) which built upon the GSL (to avoid overlapping) and which consists of 570 word families extracted from various academic texts. These lists were (and still are) usually used together to test a target corpus and measure the GE and academic vocabulary load, providing the opportunity to compare this with other scholarly findings grounded in the same methodology.

A notable contribution to the GE word lists was provided by Nation (2012), who elicited 25 GE word lists from the textual compilation of the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). In this way, 25 BNC/COCA lists of 1,000 GE word families were offered, with an additional four lists of the most frequent proper nouns, abbreviations, marginal words and transparent compounds, respectively.[2] These lists have been used to measure the coverage of GE vocabulary in various corpora.

In addition to the statistically most frequent vocabulary, it has become evident that certain words are very specific to a distinctive type of text, and the knowledge of these words facilitates understanding of the texts (Baker 2004; Al-Rawi 2017). These words of 'special status' (Stubbs 2010: 21) are considered key vocabulary. In statistical lexical analysis, they are obtained according to their frequency compared to a reference corpus, which is most often a GE one.

Considering the specific nature of the maritime lexicon, for both nautical and marine engineering, we will use both criteria, i.e. frequency and the key nature of the vocabulary, to analyse the vocabulary types and loads of both target corpora (see also Đurović 2023). In addition, we will analyse and compare them in order to provide solid and measurable answers to our research questions.

3. Research questions

The three research questions posed in this study are:

1. How many high-frequency, general-purpose and academic tokens are detected in nautical and marine engineering corpora?

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

2. How many headwords need to be mastered for an adequate reading comprehension of nautical and marine engineering corpora?
3. What does the detection of keywords and metadata reveal in terms of lexicographic similarities and differences of the two corpora referring to two maritime subdomains?

4. Corpus and research methodology

This section of the article will provide more specific detail about the corpus and the methodology used to answer the research questions given above.

4.1 Corpus details

The number of maritime communication spheres is becoming increasingly diverse, given the continuing globalisation and internalisation of the navigating profession. Maritime and marine-related literature comprises a myriad of academic, professional and technical publications covering different branches of the seafaring profession. For this study, we have deliberately chosen to analyse corpora comprised of nautical and marine engineering professional books.

Aware of the dynamic changes characteristic of the shipping industry and emerging concepts in the maritime profession, our choice was to cover the chronological range of books dating from 1990 to 2021. In the selection of the texts, we started with various conventional concepts in navigation (good seamanship, manoeuvring, vessel position and maintenance) and marine engineering (marine propulsion, engine operation and auxiliaries). Nevertheless, we also included some novel maritime- and marine-related concepts linked to smart applications, automation, propulsion systems, electronic navigation and 'ultimate concepts' in shipboard operations. In addition, we took into consideration the professional genres and the narrative character of the publications from the two professional areas. For certain other types of publications, such as marine engineering instruction books, it would be difficult to find a comparable counterpart in nautical publications.

Finally, we identified the eight selected books, available in electronic form, from the field of nautical studies, and our eight marine engineering books (Table 1 and Table 2). Totalling 768,135 and 813,429 running words (tokens) for the respective tables, the corpora can be considered both sufficient and relevant in size, as well as convenient for further comparative analysis.

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

Table 1: Corpus of Nautical Books (CONB)

No.	Book title	No. of types	No. of tokens
1	Laugier, C. and R. Chatila (Eds.). 2007. *Autonomous Navigation in Dynamic Environments.* STAR Springer Tracts in Advanced Robotics 35. Berlin: Springer.	5,194	57,656
2	Manley, P. 2008. *Practical Navigation for the Modern Boat Owner*. San Francisco: John Wiley & Sons.	3,203	33,052
3	Weintrit, A. and T. Neumann (Eds.). 2013. *Marine Navigation and Safety of Sea Transportation*. London: Taylor & Francis.	9,641	126,451
4	Dixon, C. and J.K. Spencer. 2021. *The Ocean: The Ultimate Handbook of Nautical Knowledge*. San Francisco: Chronicle Books.	10,861	90,743
5	Touche, F. 2005. *Wilderness Navigation Handbook*. Canada: Friesens.	4,344	56,649
6	Tetley, L. and D. Calcutt. 2001. *Electronic Navigation Systems*. 3rd edition. London: Routledge.	6,567	127,842
7	Grewal, M.S., L.R. Weill and A.P. Andrews. 2007. *Global Positioning Systems, Inertial Navigation, and Integration*. 2nd edition. Hoboken, New Jersey: Wiley-Interscience.	8,586	160,249
8	Barrass, C.B. and D.R. Derrett. 2012. *Ship Stability for Masters and Mates*. 7th edition. Oxford: Butterworth-Heinemann.	5,055	115,493
	Total	**53,451**	**768,135**

Table 2: Corpus of Marine Engineering Books (COMEB)

No.	Book title	No. of types	No. of tokens
1	Watson, G.O. 1990. *Marine Electrical Practice*. 6th edition. London: Butterworth-Heinemann.	6,266	108,200
2	Taylor, D.A. 1996. *Introduction to Marine Engineering*. 2nd edition. London: Butterworth-Heinemann.	5,680	92,124
3	Jackson, L. 2001. *Reed's General Engineering Knowledge for Marine Engineers*. 4th edition. London: Thomas Reed.	7,187	114,036
4	Tsinker, G.P. 1995. *Marine Structures Engineering: Specialized Applications*. Dordrecht: Springer Science+Business Media.	9,721	153,542
5	Martelli, M. 2014. *Marine Propulsion Simulation*. Warsaw/Berlin: De Gruyter.	3,061	48,759

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

6	Kantharia, R. 2013. *Marine Engineer's Handbook — A Resource Guide to Marine Engineering*. (e-Book) Marine Insight official website.	1,417	6,860
7	Hobart, H.M. 1911. *The Electric Propulsion of Ships*. London/New York: Harper and Brothers.	4,149	54,895
8	Woodyard, D. 2004. *Pounder's Marine Diesel Engines and Gas Turbines*. 8th edition. Oxford: Butterworth-Heinemann.	8,537	235,013
	Total	**46,018**	**813,429**

4.2 Research methodology

The initial method applied is known as Lexical Frequency Profiling (Laufer and Nation 1995), used for measuring vocabulary types and levels in a corpus. For this purpose, we used the latest AntWordProfiler software version 2.0.1, which is largely an upgrade of the previously used RANGE programme (Nation and Heatley 1994), developed for the lexical analysis of texts. The word lists used for testing coverage with general and academic vocabulary in our target corpora were the General Service List (West 1953) and the Academic Word List (Coxhead 2000), as they are generally used together in this kind of analysis, and for the purpose of comparison.

The same software was used to test the corpus coverage by General English word lists. Following the same reason of comparability, for measuring the coverage level of General English vocabulary, we used the Nation's (2012) BNC/COCA word lists developed from the British National Corpus and Corpus of Contemporary American English.

For keyness analysis, version 4.1.0 of the AntConc software (https://www.laurenceanthony.net/software.html) was used. As a referent GE corpus for keyness analysis, we used the Freiburg version of the Lancaster-Oslo-Bergen (FLOB) corpus. The rationale behind the selected referent GE corpus was its size, as it is close to the target corpora, as well as the fact that it was developed as a British counterpart to the Brown GE corpus of American English.

For preparing the corpora for further software analysis, we used AntFileConverter 2.0.2 by the same author (https://www.laurenceanthony.net/software/antfileconverter/), since the referent files need to be uploaded in plain text format.

In addition, the lists are generally available as headwords (word families), and thus were expanded to all family members using the Familizer + Lemmatizer programme (Cobb 2018). Special attention should here be paid to the headword entries, such as lemmas or word families, whereas the final selection should always be dictated by both user needs and expert advice (Atkins 2008; Đurović 2021).

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

In choosing the most appropriate methods, we were led by the most practical and illustrative results that can be obtained and that were tested by utilising various software options and settings.

5. Findings

5.1 Research Question 1

Aiming to test the vocabulary load and types in the target corpora, we used West's General Service List (West 1953) and the Academic Word List of Coxhead (2000) as mentioned in Section 5.2. We tested the corpora (CONB and COMEB) individually and the comparative results are presented in Table 3.

Table 3: Coverage of GSL and AWL in the CONB and COMEB

Word lists	Tokens CONB	Tokens COMEB	Coverage CONB (%)	Coverage COMEB (%)
GSL	548,287	600,580	71.38	73.83
AWL	69,468	63,289	9.04	7.78
Not in the lists	150,380	149,560	19.58	18.39
Total	**768,135**	**813,429**	**100.00**	**100.00**

The coverage of the Nautical corpus by GE vocabulary is 71.38%, which is similar to the GSL coverage measured in Marine Engineering instruction books (Đurović et al. 2021), but, as can be seen from Table 3, this is still somewhat higher in Marine Engineering general books (73.83%). As expected, considering the technical and specific nature of maritime publications, the GE coverage (GSL) is below the general coverage of 78–98% expected to be found in various types of written text (Nation and Waring 1997). Interestingly, it is slightly higher than in various academic texts, where it reached the level of 70–71.9% (Coxhead 2000).

On the other hand, the coverage of the Academic Word List solely does not reach the average of about 10% which is measured in research articles and textbooks (Coxhead 2000) or in Medical texts (Chen and Ge 2007), or the values of 11.17% in Applied Linguistics (Vongpumivitch et al. 2009), 9.96% for Chemistry texts (Valipouri and Nassaji 2013) and 9.47% for Pharmacy related material (Fraser 2007). However, the values reached are still higher than the 8.07% found in Marine Engineering instruction books (Đurović et al. 2021). Bearing all the figures in mind, as well as the genre in use, the given corpora can be considered an academic type of text to some extent, but more evidently it fits the genre of technical literature. The presence of the academic discourse can be explained by the fact that the professional books analysed in here contain an academic narrative for educational and pedagogical purposes as a means by

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

which to make the teaching of distinctive subjects to seafarers and officers in the classroom easier (cf. Franceschi 2014).

If we take into account the cumulative coverage of GSL and AWL (80.42% in the CONB and 81.61% in the COMEB) we observe that it is below the average of 86.1% found in academic texts (Nation 2000: 27). This clearly points to the demanding nature of maritime publications in terms of technical vocabulary load. Considering that, and knowing the first 2,000 GE words and most common academic vocabulary, we are left with nearly 20% of the specific vocabulary being unknown, or every fourth to fifth word of the narrative, which would hinder proper comprehension of these publications. This fact speaks in favour of the importance of maritime technical vocabulary, in the view of both Nautical Sciences and Marine Engineering.

From this specific research, it is evident that Nautical books are more demanding in terms of vocabulary load than Marine Engineering books. This again can be associated with the nature of the work on deck, including the larger scope of activities on the ship's bridge, including external communications with other ships and shore stations. In the case of marine engineers, communication rests on intra-ship communicative activities and deck-engine speech activities (De Castro 2020). However, further caution is required here since different findings have emerged from previous research conducted on Marine Engineering publications and vocabulary (Bocanegra-Valle 2013; Hsu 2014; Đurović et al. 2021), which all point to Marine Engineering technical publications being the most demanding ones. For example, in Marine Engineering instruction books, over a fifth (20.5%) of the vocabulary was "unknown", i.e. not covered by the GSL and the AWL (Đurović et al. 2021). These results point to the complexity and multidisciplinary nature of the maritime subareas, as well as the necessity of a detailed justification of the corpus and methodology applied, including their limitations.

5.2 Research Question 2

Considering the previous analysis, we raise the question regarding the amount of vocabulary needed for an adequate reading comprehension of Nautical and Marine Engineering publications. This evaluation method is one of the main advantages of word lists compared to dictionaries, which are generally much richer in entries, but at the same time, can be considered much more robust, as well.

For testing the required comprehension level of our corpora, we examined the coverage of the BNC/COCA GE word lists (Nation 2012), respectively. As usual in this type of research, we also tried to exclude the most frequent abbreviations, transparent compounds, marginal words and proper nouns (the four additional lists).

The comparative findings of the analysis conducted using the AntWordProfiler are given in Table 4.

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

Table 4: Coverage of the BNC/COCA lists in CONB and COMEB

Word lists	Coverage in CONB (%)	Coverage in COMEB (%)
2,000 + 4 additional lists	78.65	78.68
3,000 + 4 additional lists	87.74	87.12
4,000 + 4 additional lists	90.66	90.01
5,000 + 4 additional lists	92.05	92.52
6,000 + 4 additional lists	98.98	93.74
7,000 + 4 additional lists	93.51	94.66
8,000 + 4 additional lists	93.96	95.25
9,000 + 4 additional lists	94.23	95.83
10,000 + 4 additional lists	94.51	96.15
11,000 + 4 additional lists	94.69	96.33
12,000 + 4 additional lists	94.76	96.50
13,000 + 4 additional lists	94.90	96.61
14,000 + 4 additional lists	95.05	96.71
25,000 + 4 additional lists	95.60	97.32

From Table 4 it is apparent that the corpus of Marine Engineering books had better coverage in GE vocabulary. Hence, we note that the desired level of 95% required for adequate reading comprehension (Laufer 1989) is reached at the level of 8,000 GE words, which corresponds with the results obtained by Hsu (2014) for Marine Engineering textbooks in comparison with other areas of engineering. On the other hand, in Marine Engineering instruction books, that level is reached only after 12,000 GE words. Interestingly again, the corpus of Nautical books shows even lower coverage by GE vocabulary, since 95% coverage is reached at the level of 14,000 GE words. We can assume that this is the case because Marine Engineering is a branch of engineering that shares many discursive aspects with general engineering discourse (Borucinsky and Kegalj 2019; Bocanegra-Valle 2013), which makes it more multidisciplinary and more familiar to a wider range of English speakers. Furthermore, Nautical English can be considered the most technical in terms of the Maritime sciences. The only other domain sharing lexical registers with navigation is aviation. To the best of our knowledge, no similar research has been conducted on aviation publications or (text-)books, which is a potential area for further comparative research.

In addition, the ideal coverage of 98% (Hu and Nation 2000) is not reached in either corpus even with all the available 25,000 GE words, which definitely confirms that English for Maritime Purposes is a very technical ESP, the vocabulary of which requires special attention or teaching and learning efforts.

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

5.3 Research Question 3

A further distinction, and certain similarities, were analysed by testing the keyness of the two corpora.

Firstly, we tested the target corpora against each other, to obtain the key nautical terms tested against the ME corpus, and then the other way around: we analysed the ME keywords against the referent corpus of Nautical books (CONB). For this purpose, we used AntConc software (Anthony 2022). For practical reasons, we present here only the first 19 results extracted from the programme tables.

	A	B	C	D	E	F	G
1	type	freq_tar	freq_ref	range_tar	range_ref	likelihood	effect
2	you	2508	67	8	5	3094.486	0.007
3	your	1999	17	8	3	2716.386	0.005
4	gps	1368	0	7	0	1977.204	0.004
5	navigation	1602	112	8	7	1636.774	0.004
6	code	970	22	8	5	1219.845	0.003
7	error	1114	60	8	6	1215.87	0.003
8	satellite	830	2	5	2	1173.86	0.002
9	signal	1480	213	8	6	1140.987	0.004
10	draft	917	29	6	6	1104.389	0.002
11	maritime	866	32	6	6	1017.63	0.002
12	receiver	1119	121	6	6	984.936	0.003
13	compass	654	2	6	1	920.421	0.002
14	data	1569	353	8	7	903.514	0.004
15	tonnes	835	55	3	5	866.875	0.002
16	errors	723	34	8	5	812.387	0.002
17	chart	682	31	5	6	771.563	0.002
18	robot	525	1	2	1	745.299	0.001
19	map	578	16	7	2	709.135	0.002
20	center	606	24	7	2	703.533	0.002

Figure 1: Keyness of CONB vocabulary compared against COMEB

As can be seen from the columns in this programme setting, we have a list of the word types ranked by the Effect Size Measure + Threshold — which is the result of the keyness strength calculation and points to the possible cut-off points that we can choose as the thresholds for our table size, i.e. the length of the word list. Since we did not opt for a word list here but for comparative analyses of the keywords of the two corpora, we did not have to choose the cut-off point but, rather, a convenient presentation of the results. The same is true for the Likelihood Measure + Threshold column.

Regarding the list contents, most of the words belong to the nautical/navigational register (*GPS, navigation, maritime, compass*, and so on). On the other hand, we were surprised by the fact that the personal pronoun *you* was detected as one with a very distinctive frequency compared to the COMEB. Intrigued by the finding, and thanks to various other options offered by the

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

AntConc software, we explored the context of the word *you* in the nautical corpus further. The majority of the widespread use of *you* in instructions and explanations is found in the form of conditionals (Figure 2), which is obviously not the case with the COMEB, where, generally, passive forms prevail in describing systems and operations.

PADI instructors are held to the same exact standards whether	you'	re in the Caribbean, Germany, or Japan. It's
if you're in the eastern hemisphere and west if	you'	re in the western hemisphere. For example, if you'
if you're in the western hemisphere. For example, if	you'	re in the eastern hemisphere and the official time
you a precisely interpolated but uncorrected bearing of 225° + 107/4 = 252°.	You'	re in the eastern hemisphere and the official time
you a precisely interpolated but uncorrected bearing of 45° - 72/4 = 27°.	You'	re in the western hemisphere and the official time
should avoid them all. [Pufferfish are extremely poisonous; avoid unless	you'	re in a restaurant being served by a chef
the rushing water. Do not stay in a car. If	you'	re in a boat at sea, avoid shoals, shallow
the surf zone for the lull between wave sets. If	you'	re in a life raft or boat, see "Be
risks. Kidnapping on land—all pirates can do that. If	you'	re in a pirate gang, and you decide kidnapping
use an acoustic device that sounds like a bird if	you'	re in a bird sanctuary. Do not depend on

Figure 2: The word "you" used in the CONB context

These variations concerning the second-person pronoun *you* can be explained by the differences in the two discourses: nautical narrative storytelling, since the books are aimed at Nautical students and seafarers and cover countless "if" navigational situations that a navigator may encounter at sea (piracy, determining bearing and heading, navigation in shallow waters, etc.). Conversely, the discourse prevalent in Marine Engineering narratives is instructional, and focuses on lexical words (mainly nouns and verbs) which carry their meaning in collocations (e.g. *run the engine*) or are passive in the written texts (*the valve is opened*).

Keeping in mind the mentioned specificities of the navigational/engineering context, in Figure 3, specifically in the Range columns, we show that some nautical terms rarely appear in ME books, such as *GPS* (Global Positioning System) or *compass,* whilst some are present in both corpora, but much more frequently in Nautical books *(navigation, error, maritime, receiver).* Moreover, if we count the frequencies and keyness cumulatively for word families, some words such as *error(s*) would be even higher-ranked than they are in what is presented here.

Unlike the results given in Figure 1, Figure 3 presents typical Marine Engineering vocabulary that rarely appears in Nautical books. This includes words such as *valve* and *turbine*, but also words that appear in both corpora in the same range (or number of publications, respectively), such as *air* and *temperature*, but with relatively high frequency in Marine Engineering books. This makes them prevalent lexemes in this kind of maritime publication.

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

	A	B	C	D	E	F	G
1	type	freq_tar	freq_ref	range_tar	range_ref	likelihood	effect
2	engine	3569	136	8	6	3783.747	0.009
3	fuel	2701	81	8	6	2980.409	0.007
4	cylinder	2138	9	7	4	2742.362	0.005
5	engines	2147	14	7	5	2709.058	0.005
6	pressure	2513	239	7	7	2065.445	0.006
7	oil	2419	218	8	6	2030.218	0.006
8	valve	1618	17	7	2	1988.664	0.004
9	gas	1599	72	8	6	1638.106	0.004
10	piston	1161	0	6	0	1544.728	0.003
11	diesel	1209	9	8	2	1515.341	0.003
12	injection	1051	0	5	0	1398.303	0.003
13	exhaust	1057	8	6	3	1323.633	0.003
14	steam	997	6	7	3	1261.693	0.002
15	air	1874	272	8	8	1254.837	0.005
16	pump	1120	40	8	4	1199.882	0.003
17	turbine	911	3	7	1	1175.988	0.002
18	temperature	1327	131	7	7	1073.553	0.003
19	turbines	794	0	7	0	1056.255	0.002
20	propeller	932	25	7	4	1044.406	0.002

Figure 3: Keyness of COMEB vocabulary compared against CONB

Finally, we wanted to obtain the keyword lists for both areas/types of publications with reference to GE English, and compare the results. Considering the software requirements, best practice and the size of the corpora, we used, as mentioned, the FLOB corpus as the referent GE corpus.

Through separate analysis procedures, we obtained keyword lists from the CONB and COMEB, represented by the first 50 keywords for each of the vocabulary types (Table 5). For practical reasons, in this table, we put family members together (e.g. *signal* and *signals*, *user* and *use*, and so on)

Table 5: Keywords of the CONB and COMEB against the FLOB

No.	Nautical keywords	Marine engineering keywords
1	ship(s)	engine(s)
2	Navigation	fuel
3	GPS	speed
4	signal(s)	cylinder
5	Data	oil
6	receiver(s)	system
7	Figure	pressure
8	error(s)	figure
9	system, systems	water
10	Position	ship

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

11	Maritime	ice
12	Draft	valve
13	Code	marine
14	Tonnes	load
15	satellite, satellites	gas
16	Vessel	air
17	Using	diesel
18	Frequency	piston
19	Compass	pump
20	Water	injection
21	Angle	exhaust
22	Speed	control
23	Stability	temperature
24	Centre	propeller
25	Chart	low, lower
26	Matrix	steam
27	Velocity	turbine
28	meter(s)	current
29	Safety	shaft
30	Map	propulsion

Here it is clear that the key/technical vocabulary is different not only when directly compared against each other, but also when tested against GE. Considering that the two maritime areas share the language for general maritime purposes (IMO 2015), we can say that they are particularly distinct in terms of their lexical registers.

In addition, we have to pay special attention even to seemingly common technical vocabulary. For example, we have *bearing* in both lists, but in Nautical English it points to a position (e.g. 'Our vessel is bearing 215 degrees from you'), while in Marine Engineering it refers to a machinery component ('We must replace the main bearing'). To make it even more complicated, we have words from GE that have a completely different meaning in Maritime English. Another example of an ambiguous word detected in both corpora is the term *pressure*, since in Marine Engineering it refers to the physical force in the elements, while it is connected with weather systems in Nautical English.

The stated difference in the keywords in both corpora confirms the "polarity" of these two discourse communities and illustrates the problematic process of denotations in technical dictionaries. This brings us to the standing "conflict" among linguists and lexicographers with regard to the lexical vs. contextual meaning of entries (cf. Abdelzaher 2022). Taking into consideration the pronounced technical nature of the target corpus, as examined above, an eclectic approach is required in building technical dictionaries, still inclining towards contextual considerations, or even specific models such as collocate-to-sense mapping, as proposed by, for example, Kilgarriff (2005).

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

Another distinction worth mentioning in the specific case is the distinction between marine and maritime, as the two terms are often used interchangeably although they are related to different aspects of shipping. The word *maritime* occurs in the Nautical corpus but not in the Marine Engineering one, suggesting that the adjective *maritime* refers to activities associated with the navigational actions happening on deck. On the other hand, the word *marine* refers to operations relating to machinery, ship engines, environmental protection, or other activities under the ship's hull or in the sea (e.g. *marine engine, marine propulsion, marine fuels, marine environment, marine species*) (Dževerdanović-Pejović 2020a). In addition, the most frequent noun in the Nautical corpus, the word *ship*, has the most generic meaning in the Nautical corpus, as it refers to sailing, type of vessel, legal framework, conventions, and so on. Finally, the term *water* prevails in the Marine Engineering corpus as it is associated with the proper operation of the vital elements in the engine room, as in *the level of water in boilers, fresh water, bilge water, sea water*, and other similar phrases, whereas in the Nautical corpus it is mostly connected with navigational conditions (*calm water, safe waters, busy waters, high water*).

6. Limitations of the study

With using the above statistical methods, we should bear in mind not to strictly abide by the computational counts only. A further investigation into the obtained results and statistics, including as they relate to technical expertise and experience, is indispensable in the process.

As we can infer from the above analysis and discussion, there are numerous polysemous or cryptotechnical words (Fraser 2009) found in maritime publications that limit the accuracy of the results. For example, the noun *list*, found in the most frequent GE words in Maritime English, may refer to the inclination of the ship to one side or the other. In addition, some GE words in collocations with others refer to specific marine systems, such as *jacket water cooling system* or *guide shoe*, again referring to specific engine components. The same is found with nautical vocabulary such as *draft, stability* and *code*.

In addition, as was also found by Bocanegra-Valle (2013), seafaring- and ship-related words, especially nouns, are prone to being merged into compounds. For example, in Table 5 we find *shaft* as one of the most frequent keywords. If we were to consider the fact that it is frequently used in, for example, *crankshaft* and *camshaft*, by grouping these together with *shaft*, it would rank even higher on the list. The same criteria, generally not recognised by the lexical analysis software, would bring additional words beyond the decided cut-off point for frequency word lists. One additional detail should be borne in mind. The definition of *word* should be defined in advance, as it is (differently) the case in different programmes and their settings (e.g. headword/word family, lemma, word type or token/running word).

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

In addition, the specific results are always limited and conditioned by the selected corpus, as with the example of the results obtained from our research using Marine Engineering books and the results obtained from Marine technical manuals (Đurović et al. 2021). This means that even within a unique area of ESP we can obtain a variety of results. Therefore some authors (e.g. Gizatova 2018) only partly rely on corpus linguistics in building lexical macrostructures, such as in checking the frequency of otherwise selected terms, collocations or idioms. Conversely, our intention was to show the further prospective utility of technical corpora and their lexical analysis.

Although frequency is generally considered a solid base for headword selection, Rundell and Kilgarriff (2011) rightly mention that it is not an adequate criterion when, for example, extracting multiword items. In addition, there is always the dilemma on the right cut-off point, which, again, suggests that frequency on its own cannot guarantee that all the relevant words will be selected (Nielsen 2018; Vuković-Stamatović and Živković 2022). That is why, for these analyses, a detailed presentation of the methodology and the interpretation of the derived specific vocabulary in a specific context are required and recommended. In order to avoid the above-mentioned limitations, quantitative methods should be combined with qualitative expertise. In that way, a comprehensive study of the elicited figures regarding frequency, coverage and keywords should include both the data and a 'knowledge of the world' interface (Van Dijk 2014: 5).

7. Practical implications of the research

Regardless of the limitations of the methodology, the study still provides a very useful insight and measurable results in terms of the types of vocabulary, vocabulary load and specificity that this ESP may feature in comparison with similar research. In addition, lexicographers are provided with solid and justifiable methods for headword extraction, while language instructors and learners are offered corpus-based technical vocabulary and a methodology to focus on.

Our research aimed primarily to compare the discourses of two specific maritime communicative domains in terms of their technical vocabulary. However, the methodology and software used could be employed to derive a number of further lexicographic benefits. For example, the AntConc software can detect the most frequent collocations and *n*-grams, which enabled us to reveal the words in a particular context and isolate their contextual instances, including their specific and often different semantic aspects and connotations.

AntWordProfiler, on the other hand, could be used to develop frequency lists, such as the one generated from Marine Engineering instruction books (Đurović et al. 2021). The purpose of those lists (and their evaluation, at the same time) is to reach the level of 95% coverage in a target text sooner than

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

with GE word lists, and these lists are usually built upon the first 2,000–3,000 GE words, which are considered the most commonly known among English language learners. On the other hand, corpus-based frequency is a solid criterion for any word list, including glossaries and more complex lexicographic endeavours such as dictionaries. This has been the case with GE dictionaries, such as, for example the Collins COBUILD English Language Dictionary, Macmillan's Dictionary or many others produced in the meantime (Abdelzaher 2022; Đurović 2021). Having said that, the utilization of this criterion and generally corpus-linguistics methods has been rather slow with ESP dictionaries (Bowker 2010). One of the most widely cited reasons for this is the less abundant corpora compared to the one for GE. Thus, the frequency-based methodology of vocabulary detection has not been commonly used for technical dictionaries and glossaries, except for the numerous specialized word lists tailored to meet the specific needs of the language learners or professionals.

Frequency and keyword lists might have still further lexicographical implications. In addition to building frequency-based technical glossaries or dictionaries, both criteria can be combined for the purpose of comprehensiveness. In addition, knowing that deck and engine officers are in favour of patterns or schemes in the acquisition and the production of restricted verbal and written genres (Dževerdanović-Pejović 2020b), frequency and keyword lists, as well as certain other options offered by modern lexical software, can also serve to establish the co-occurrence and mapping of a specific genre, contributing to the accurate and reliable metalanguage of a technical dictionary or similar segmental lexicographical product.

In addition, the methodology (or methodologies) can be combined and utilized not only for segmental lexicography, but also for building hybrid dictionaries (cf. Bowker 2010). For example, a hybrid ME dictionary could include the most frequent nautical and/or marine engineering words, but also the most frequent GE ones as extracted from the same corpus, with additional semantic notions in the given corporal context due to the polysemous character of some GE and "cryptotechnical" words. As such, it would cover the full English vocabulary load of the specific genre(s).

Finally, the possibility of using modern software tools in the analysis of large corpora and thus of tackling a huge amount of data enables us not only to dig into a particular professional communicative domain, but also to engage in comparative and contrastive research on specific lexes and to explore the worlds in which their semantic matters are imprinted. In this way, the specific lexicographic metalanguage would reflect the peculiarities of distinctive technical subareas, such as those shown in the nautical vs. marine engineering context.

8. Conclusion

Intrigued and inspired by the distinctiveness of the lexical registers characteristic of the nautical and marine engineering genres, which at the same time share

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

the General Maritime vocabulary, we were looking for adequate corpus linguistics and statistical lexical methods that could provide us with measurable results in terms of the differences between the two types of vocabulary. For that purpose, we sought to provide answers to three research questions related to the compiled corpora of Nautical and Marine Engineering books. The aim of the three-step research was to provide a solid foundation for the separate treatment of specialized maritime lexicons dedicated to the professionals in those specific fields.

Firstly, both genres proved to be very challenging in terms of the technical vocabulary load, since the coverage of the corpora with GE vocabulary together with academic vocabulary was lower than in other types of texts. This was even more the case for Nautical books.

Similar results were found in measuring the coverage of the corpora by GE vocabulary only (BNC/COCA word lists). This part of the lexical analysis showed that adequate reading comprehension (expected at a level of 95% coverage with GE vocabulary) would be achieved with no fewer than 14,000 GE words in the case of Nautical books and with 8,000 words in the case of Marine Engineering books. Again, Nautical books proved to be more technical vocabulary-wise, whilst the ideal threshold of 98% for ideal reading comprehension was not reachable even with all the available word lists covering GE vocabulary.

In addition, we extracted the keyword lists from both corpora (in comparison with a referent GE corpus) and examined them for potential similarities and differences. This provided us with additional evident differences in the two registers, genres and discourses. Some common terms, such as *pressure, bearing* and so on, hold different meanings in the two respective corpora, pointing to the highly represented phenomenon of polysemy in maritime lexis.

Considering the decisions made by throughout the process, we followed the general recommendations for the methodologies applied, but also stated the limitations of the study, primarily those related to the selected genres i.e., the specific content of the corpora. It was also noted that the entire process of analysis cannot rest on statistical results only, but also requires expert knowledge with regard to the two maritime areas and their respective registers. In addition, the research findings point to the great challenge imposed on Maritime English lexicographers and the special attention required when dealing with these demanding and intriguing areas of English for Specific Purposes.

Finally, having at hand very meticulous methodologies for technical vocabulary extraction, a more comprehensive project could be conducted to comprise various genres and combined methodologies for a more complex lexicographical product such as, for example, a (mono-, bi- or multi-lingual) Marine Engineering dictionary.

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

Endnotes

1. A word family comprises the head word with all its inflected and derived forms.
2. Transparent compounds are compounds where the meaning can be understood from the separate meanings of their constituents.

References

Abdelzaher, E.M. 2022. An Investigation of Corpus Contributions to Lexicographic Challenges over the Past Ten Years. *Lexikos* 32(1): 162-179.

Al-Rawi, M.K.S. 2017. Using AntConc: A Corpus-based Tool to Investigate and Analyse the Keywords in Dickens' Novel 'A Tale Of Two Cities'. *International Journal of Advanced Research* 5(2): 366-372.

Atkins, S. 2008. Theoretical Lexicography and its Relation to Dictionary-making. Fontenelle, T. (Ed.). 2008. *Practical Lexicography: A Reader*: 31-50. Oxford: Oxford University Press.

Baker, P. 2004. Querying Keywords: Questions of Difference, Frequency and Sense in Keywords Analysis. *Journal of English Linguistics* 32(4): 346-359.

Bisht, R.K., Dhami, H.S. and N. Tiwari. 2011. An Evaluation of Different Statistical Techniques of Collocation Extraction Using a Probability Measure to Word Combinations. *Journal of Quantitative Linguistics* 13(2–3): 161-175.

Bocanegra-Valle, A. 2013. Maritime English. Chapelle, C.A. (Ed.). 2013. *The Encyclopedia of Applied Linguistics:* 3570-3583. Hoboken: Blackwell Publishing.

Borucinsky, M. and J. Kegalj. 2019. Syntactic Ambiguity of (Complex) Nominal Groups in Technical English. *International Journal of English Studies* 19(2): 83-102.

Bowker, L. 2010. The Contribution of Corpus Linguistics to the Development of Specialised Dictionaries for Learners. Fuertes-Olivera, P. (Ed.). 2010. *Specialised Dictionaries for Learners*: 155-168. Berlin: De Gruyter.

Chen, Q. and G. Ge. 2007. A Corpus-based Lexical Study on Frequency and Distribution of Coxhead's AWL Word Families in Medical Research Articles (RAs). *English for Specific Purposes* 26: 502-514.

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.

Cole, C., Pritchard, B. and P. Trenkner. 2007. Maritime English Instruction — Ensuring Instructors' Competence. *Ibérica* 14: 123-148.

Coxhead, A. 2000. A New Academic Word List. *TESOL Quarterly* 34(2): 213-238.

Čulić-Viskota, A. and S. Kalebota. 2013. Maritime English — What Does It Communicate? *Transactions on Maritime Science* 2(2): 109-114.

Čulić-Viskota, A. and H. Rummel. 2022. Tribute to Kurt Opitz (1931–2021) — German Contributor to Maritime English — What Does It Communicate? *Transactions on Maritime Science* 11(2). https://doi.org/10.7225/toms.v11.n02.019

De Castro, G. 2020. A Needs Analysis Study of Engine-Deck Communication: Towards Improving Syllabus Design. *Taiwan International ESP Journal* 11(2): 35-53.

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

Dževerdanović-Pejović, M. 2020a. Discourse Analysis of the Research Articles About Marine Environment Relating to the Adriatic Coast. Joksimović, D., M. Đurović, I.S. Zonn, A.G. Kostianoy and A.V. Semenov (Eds.). 2020. *The Montenegrin Adriatic Coast. The Handbook of Environmental Chemistry* 110: 175-189. Cham: Springer.

Dževerdanović-Pejović, M. 2020b. Learning Technical Genres — A Blended Learning Approach. *Pomorstvo: Scientific Journal of Maritime Research* 34(2): 212-222.

Đurović, Z. 2021. Corpus Linguistics Methods for Building ESP Word Lists, Glossaries and Dictionaries on the Example of a Marine Engineering Word List. *Lexikos* 31: 259-282. http://dx.doi.org/10.5788/31-1-1647

Đurović, Z. 2023. Frequency or Keyness? *Lexikos* 33: 184-206. https://doi.org/10.5788/33-1-1807

Đurović, Z., M. Vuković Stamatović 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: 123-133. https://doi.org/10.5209/clac.78300

Franceschi, D. 2014. The Features of Maritime English Discourse. *International Journal of English Linguistics* 4(2): 78-87.

Fraser, S. 2007. Providing ESP Learners with the Vocabulary They Need: Corpora and the Creation of Specialized Word Lists. *Hiroshima Studies in Language and Language Education* 10: 127-143.

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.

Gizatova, G. 2018. A Corpus-based Approach to Lexicography: A New English–Russian Phraseological Dictionary. *International Journal of English Linguistics* 8(3): 357-363.

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: 403-430.

IMO Model Course 3.17, Maritime English. 2015. London: International Maritime Organization.

Kilgarriff, A. 2005. Putting the Corpus into the Dictionary. *Proceedings of the Second MEANING Workshop*. Trento, Italy, 3–4 February 2005.

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.

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. 2012. *The BNC/COCA Word Family Lists.* Available at: http://www.victoria.ac.nz/lals/about/staff/paul-nation.

Nation, I.S.P. and A. Heatley. 1994. *Range: A Program for the Analysis of Vocabulary in Texts (Computer Software).* Retrieved from: http://www.victoria.ac.nz/lals/staff/paul–nation/nation.aspx

http://lexikos.journals.ac.za; https://doi.org/10.5788/33-1-1814 (Article)

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.

Nielsen, S. 2018. LSP Lexicography and Typology of Specialised Dictionaries. Humbley, J., G. Budin and C. Laurén (Eds.). 2018. *Languages for Special Purposes: An International Handbook*: 71-95. Berlin/Boston: De Gruyter.

Pritchard, B. 2013. The English Element in the Development of Croatian Maritime Terminology. *Pomorstvo: Scientific Journal of Maritime Research* 27(1): 247-259.

Rundell, M. and A. Kilgarriff. 2011. Automating the Creation of Dictionaries: Where Will it All End? Meunier, F. et al. (Eds.). 2011. *A Taste for Corpora. In Honour of Sylviane Granger*: 257-282. Amsterdam/Philadelphia: John Benjamins.

Sinclair, J. 1992. *Collins COBUILD English Language Dictionary*. London: HarperCollins.

Stubbs, M. 2010. Three Concepts of Keywords. Bondi, M. and M. Scott (Eds.). 2010. *Keyness in Texts: Corpus Linguistic Investigations*: 21-42. Amsterdam: John Benjamins.

Valipouri, L. and H. Nassaji. 2013. A Corpus-based Study of Academic Vocabulary in Chemistry Research Articles. *Journal of English for Academic Purposes* 12(4): 248-263.

Van Dijk, T.A. 2014. *Discourse and Knowledge: A Sociocognitive Approach*. Cambridge/New York: Cambridge University Press.

Vongpumivitch, V., J.-Y. Huang and Y.-C. Chang. 2009. Frequency Analysis of the Words in the Academic Word List (AWL) and Non-AWL Content Words in Applied Linguistics Research Papers. *English for Specific Purposes* 28: 33-41.

Vuković-Stamatović, M. and B. Živković. 2022. Corpus-based Headword Selection Procedures for LSP Word Lists and LSP Dictionaries. *Lexikos* 32(1): 141-161.

West, M. 1953. *A General Service List of English Words*. London: Longman, Green & Co.