Science for Education Today

2023, vol. 13, issue 2 http://en.sciforedu.ru/ ISSN 2658-6762

DOI: 10.15293/2658-6762.2302.08 Research Full Article/ Article language: English

The potential of using corpus-based technologies for incidental vocabulary learning of General Maritime English

A. A. Mironenko[1], E. V. Mironenko[1]

[1] Admiral Makarov State University of Maritime and Inland Shipping, Saint Petersburg, Russian Federation

Abstract

Introduction. *The article addresses the problem of expanding the potential vocabulary of cadets of maritime educational institutions in the process of listening comprehension determined by the need to develop strategies for independent vocabulary acquisition necessary for autonomous learning of maritime English.*

The purpose of this article is to justify the application of corpus technologies for the selection of vocabulary items, which will allow students to master the strategies required to expand vocabulary in the process of autonomous listening, and increase the efficiency of learning and teaching maritime English.

Materials and Methods. *The study follows the two-level Maritime English learning model incorporating General Maritime English and Specialized Maritime English. The first component focuses on the linguistic content embedded in the generalized maritime context and prioritizes the lexical and cognitive approaches highlighting the necessity of their integration and emphasizing corpus tools the application of which is helpful in topic-related vocabulary selection through instructional, linguistic and statistic principles on the basis of authentic texts.*

Results. *The article describes the preliminary stage of potential vocabulary acquisition by junior cadets through listening comprehension and proposes the procedure of text selection for context-specific raw corpus design and thematic word list generation. Through the computer processing of the wordlist and the analysis of the concordance lines, the independent vocabulary learning strategies were identified and ranked. The conclusion was drawn about the correlations between the basic word learning potential and the number of independent word learning strategies applied.*

Conclusions. *The findings of the research add to the understanding of the potential of corpus-based technologies for the topical vocabulary selection and analysis in order to develop cadets' independent word learning strategies that are significant for potential vocabulary acquisition through listening comprehension.*

For citation

Mironenko A. A., Mironenko E. V. The potential of using corpus-based technologies for incidental vocabulary learning of General Maritime English. *Science for Education Today*, 2023, vol. 13 (2), pp. 172–191. DOI: http:/dx.doi.org/10.15293/2658-6762.2302.08

✉ *Corresponding Author: E. V. Mironenko, mironenkoev@gumrf.ru*

The results achieved suggest that the word lists generated from the subject corpus should be integrated into the basic vocabulary of General Maritime English, which should be mastered under the teacher's guidance.

Keywords

General Maritime English; Corpus-based technologies; Incidental vocabulary learning; Vocabulary size; Context-specialized corpus; Word learning strategies.

Introduction

In compliance with the updated IMO model course 3.17 Maritime English (ME) curriculum design and content selection should develop a feasible, but flexible paradigm, which meets the professional needs while setting language learning in social and institutional contexts and scenarios. For that reason, the latest edition of the modal course has been revised into two "new" core sections: General Maritime English (GME) and Specialized Maritime English (SME) where the former focuses on the linguistic content embedded in the generalized maritime context and the latter focuses on the professional content realized through the English language [1]. Thus, the teaching goal of GME is in preparing junior maritime cadets to acquire the Specialized Maritime English aimed at "achieving the effective communication competences related to the English language in the STCW Convention, 1978, as amended"[1].

The language content for the GME course integrates the three areas of language input that is phonology, vocabulary and grammar with practice of the four language communicative skills namely listening, speaking, reading and writing at the elementary and lower-intermediate levels of proficiency in ME covering typical sets of communicative events that seafarers of all ranks may become involved in. By the concept of teaching EFL at "Maritime academy" institute that is a structural unit of Admiral Makarov State University of Maritime and Inland Shipping the minimum entry level of the prospective cadets' General English should not be lower than Elementary (A2 on the CERF scale), however the results of the diagnostic testing which is regularly arranged at the Department of English for Navigation and Radio Communication at the beginning of the 1st academic year show that a third of newly-entered cadets do not confirm that level. When it comes to the assessment of the particular language aspects and skills, listening comprehension remains one of the most challenging activities[2]. Acknowledging the fact that lexis is of primary importance in providing effective communicative activities, both receptive and productive[3], an insufficient level of audio comprehension can witness to limited or inadequate vocabulary [2; 3; 4].

Being guided by the practicable and internationally accepted descriptive scale that clearly identifies the Maritime English communication performance required for the STCW Operational and Management Levels where the abilities "to communicate fluently on radio complying with the Radio Regulations", "respond competently in emergencies", "comprehend the content of a message",

[1] *Model Course 3.17. Maritime English*, 2nd ed. International Maritime Organization: London, UK, 2015, p. 11–12, ISBN 978-92-801-1622-9

[2] Sherbakova I. O. *On the concept of teaching EFL at the institute "Maritime academy"*. (In Russian) URL: https://www.elibrary.ru/item.asp?id=47302594

[3] Kuznetsova T. N. *Key issues in teaching lexis*. (In Russian) URL: https://elibrary.ru/item.asp?id=48869991

"understand written and oral instructions" [4] are among the key descriptors,, GME teacher should not ignore developing cadets' listening skills. But in conditions of in-class hour deficiency (according to the current GME syllabus 1st year deck cadets have on average 3 academic hours per week) it is a big challenge. However, the mainstreaming of e-learning the essence of which is the organization of educational activities using various multimedia tools and distant learning technologies provides many incentives for learners enhancing their readiness for independent work in an electronic educational environment and making the process of foreign language acquisition more learner-centered [5; 6; 7; 8]. To benefit from it, the educational content of GME should be properly designed and the balance between explicit (teacher-guided) and incidental (autonomous) learning should be carefully measured. Bearing in mind that vocabulary size is among the major contributors to SL learners' proficiency and tackling vocabulary incidentally is a time and energy-consuming process [9, p. 208], we intend looking closely at the preparatory stage of incidental vocabulary learning from listening considering the preconditions and corpora tool-assisted procedure of vocabulary selection.

Methods

The degree of cadets' autonomy depends on appropriate organization of the educational process grounded on the combination of carefully selected traditional and innovative approaches.

In the context of Communicative Language Teaching approach which is implemented as the principal means of maritime cadets' foreign language learning the maritime education and training (MET) instructors should teach in a way that develops communicative competence emphasizing authenticity and grounding the language in the environment of the occupation [10; 11]. For seafarers to communicate effectively, they need to be able to use and understand English in a range of situations where "being able to use English" means that the seafarer can combine the 'building blocks' of language (grammar, vocabulary, phonology) to express him/herself clearly and appropriately in speech and writing and "being able to understand English" means their ability to interpret messages they hear or read correctly and respond to them appropriately and comprehensibly[5].

The importance of lexis in ME teaching has ever been well-realized, as the reach language of the sea has developed through the centuries and accumulated a huge number of terms, which are the core of any ESP. The reciprocating influences of General English and Maritime English have brought into existence ambiguous layers of lexical units which wrong interpretation can lead to the professional communication breakdown and which acquisition should be prioritized[6]. In the light of the communicative approach the role of lexis has been reconsidered and its meaning-making potential has been widely recognized [12; 13]. It is enforced by the basic principle of the Lexical approach according to which "Language consists of grammaticalized lexis, not lexicalized grammar" and according to which stress is laid on

[4] Cole C. W., Trenkner P. *The Yardstick for Maritime English STCW assessment purposes*, 2008, pp. 1–15. URL: http://imla.co/sites/default/files/imec20.pdf

[5] Richards J. C. *Communicative language teaching today*, 1st ed., New York: Cambridge University Press, 2006, pp. 2–14. ISBN-139778-0-521-92512-9

[6] Mironenko E. V., Mironenko A. A. E-thesaurus design principles on the basis of maritime pseudo-equivalent. *IMLA International Maritime English Conference (IMEC31).* 2019, P. 111–121. URL: http://www.imla.co/imec/IMEC31Proceedings.pdf

lexical chunks rather than individual words[7]. Shifting the focus to the cognitive vocabulary approach incites the teachers to harness learners' analytic abilities. Cognitive linguistics treats language and its acquisition as usage-based and as reflecting the general cognitive abilities that operate in L2 learners' interaction with the world. In this view linguistic phenomena are considered to be 'motivated' [9, p. 211]. Some things are more likely to happen in language than others because they are more congruent with habitual learners' perceptual and cognitive experience. For example, helping the students develop "independent word learning strategies that will serve them well in subsequent encounters with unfamiliar words"[8] is a way to motivate learners for incidental learning [14; 15].

The advent of the corpus linguistics gave a new impetus to a lexical view of language acquisition accelerating and improving the processing of bulk information. Large general-purpose corpora which are freely available on-line provide the course designers with reliable data as to word frequency and habitual co-occurences of some course-related words taken from a wide range of natural language text files[9]. Moreover, using up-to-date software and "web as a corpora" technologies researchers can compile smaller, carefully collected, context-specific corpora containing linguistic material consistent with specific language register or situationally defined text categories such as 'fiction', 'news article' etc.[10] [16; 17].

Extensive listening or viewing audiovisual media leads to incidental vocabulary learning and inspires bottom-up process leading to greater automatic word recognition. Prior vocabulary knowledge is among the factors that affect deducing the meaning of new words without the assistance of an instructor [18], so to design more effective resources, the primary focus of materials developers should be on the compiling of the core wordlist which would serve as the framework for incidental vocabulary learning. While selecting the words for the wordlist, we were guided by a number of principles broadly grouped into three categories: a) instructional, b) linguistic and c) statistic[11].

a) *Instructional principles* are targeted at the educational program specialization, learning needs, and relevance of the lexis to the topics stipulated by the academic course syllabus. In light of these principles, the teacher should have a clear understanding of the size of the vocabulary, its composition and what vocabulary knowledge is needed for autonomous listening.

As specified by the English model syllabus of the Russian Federation complete secondary education, initially the first-year cadets, who are by and large the recent leavers of Russian comprehensive secondary schools, should be able to use orally or in writing 1300 lexical units and recognize in reading or listening 1400 lexical

[7] Lewis M. *Implementing the lexical approach: Putting theory into practice*. Heinle, Cengage Learning: Andover, Hampshire, UK, 2008, pp. 17–59. ISBN: 978-1-899396-60-3

[8] Boers F., Lindstromberg S. *Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology*. Mouton de Gruyter: Berlin-New York, 2008, pp. 17–27. ISBN 978-3-11-01963333330-6

[9] McCarten J. *Teaching Vocabulary: Lessons from the Corpus, Lessons for the Classroom*. USA, New York: Cambridge University Press, 2007, pp. 2–1., ISBN 13 978-0-521-94325-3

[10] Vaughan E., Clancy B. Small corpora and pragmatics. In: Romero-Trillo J. (eds) *Yearbook of Corpus Linguistics and Pragmatics*, 2013, vol. 1. Springer, Dordrecht, pp. 53–73. DOI: http://dx.doi.org//10.1007/978-94-007-6250-3_4

[11] Gal'skova N. D., Gez N. I. *Teoriya obucheniya inostrannym yazykam. Lingvodidaktika i metodika*, izd. 3-e., ster. M.: Akademiya, 2006, pp. 294–296. ISBN 5-7695-2969-5

units [12]. The vocabulary size of the technical university graduates is not fixed and may vary from 2300 to 7000 lexical units. Taking into account the linguo-statistical data, which speak that for successful retrieval of information from receptive activities (particularly, skim reading or listening) the graduate's vocabulary should be about 2500 lexical units which provide 75–80 % of reading or listening comprehension [13], we understand the importance of the autonomous learning in the process of which the lexis beyond the university English syllabus can be acquired.

When it comes to the composition of the topical wordlist, it depends on the educational program specialization. In compliance with the IMO model course[14], the topology of the ME is based on word frequency research. Accordingly, all the words are ranged from not technical words to strictly technical ones and categorized as follows:

1st level: high frequency words – usually the 2000 – word level is set as the most suitable limit for high frequency words (80% of running words of academic texts and newspapers and 90% of conversations and novels) [15]. This list includes about 165 families of function words (e.g. *the, is, in, two, because*), and the majority of content words, that is nouns (*globe, heat*), verbs (*explore, protect*), adjectives (*sandy, coastal*) and adverbs (*ashore, highly*). These are not technical words.

2nd level: academic vocabulary – a specialized extension of 1st level words which is also termed 'semi-technical' (8.5% of academic texts; 4 % of newspapers). These words are often highly polysemous and can be disambiguated in the maritime context only (*bank, dolphin, palm*)[16].

3rd level: a very limited number (up to 5 – 7%) of nautical terms, which central lexical meaning is restricted to maritime use only and thus 'unambiguous' by nature (*double-hull, log-book, seaworthiness*)[17].

In MET institutions, the 1st and 2nd years of English training is a transition stage from GME to SME. Over this period particular attention is paid to thematic words which are beyond maritime terminology but will contribute to its comprehension in the course of study. Hence, it is quite evident that non-technical vocabulary with the 'speckles' of semi-technical words is in the focus of both explicit and incidental learning.

Answering the question what vocabulary knowledge is needed for listening we considered that at the most general level, knowing a word is presented as an indivisible polyfunctional system which involves form, meaning and use[18, 19]. An effective acquisition of a sizable vocabulary requires its differentiation. Traditionally language teaching method scholars distinguish receptive (passive) and productive (active) vocabulary. "Receptive vocabulary use involves the form of a

[12] Primernaya rabochaya programma srednego obshhego obrazovaniya predmeta «Anglijskij yazyk». URL: https://edsoo.ru/Primernaya_rabochaya_programma_srednego_obschego_obrazovaniya_predmeta_Anglijskij_yazik_.htm

[13] Folomkina S. K. *Obuchenie chteniyu na inostrannom yazyke v neyazykovom vuze: ucheb.-metod. posobie*, izd. 2-e, ispr. M.: Vysshaya shkola, 2005, pp. 49–71. ISBN 5-06-005417-9

[14] *Model Course 3.17. Maritime English*. Annex, pp. 238 – 239.

[15] Nation I. S. P. *Learning Vocabulary in Another Language*, 2nd ed. Cambridge: Cambridge University Press, 2013, pp. 22. ISBN 9781139858656 DOI: https://doi.org/10.1017/CBO9781139858656

[16] Mironenko E. V. Inoyazychnaya kommunikativnaya kompetenciya sudovoditelya kak sredstvo obespecheniya bezopasnosti v mezhdunarodnom sudoxodstve. *Vysshee obrazovanie segodnya*, 2012, no. 7, pp. 31. (In Russian) URL: https://www.elibrary.ru/item.asp?id=18249984

[17] Ibid., p. 30.

[18] Nation I. S. P. *Learning Vocabulary in Another Language*, p. 39.

[19] Passov E.I., *Osnovy kommunikativnoj metodiki obucheniya inoyazychnomu obshheniyu*. M.: Russ.yaz., 1989, pp. 132–246. ISBN 5-200-00717-8

word while listening or reading and retrieving its meaning. Productive vocabulary use involves wanting to express a meaning through speaking or writing and retrieving and producing the appropriate spoken or written word form"[20].

There is on more type of the vocabulary that comprises partly known words which are the words which meaning can be guessed from the context while reading or listening though they haven't been present in the learner's language background before. This vocabulary is recognized as "potential vocabulary" and it is the contextual guess which makes this type of the vocabulary controllable while its incidental acquisition through receptive activities[21].

b) Among *linguistic* principles the most research-relevant ones are:

– co-occurrence, according to which the value of lexical unit depends on its ability to collocate with other words,

– ord-formation potential that is the possibility of the word to form derivatives and to predetermine contextual guess and independent semantization,

– polysemy that is characterized by the coexistence of several related meanings for a word or phrase[22].

It is of great importance here to decide what we count as a word and which independent word learning strategies can be applied.

In compliance with the Lexical approach, the term "lexical item" applies to all vocabulary, not just single words but also 'prefabricated multi-word chunks'[23] which include items such as collocations (*heavy rain*), fixed and semi-fixed expressions (*nice day for it*) and idioms (*to have wind in sails*).

To deduce the meaning of new words without the assistance of an instructor cadets should be encouraged to use a range of implicit vocabulary learning strategies or determination strategies when an individual discovers the meaning of a new word "through guessing from one's structural knowledge of a language, guessing from an L1cognate, guessing from context, or using reference material"[24]. This requires from learners to be active constructors of the new knowledge on the basis of the familiar one. Among a number of foreign and domestic scholars (M. Lewis[25], I. S. P. Nation[26], N. Schmitt[27], N. D. Gal'skova, N. I. Gez[28], S. K. Folomkina[29]) there is a consensus of opinions that the choice of word learning strategies depends on the nature of the lexical unit and is aimed at the developing of learners' ability to guess the meaning of the word on their own while reading or listening [19; 20].

Thus, the main strategies applicable for potential vocabulary extension are as follows:

– guessing the word's meaning from context (true cognates),

– analysing affixes and roots (affixation),

– analysing part of speech (conversion),

[20] Nation I. S. P. *Learning Vocabulary in Another Language*, p. 37.

[21] Folomkina S. K. *Obuchenie chteniyu na inostrannom yazyke v neyazykovom vuze: ucheb.-metod. posobie*, p. 55.

[22] Gal'skova N. D., Gez N. I., *Teoriya obucheniya inostrannym yazykam. Lingvodidaktika i metodikaj*, p. 295.

[23] Lewis M. *Implementing the lexical approach: Putting theory into practice*, p. 20.

[24] Schmitt N. *Vocabulary in Language Teaching.* Cambridge: Cambridge University Press, Cambridge, United Kingdom, 2000, pp. 135. ISBN 0-521-66048-3

[25] Lewis M. *Implementing the lexical approach: Putting theory into practice*, p. 47–49.

[26] Nation I. S. P. Learning Vocabulary in Another Language, p. 421–422.

[27] Schmitt N. *Vocabulary in Language Teaching*, p. 132–138.

[28] Gal'skova N. D., Gez N. I., *Teoriya obucheniya inostrannym yazykam. Lingvodidaktika i metodikaj*, p. 298–299.

[29] Folomkina S. K. Obuchenie chteniyu na inostrannom yazyke v neyazykovom vuze, p. 64–68.

– analysing the different meanings of one word (polysemy),

– analyzing the composition of lexical chunks, the elements of which are known to the learners.

The belief of incidental lexis uptake rests on analogy with native language acquisition. Since most vocabulary in the mother-tongue is acquired, not taught, the ideal method of acquiring new words is to be exposed to enough suitable input applying an appropriate independent word learning strategy.

c) The main *statistic* principle of vocabulary selection is the frequency of word usage, which is defined by N. D. Gal'skova and N. I. Gez as "the property of the word to be used in a number of texts with a definite frequency. Consequently, frequency and usage of the word are considered as two equally-valued components of one parameter"[30].

One of the most efficient ways which helps to process a considerable amount of statistical information is corpus software tools [17; 21]. In addition to selecting and ranging the most frequently used words (a quantitative analysis), the corpus also gives us access to hundreds of texts in order to observe how the targeted vocabulary is used in context that is the subject of a qualitative analysis.

On the one hand, there are a number of multi-million-word professional general English corpora, such as the pioneering Brown Corpus (1961–64), Collins COBUILD Bank of English (1987), British National Corpus (1994), The Corpus of Contemporary American English (1990)[31], etc. which can generate thousands of concordance lines with any 'key-word-in-context' to provide the researchers with huge amount of information about the particular word (syntactic, semantic and extra linguistic information about the target word, its grammatical use, habitual co-occurrences with the other words). However, very often a relatively small amount of well-chosen vocabulary can allow learners to accomplish learning goals more efficiently. In such a case, tailor-made corpora compiled from disciplinary specialized materials are of great value.

We believe that the rational implication of all above-mentioned approaches into potential vocabulary acquisition and principles of its selection for developing maritime cadets' independent word learning strategies will contribute to qualitative changes in GME teaching and make incidental learning more cadet-friendly.

Results

Driven by the wish to devise a syllabus which would enable learners to acquire language within a wider context through listening more productively, we staged the process as follows.

First, it was necessary to compile a context-specialized corpus in order to retrieve the wordlist which would serve as a starting point in enlarging cadets' potential vocabulary. Following the lexical approach claim that "more meaning is carried by lexis in naturally occurring language than merely by grammatical structures" [12, p. 5] we analyzed the relevant ME educational resources to select the texts which can be exposed to the cadets. The search for podcast hosting sites from the list of "General English Language teaching and testing websites" of the Model course[32] was made and the following criteria for the podcast selection were applied:

– an open access to the podcast;

– topic relevance;

– availability of a transcript;

[30] Gal'skova N. D., Gez N. I., *Teoriya obucheniya inostrannym yazykam. Lingvodidaktika i metodika*, p. 294.

[31] Schmitt N. *Vocabulary in Language Teaching*, pp. 68–71.

[32] *Model Course 3.17. Maritime English*, pp. 21–22.

– code-level of the English language proficiency (according to CEFR);

– style heterogeneity.

Once the corpus is 1st academic year English syllabus oriented, the priority was given to the topic "Geography and Weather" as it allows to immerse the cadets in the atmosphere of seafaring to a large extent. Using the continuous sampling method, we selected 100 topic-related audio files provided with the transcripts. Considering the objective necessity to implement a two-level approach to the GME teaching resulted from the fact that the first-year cadets' level of the English language competence may vary from beginner (A1) to intermediate (B1) and that the process of foreign language acquisition is successful when the learner understands a message or receives "comprehensible input" that is a step beyond his current stage of linguistic competence [12, p. 6], the selected texts meet A1 – B2+ level. The duration of the audio sample varies from 45 sec. (0′45″) till 15 min. 55 sec. (15′55″) – transcripts length – from about 80 to 1400 running words. The samples cover 4 text categories which are ranged from popular science and news reports to fiction prose and dialogues on the street. The quantity distribution of the geography-and-weather-related texts over the podcasts is shown in table 1.

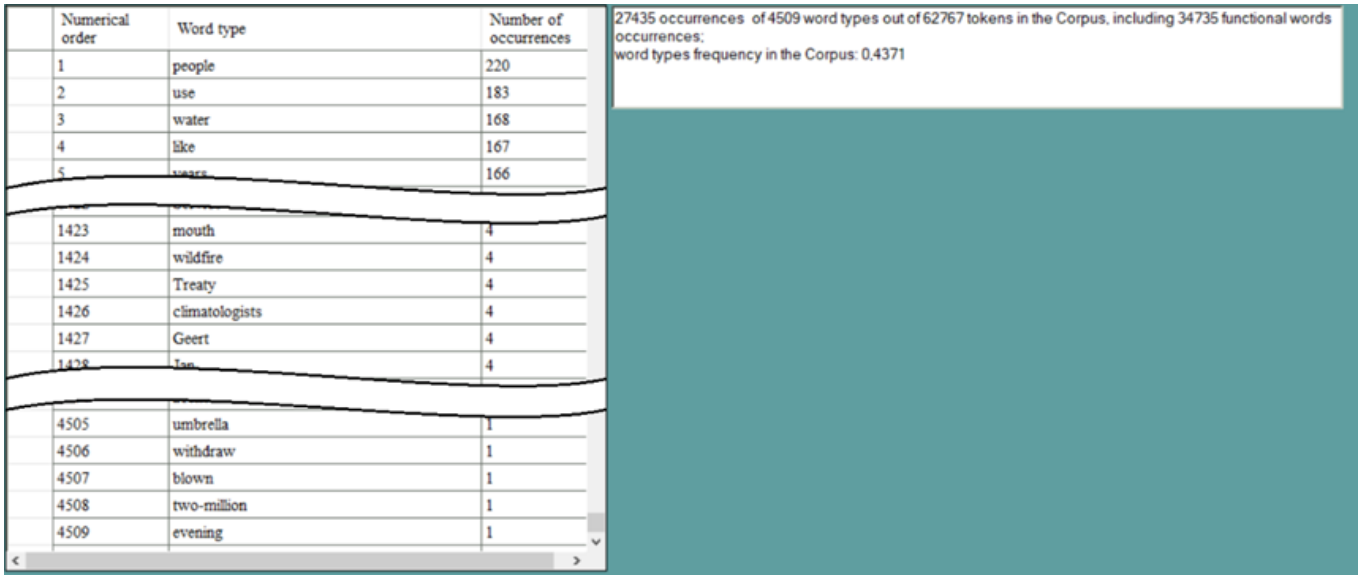
Table 1

The quantity distribution of the geography-and-weather-related transcripts over the selected podcasts

Site	Language level	Podcast	Duration range	Number of transcripts
English club	Beginner (A1)	*Listen and Learn*	0′57″ – 1′24″	5
		Interesting facts	1′03″ – 1′29″	6
		This Week in History	0′45″ – 1′02″	3
BBC Learning English	Elementary (A2)	*Course*	1′37″	1
	Intermediate (B1)	*The English We Speak*	2′02″ – 2′49″	7
		6 Minute English	6′06″ – 6′46″	12
	Intermediate (B1+)	*6 Minute Vocabulary*	5′45″	1
		Drama	6′18″ – 9′02″	16
British Council	Intermediate (B1)	*Skills-Listening*	2′08″	1
	Intermediate (B1)/ Upper Intermediate (B2)	*Video series – Word on the street*	2′28″ – 5′25″	9
	Upper intermediate (B2+)	*Video Zone*	1′38″ – 3′41″	3
		Video series – Britain is Great	3′51″ – 4′43″	2
VOA Learning English	Intermediate (B1)	*Science and Technology*	2′47″ – 6′16″	10
		Words and their stories	4′17″ – 6′32″	12
Many Things	Upper Intermediate (B2)	*MT_Science in the News*	15′30″	3
		MT_Places	13′48″ – 15′55″	8

The transcripts of the selected 100 audio samples were downloaded, converted into *.txt format and collected in one database. The processing of this massive with the help of the program developed on the basis of Microsoft Visual Studio platform resulted in compiling a raw corpus. After preliminary cleaning, specifically, elimination of the running titles as well as breaks in words the average size of the raw corpus was about 63 thousand words.

It is worthwhile noting that there are several ways of counting words in a corpus: if we count every word form and its every occurrence in the text, we deal with tokens; if we do not count occurrences but only word forms, we deal with types; if a headword and some of its inflected and reduced forms are counted, we deal with a lemma[33]. At the pre-annotated stage, we found it reasonable to measure the size of the database in tokens. On filtering functional words, figures and proper names, the number of tokens was nearly halved. Then using a "wordlist" tool a rank frequency list of 4500 word types was generated. The extremes of the list with the number of each word type occurrences is shown in figure 1.



Numerical order	Word type	Number of occurrences
1	people	220
2	use	183
3	water	168
4	like	167
5	years	166
1423	mouth	4
1424	wildfire	4
1425	Treaty	4
1426	climatologists	4
1427	Geert	4
1428	Jan	4
4505	umbrella	1
4506	withdraw	1
4507	blown	1
4508	two-million	1
4509	evening	1

27435 occurrences of 4509 word types out of 62767 tokens in the Corpus, including 34735 functional words occurrences;
word types frequency in the Corpus: 0.4371

Figure 1. A rank frequency wordlist from the specialized corpus

In the following, the manual sorting of topic related words was carried out. Bearing in mind that the first 1000 high-frequency English words cover about 74 %, the second 1,000 – about 7 % and the third 1,000 – about 4 % of the running words in academic texts[34], it is fair to assume that the coverage of the text by every next thousand of words decays in exponential dependence. As it is shown in the graph (figure 2) the first 1350 words of the wordlist were selected for further analysis seeing that they cover 82 % of all the texts from the corpus.

[33] Schmitt N. *Vocabulary in Language Teaching*, pp. 71–73.

[34] Nation I. S. P. *Learning Vocabulary in Another Language*, p. 20.

2023, vol. 13, issue 2 http://en.sciforedu.ru/ ISSN 2658-6762

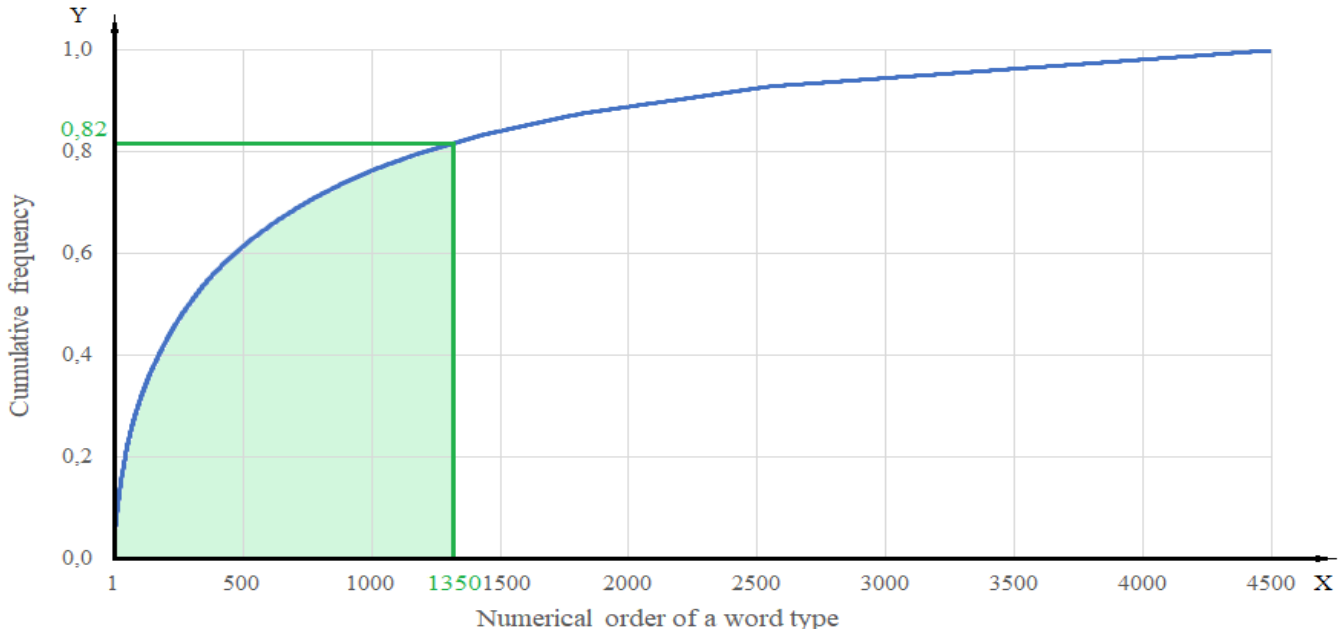


Figure 2. Word type cumulative frequency function

One more criterion for the wordlist length determination is a number of word occurrences. The borderline of the preliminary selected wordlist hits the word type with 4 occurrences in the corpus. Thus, the word list was extended up to 1442 by adding all the word types occurred four times, as in figure 3.

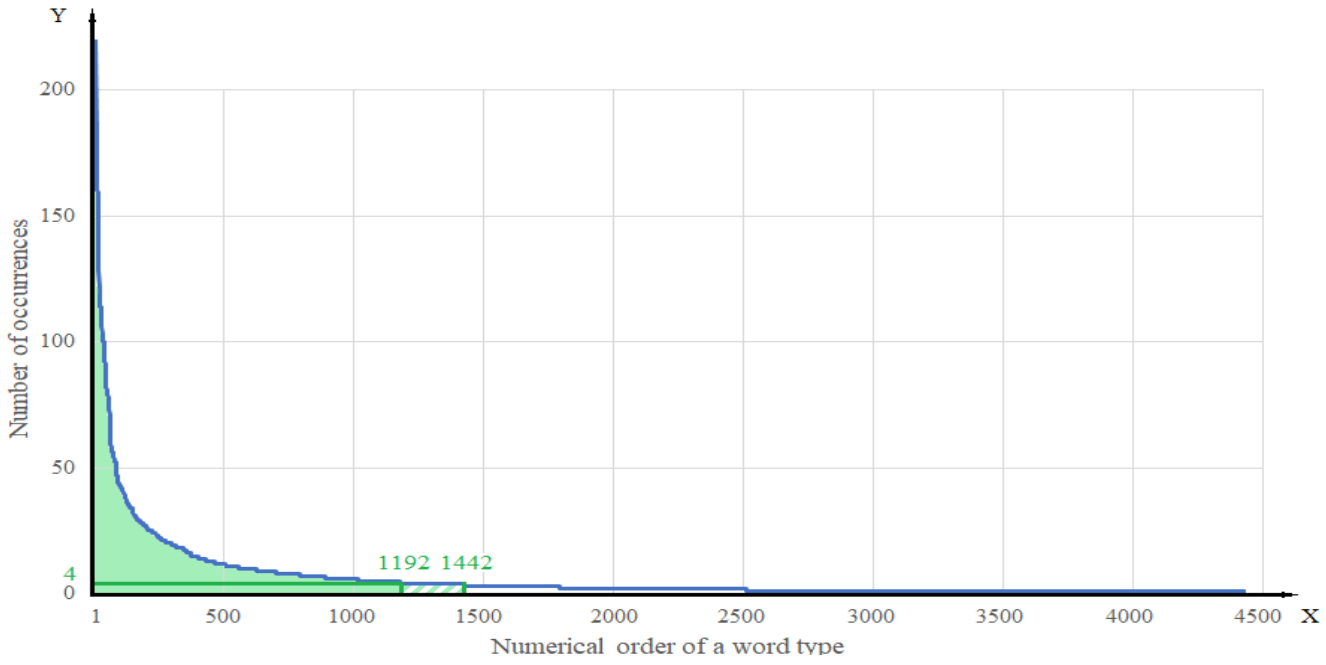


Figure 3. Number of occurrences of word types in the Corpus

At a further stage, the selected wordlist was analyzed from the point of "weather and geography" topic relevance that resulted in compiling a 237-word core topical vocabulary where the highest frequency word type "water" has 168 occurrences and the lowest frequency word type "climatologists" – 4 occurrences.

While incidental listening, it is expected that students may not understand every particular word in the text. It is the language guessing that enables the cadets to learn autonomously. In view of this we focused our attention on pre-listening stage where the language input was aimed at:

– seeing how new words are built;

– recognizing international words and pseudo international words;

– guessing the correct meaning of polysemantic words from the context;

– conversion;

– recognizing the idioms and collocations.

To check to which extent all the above-mentioned independent word learning strategies could be applicable to the semantization of new vocabulary, all the selected topic words were computer processed. With the help of a self-designed concordancer that is a software which allows a researcher search natural language text files for words, phrases and patterns all topical-word-in-context lines were analyzed. As an example, in figure 4 the results of the highest frequent word "water" processing is presented.

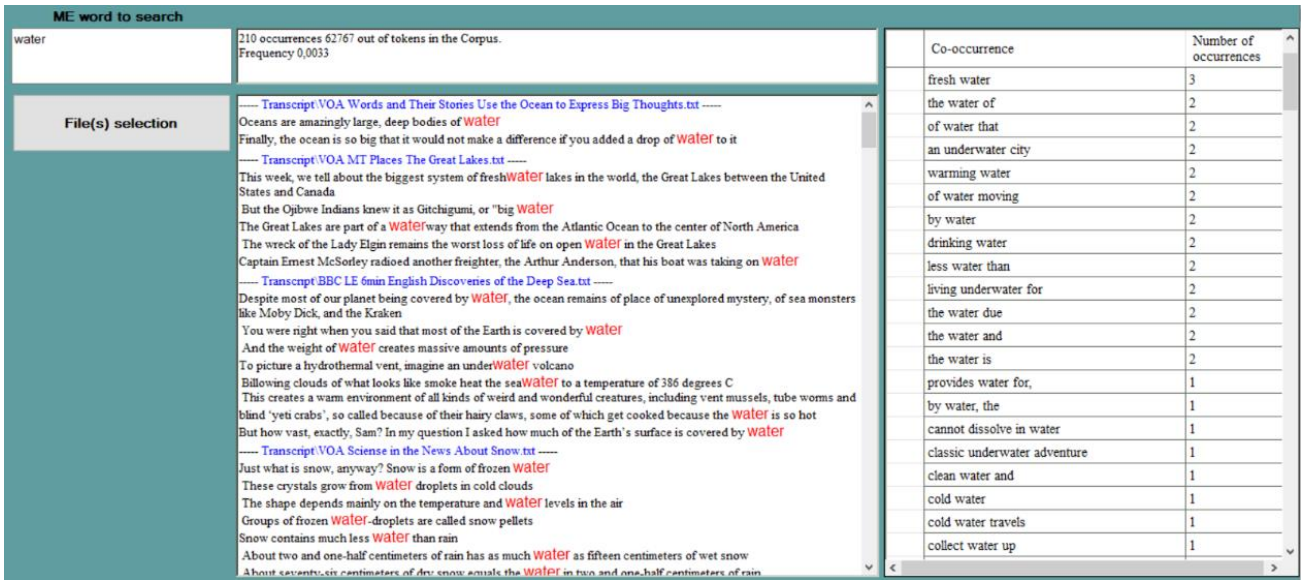


Figure 4. Concordance processing of the word type "Water"

The analysis of the "water" concordance lines shows that this word has derivatives (*underwater, waterless*), is a part of many compounds (*waterway, freshwater, seawater*) and composes a number of collocations (*body of water, water droplet, big water*), thus this lexical unit is a likely candidate to expand a cadet's potential vocabulary while applying three independent word learning strategies.

The concordance lines analysis resulted in the topical words distribution according to their semantization potential on condition of incidental learning. As it shown in figure 5, affixation and lexical chunks share the top position, while polysemantic words and true cognates (internationalisms) are less frequently occurred.

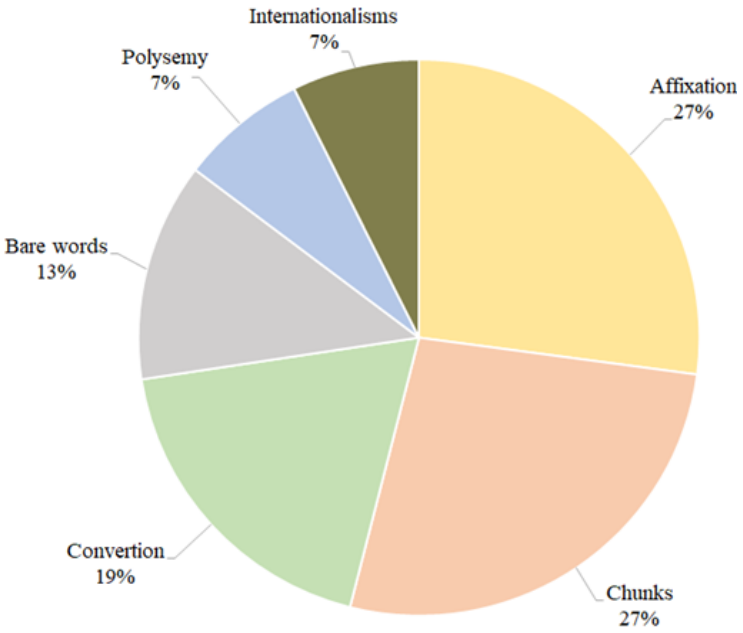


Figure 5. Topical words distribution in dependence to learning strategies

It is worth noting that a considerable number of topical words were distributed across several categories (fig. 6), thus more than 60 words can be grounds for applying two strategies, 26 words – three strategies and the words *ship* and *head* might help cadets to semantisize a number of words applying four strategies. For example, in "weather and geography" related texts *ship* is used both as a noun and a verb, is a stem of multi meaning derivative *shipping*, a part of the compound noun *shipwrecks* and the collocation *to abandon a ship*.

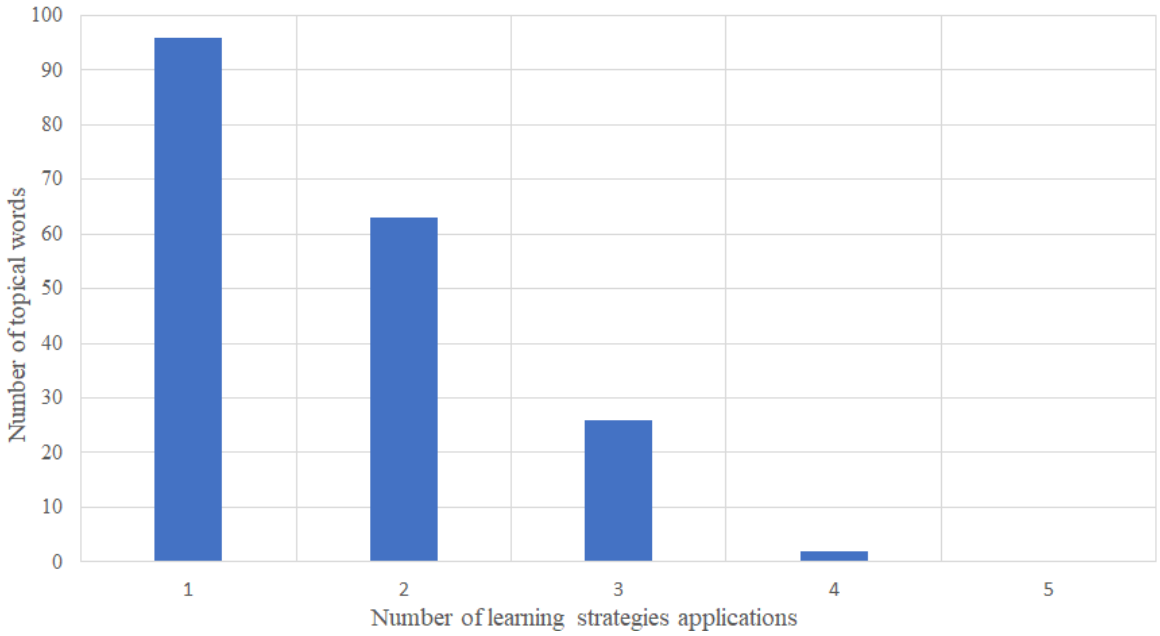


Figure 6. The distribution of the topical words across learning strategies

This outcome allows us to suggest that the more independent word learning strategies be applied the higher the potential of the unknown word be contextually guessed. It means that the acquisition of implicit vocabulary supported by the corpus-based technologies makes the process of incidental GME learning more cadet-friendly.

Conclusions

If we speak about the potential vocabulary acquisition, listening is nothing if not the best provides comprehensible input. Answering the research question what lightens the burden of trainees' incidental vocabulary learning through listening and, reasonably, enhance the efficiency of General Maritime English teaching and learning we realize the importance of keeping the balance between explicit and incidental learning. The incidental vocabulary acquisition is not an easy process which requires of the cadets the skillful application of independent word learning strategies. To lighten the burden of the process, the teacher should forecast potential difficulties in unknown vocabulary acquisition and carefully plan the meeting with the new words. And a great contributor here is the corpus-based technologies as they:

– improve the processing of natural texts,

– increase the topic relevance of the wordlist compiled,

– contextualize the target vocabulary in natural texts and by that help in determining those independent word learning strategies which most fully conform to the lexical material.

In view of the above, the study laid the focus on the *preparatory stage of the vocabulary learning process* which includes the authentic text selection for a raw context-specialized corpus design and a topical wordlist compilation. Through the computer processing of the wordlist and the analysis of the concordance lines the independent vocabulary learning strategies were identified and ranked. The conclusion was made on the dependence of a basic word learning potential and the number of independent word learning strategies to apply to its derivatives semantization. All of this justifies the potential of corpus-based technologies for arranging incidental vocabulary learning.

Thereto, the fulfilment of these preconditions is a critical stage for the extension of the marine cadets' potential vocabulary as it provides a basis for developing independent word learning strategies the use of which shifts the focus from declarative knowledge to knowledge of process character, in particular contextual guess while autonomous listening.

The main limitation of the study is that it required time-consuming manual processing of the wordlist as the compiled corpus had not been annotated but only partially cleaned. Therefore, for future research we are planning to develop the software to deal with lemmas, that will make the word-list analysis more computer-controlled. GME corpus compilation is a continuous project intending to embrace all the 1st-year GME syllabus topics. Being properly compiled and annotated the specialized corpus under research will be an asset for GME course teachers as it may contribute to the designing of the special language-focus activities and exercises which will take rightful place during explicit learning and accelerate the pace of learners' cognitive skills development emphasizing on independent word contextual guesswork. That will help junior cadets make steps towards learner autonomy, because, without direction, the learner is unlikely to be empowered.

REFERENCES

1. Zhang Y., Cole C. Maritime English as a code-tailored ESP: Genre-based curriculum development as a way out. *Ibérica*, 2018, vol. 35, pp. 145–170. URL: http://www.aelfe.org/documents/35_06_IBERICA.pdf
2. Matthews J. Vocabulary for listening: Emerging evidence for high and mid-frequency vocabulary knowledge. *System*, 2018, vol. 72, pp. 23–36. DOI: https://doi.org/10.1016/j.system.2017.10.005
3. Li Y., Zhang X. L2 Vocabulary knowledge and L2 listening comprehension: a structural equation model. *Canadian Journal of Applied Linguistics*, 2019, vol. 22 (1), pp. 85–102. DOI: https://doi.org/10.7202/1060907ar
4. Zhang P., Graham S. Learning vocabulary through listening: The role of vocabulary knowledge and listening proficiency. *Language Learning*, 2020, vol. 70 (4), pp. 1017–1053. DOI: https://doi.org/10.1111/lang.12411
5. Krasnov S. V., Kalmykova S. V., Krasnova S. A. Methodological aspects of the educational process in the context of e-learning. *Global Scientific Potential*, 2020, no. 5, pp. 23–28. (In Russian) URL: https://www.elibrary.ru/item.asp?id=43310045
6. Ermolova T. V., Savitskaya N. V., Dedova O. V., Guzova A. V. The problem of choosing universal tools for foreign language teaching in conditions of changing educational paradigms. *Science for Education Today*, 2021, vol. 11 (6), pp. 179 –194. (In Russian) DOI: http://dx.doi.org/10.15293/2658-6762.2106.10 URL: https://www.elibrary.ru/item.asp?id=47447644
7. Krylova E. A. Blended learning in higher education. *Tomsk State Pedagogical University Bulletin*, 2020, no. 1, pp. 86–93. (In Russian) DOI: http://dx.doi.org/10.23951/1609-624X-2020-1-86-93 URL: https://elibrary.ru/item.asp?id=41722368
8. Sari L. I., Sari R. H. Lecturers' challenges and strategies in teaching Maritime English online to students with low English proficiency. *Register Journal*, 2022, vol., 15 (2), pp. 222–244. DOI: http://dx.doi.org/10.18326/rgt.v15i2.222-244
9. Boers F. Cognitive linguistic approaches to teaching vocabulary: Assessment and integration. *Language Teaching*, 2013, vol. 46 (2), pp. 208–224. DOI: http://dx.doi.org/10.1017/S0261444811000450
10. James A. J., Schriever U. G., Jahangiri S., Girgin S. C. Improving maritime English competence as the cornerstone of safety at sea: a focus on teaching practices to improve maritime communication. *WMU Journal of Maritime Affairs*, 2018, vol. 17, pp. 293–310. DOI: https://doi.org/10.1007/s13437-018-0145-4
11. Dorfman O., Chernova O. Development of the communicative competence of students of technical universities as a strategy for teaching effective communication. *Litera*, 2022, vol. 4, pp. 27–35. (In Russian) DOI: http://dx.doi.org/10.25136/2409-8698.2022.4.37806 URL: https://www.elibrary.ru/item.asp?id=48309034
12. Dervić M., Bećirović S. Prerogative of the lexical approach in communicative language teaching. *European Journal of Education Studies*, 2020, vol. 7 (3). DOI: https://doi.org/10.5281/zenodo.3748039
13. Igolkina M. I., Yazynina V. S. Lexical approach to foreign language teaching in a technical university. *Humanities Bulletin of BMSTU*, 2022, no. 1, pp. 7. (In Russian) DOI: http://dx.doi.org/10.18698/2306-8477-2022-1-763 URL: https://www.elibrary.ru/item.asp?id=48415204

14. Sun B. Cognitive approach of vocabulary learning. *Proceedings of the International Conference on Arts, Design and Contemporary Education*, 2016, pp. 1517–1521. DOI: http://dx.doi.org/10.2991/icadce-16.2016.373
15. Chiew M. T. L., Ismail H. H. Exploring vocabulary learning strategies in a second language setting: A review. *International Journal of Academic Research in Business and Social Sciences*, 2021, vol. 11 (12), pp. 1298–1309. DOI: http://dx.doi.org/10.6007/IJARBSS/v11-i12/11376
16. Nekrasova-Beker T., Becker A., Sharpe A. Identifying and teaching target vocabulary in an ESP course. *TESOL Journal*, 2019, vol. 10 (1), pp. e00365. DOI: http://dx.doi.org/10.1002/tesj.365 URL: https://onlinelibrary.wiley.com/doi/full/10.1002/tesj.365
17. Yakubson V., Zakharov V. Initial step of specialized corpora building: Cleaning procedures. *NORDSCI Conference proceedings: Language and linguistics, 2020*, pp. 155–164. DOI: https://doi.org/10.32008/NORDSCI2020/B1/V3/16 URL: https://www.nordsci.org/nordsci-library/p/initial-step-of-specialized-corpora-building%3A-cleaning-procedures
18. Ahrabi Fakhr M., Borzabadi Farahani D., Khomeijani Farahani A. A. Incidental vocabulary learning and retention from audiovisual input and factors affecting them. *English Teaching & Learning*, 2021, vol. 45 (1), pp. 167–188. DOI: https://doi.org/10.1007/s42321-020-00066-y
19. Logvina S. A., Yagenich L. V., Saxno E. M. Potential teaching vocabulary forming methods retrospective analysis. *Bulletin of N.A. Nekrasov Kostroma State University: Pedagogy, Psychology, Social Work, Juvenile Studies*, 2015, vol. 21 (2), pp. 181–184. (In Russian) URL: https://elibrary.ru/item.asp?id=24095168
20. Tinkel I. The interconnectedness of mindset and learning strategies: How teaching the lexical approach could support deep learning. *MAP Education and Humanities*, 2022, vol. 2 (2), pp. 30–44. DOI: https://doi.org/10.53880/2744-2373.2022.2.2.30
21. Hussein K. S. The potentialities of corpus-based techniques for analyzing literature. *Journal of Literature, Language & Culture*, 2020, vol. 1 (2), pp. 28–43. DOI: https://doi.org/10.25255/2378.3591.2020.1.2.28.43

Submitted: 6 January 2023 Accepted: 10 March 2023 Published: 30 April 2023

The authors' stated contribution:

A. A. Mironenko
Contribution of the co-author: performing statistical procedures, software development, data processing and visualization.

E. V. Mironenko
Contribution of the co-author: organization and concept of the study, literary review, collecting empirical material, interpretation of the results and general guidance of the study, writing the text of the article.

Information about competitive interests:

The authors declare no apparent or potential conflicts of interest in connection with the publication of this article.

Information about the Authors

Aleksandr Anatolyevich Mironenko
Doctor of Technical Science, Professor,
Department of Navigation,
Admiral Makarov State University of Maritime and Inland Shipping,
5/7 Dvinskaya street, 198035, Saint Petersburg, Russian Federation
ORCID ID http://orcid.org/0000-0003-0308-4310
E-mail: mironenkoaa@gumrf.ru

Elena Valeryevna Mironenko
Candidate of Pedagogical Sciences, Associate Professor,
Department of English for Navigation and Radio Communication,
Admiral Makarov State University of Maritime and Inland Shipping,
5/7 Dvinskaya street, 198035, Saint Petersburg, Russian Federation,
ORCID ID http://orcid.org/0000-0003-1315-9697
E-mail: mironenkoev@gumrf.ru

Science for Education Today

2023. Том 13. № 2 http://sciforedu.ru ISSN 2658-6762

УДК 811+378
DOI: 10.15293/2658-6762.2302.08

Научная статья / Research Full Article
Язык статьи: английский / Article language: English

Потенциал использования корпусных технологий при автономном овладении лексикой на начальном этапе обучения морскому английскому языку

А. А. Мироненко[1], Е. В. Мироненко[1]

[1] Государственный университет морского и речного флота имени адмирала С. О. Макарова, Санкт-Петербург, Российская Федерация

Проблема и цель. *В статье рассматривается проблема расширения потенциального словаря курсантов морских учебных заведений в процессе аудирования, обусловленная потребностью формирования стратегий самостоятельного овладения лексикой, необходимых для автономного обучения морскому английскому языку. Цель статьи – обосновать целесообразность использования корпусных технологий при отборе лексического материала, что позволит обучающимся овладеть стратегиями, необходимыми для расширения словарного запаса в процессе автономного аудирования, и в целом оптимизировать процесс обучения морскому английскому языку.*

Методология. *Методологической основой исследования является методическая модель двухуровнего обучения морскому английскому языку, при которой на начальном этапе формирования коммуникативной компетенции курсантов доминируют лексический и когнитивный подходы. В статье обосновывается необходимость их интеграции при активном использовании корпусных технологий, применение которых позволяет осуществлять отбор тематической лексики с учетом дидактических, лингвистических и статистических принципов на основе актуальных аутентичных текстов.*

Результаты. *В исследовании, описывающем подготовительный этап работы по расширению потенциального словаря курсантов посредством аудирования, представлена процедура отбора текстов для специального корпуса, при помощи авторского конкордансера составлен и проанализирован тематический список слов, приведены результаты анализа конкорданса «ключевое слово в контексте», выявлена зависимость дидактического потенциала базовых слов и количества стратегий, применяемых при контекстуальной догадке значений их производных, владение которыми является обязательным условием автономного овладения лексикой.*

Библиографическая ссылка: Мироненко А. А., Мироненко Е. В. Потенциал использования корпусных технологий при автономном овладении лексикой на начальном этапе обучения морскому английскому языку // Science for Education Today. – 2023. – Т. 13, № 2. – С. 172–191. DOI: http://dx.doi.org/10.15293/2658-6762.2302.08

Автор для корреспонденции: Е. В. Мироненко, mironenkoev@gumrf.ru

Заключение. *Результаты исследования позволили определить потенциал корпусных технологий при отборе и анализе тематической лексики для овладения стратегиями, являющимися ключевыми при расширении потенциального словаря в процессе автономного аудирования, и высказать предположение о необходимости включения отобранных из специального корпуса тематических списков в лексический минимум курса общего английского языка в морском контексте, изучение которого должно осуществляться под контролем преподавателя.*

Ключевые слова: *морской английский; корпусные технологии; автономное овладение лексикой; словарный запас; специальный корпус; стратегии расширения словарного запаса.*

СПИСОК ЛИТЕРАТУРЫ

1. Zhang Y., Cole C. Maritime English as a code-tailored ESP: Genre-based curriculum development as a way out // Ibérica. – 2018. – Vol. 35. – P. 145–170. URL: http://www.aelfe.org/documents/35_06_IBERICA.pdf
2. Matthews J. Vocabulary for listening: Emerging evidence for high and mid-frequency vocabulary knowledge // System. – 2018. – Vol. 72. – P. 23–36. DOI: https://doi.org/10.1016/j.system.2017.10.005
3. Li Y., Zhang X. L2 Vocabulary knowledge and L2 listening comprehension: a structural equation model // Canadian Journal of Applied Linguistics. – 2019. – Vol. 22 (1). – P. 85–102. DOI: https://doi.org/10.7202/1060907ar
4. Zhang P., Graham S. Learning vocabulary through listening: the role of vocabulary knowledge and listening proficiency // Language Learning. – 2020. – Vol. 70 (4). – P. 1017–1053. DOI: https://doi.org/10.1111/lang.12411
5. Краснов С. В., Калмыкова С. В., Краснова С. А. Методологические аспекты образовательного процесса в условиях электронного обучения // Глобальный научный потенциал. – 2020. – № 5. – С. 23–28. URL: https://www.elibrary.ru/item.asp?id=43310045
6. Ермолова Т. В., Савицкая Н. В., Дедова О. В., Гузова А. В. Аргументация выбора универсального инструментария обучения иностранным языкам в условиях перехода к digital-компетентностной парадигме // Science for Education Today. – 2021. – Т. 11, № 6. – С. 179–194. DOI: http://dx.doi.org/10.15293/2658-6762.2106.10 URL: https://www.elibrary.ru/item.asp?id=47447644
7. Крылова Е. А. Технология смешанного обучения в системе высшего образования // Вестник Томского государственного педагогического университета. – 2020. – № 1. – С. 86–93. DOI: http://dx.doi.org/10.23951/1609-624X-2020-1-86-93 URL: https://elibrary.ru/item.asp?id=41722368
8. Sari L. I., Sari R. H. Lecturers' challenges and strategies in teaching Maritime English online to students with low English proficiency // Register Journal. – 2022. – Vol. 15 (2). – P. 222–244. DOI: http://dx.doi.org/10.18326/rgt.v15i2.222-244
9. Boers F. Cognitive linguistic approaches to teaching vocabulary: assessment and integration // Language Teaching. – 2013. – Vol. 46 (2). – P. 208–224. DOI: http://dx.doi.org/10.1017/S0261444811000450
10. James A. J., Schriever U. G., Jahangiri S., Girgin S. C. Improving maritime English competence as the cornerstone of safety at sea: a focus on teaching practices to improve maritime communication // WMU Journal of Maritime Affairs. – 2018. – Vol. 17. – P. 293–310. DOI: https://doi.org/10.1007/s13437-018-0145-4

11. Дорфман О. В., Чернова О. Е. Развитие коммуникативной компетенции студентов технических вузов как стратегия обучения эффективной коммуникации // Litera. – 2022. – № 4. – С. 27–35. DOI: http://dx.doi.org/10.25136/2409-8698.2022.4.37806 URL: https://www.elibrary.ru/item.asp?id=48309034
12. Dervić M., Bećirović S. Prerogative of the lexical approach in communicative language teaching // European Journal of Education Studies. – 2020. – Vol. 7 (3). DOI: https://doi.org/10.5281/zenodo.3748039
13. Иголкина М. И., Язынина В. С. Лексический подход к обучению иностранному языку в техническом вузе // Гуманитарный вестник МГТУ им. Н.Э. Баумана. – 2022. – № 1. – С. 7. DOI: http://dx.doi.org/10.18698/2306-8477-2022-1-763 URL: https://elibrary.ru/item.asp?id=48415204
14. Sun B. Cognitive approach of vocabulary learning // Proceedings of the 2016 International Conference on Arts, Design and Contemporary Education. – 2016. – P. 1517–1521. DOI: http://dx.doi.org/10.2991/icadce-16.2016.373
15. Chiew M. T. L., Ismail H. H. Exploring vocabulary learning strategies in a second language setting: A review // International Journal of Academic Research in Business and Social Sciences. – 2021. – Vol. 11 (12). – P. 1298–1309. DOI: http://dx.doi.org/10.6007/IJARBSS/v11-i12/11376
16. Nekrasova-Beker T., Becker A., Sharpe A. Identifying and teaching target vocabulary in an ESP course // TESOL Journal. – 2019. – Vol. 10 (1). – P. e00365. DOI: http://dx.doi.org/10.1002/tesj.365 URL: https://onlinelibrary.wiley.com/doi/full/10.1002/tesj.365
17. Yakubson V., Zakharov V. Initial step of specialized corpora building: cleaning procedures // NORDSCI Conference proceedings: Language and linguistics. – 2020. – P. 155–164. DOI: https://doi.org/10.32008/NORDSCI2020/B1/V3/16 URL: https://www.nordsci.org/nordsci-library/p/initial-step-of-specialized-corpora-building%3A-cleaning-procedures
18. Ahrabi Fakhr M., Borzabadi Farahani D., Khomeijani Farahani A. A. Incidental vocabulary learning and retention from audiovisual input and factors affecting them // English Teaching & Learning. – 2021. – Vol. 45 (1). – P. 167–188. DOI: https://doi.org/10.1007/s42321-020-00066-y
19. Логвина С. А., Ягенич Л. В., Сахно Е. М. Ретроспективный анализ методов формирования потенциального словаря // Вестник Костромского государственного университета им. Н.А. Некрасова. Серия: Педагогика, Психология, Социальная работа, Ювенология. – 2015. – Т. 21, № 2. – С. 181–184. URL: https://elibrary.ru/item.asp?id=24095168
20. Tinkel I. The Interconnectedness of Mindset and Learning Strategies: How Teaching the Lexical Approach could Support Deep Learning // MAP Education and Humanities. – 2022. – Vol. 2 (2). – P. 30–44. DOI: https://doi.org/10.53880/2744-2373.2022.2.2.30
21. Hussein K. S. The potentialities of corpus-based techniques for analyzing literature // Journal of Literature, Language & Culture. – 2020. – Vol. 1 (2). – P. 28–43. DOI: https://doi.org/10.25255/2378.3591.2020.1.2.28.43

Поступила: 6 января 2023 Принята: 11 марта 2023 Опубликована: 30 апреля 2023

Заявленный вклад авторов:

Мироненко А. А.: выполнение статистических процедур, разработка программного обеспечения, анализ и визуализация данных.

Мироненко Е. В.: организация и концепция исследования, литературный обзор, сбор эмпирического материала, интерпретация результатов и общее руководство, оформление текста статьи.

Информация о конфликте интересов:

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов в связи с публикацией данной статьи.

Информация об авторах

Мироненко Александр Анатольевич
доктор технических наук, профессор,
кафедра навигации,
Государственный университет морского и речного флота имени адмирала С.О. Макарова,
198035, г. Санкт-Петербург, ул. Двинская, 5/7
ORCID ID: http://orcid.org/0000-0003-0308-4310
E-mail: mironenkoaa@gumrf.ru

Мироненко Елена Валерьевна
кандидат педагогических наук, доцент,
кафедра английского языка навигации и связи,
Государственный университет морского и речного флота имени адмирала С.О. Макарова
198035, г. Санкт-Петербург, ул. Двинская, 5/7
ORCID ID: http://orcid.org/0000-0003-1315-9697,
E-mail: mironenkoev@gumrf.ru