Yaskevich Daniil Mikhailovich
student
Gladyshev Aleksandr Olegovich
student
Sokolov Daniil Antonovich
student
Misolin Timofey Iurevich
student
Scientific adviser
Ioffe Nina Evgenyevna
senior lecturer
MIREA – Russian Technological University
Moscow

TEACHING READING OF ENGLISH TECHNICAL DOCUMENTATION AS AN INNOVATIVE METHOD: FROM THE LANGUAGE BARRIER TO SUSTAINABLE SKILLS

Abstract: *IT students have to read technical documentation in English, but general English courses do not prepare them for its vocabulary and structure. In the study, the documentation of six open-source projects (1.32 million words) was analyzed to determine how much of its vocabulary is covered by general and technical word lists. The 2,000 most frequent general English words covered only 61–75% of the running words, while 300 frequent technical words added 12–14 percentage points, more than 3,000 additional general words. On this basis, a staged method of teaching documentation reading is proposed.*

Keywords: *English for specific purposes, technical documentation, reading skills, vocabulary coverage, IT students, teaching method, open-source software.*

Яскевич Даниил Михайлович
студент
Гладышев Александр Олегович

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

студент
Соколов Даниил Антонович
студент
Мисолин Тимофей Юрьевич
студент
Научный руководитель
Иоффе Нина Евгеньевна
старший преподаватель
ФГБОУ ВО «МИРЭА – Российский технологический университет»
г. Москва

ОБУЧЕНИЕ ЧТЕНИЮ АНГЛОЯЗЫЧНОЙ ТЕХНИЧЕСКОЙ ДОКУМЕНТАЦИИ КАК ИННОВАЦИОННАЯ МЕТОДИКА: ОТ ЯЗЫКОВОГО БАРЬЕРА К УСТОЙЧИВЫМ НАВЫКАМ

Аннотация: *студентам IT-направлений приходится читать техническую документацию на английском языке, однако курсы общего английского не готовят их к её лексике и структуре. В исследовании проанализирована документация шести проектов с открытым исходным кодом (1,32 млн слов), чтобы определить, какую долю её словоупотреблений покрывают общие и технические списки слов. 2000 наиболее частотных слов общего английского покрыли лишь 61–75% словоупотреблений, тогда как 300 частотных технических слов добавили 12–14 процентных пунктов – больше, чем 3000 дополнительных общеупотребительных слов. На этой основе предложена поэтапная методика обучения чтению документации.*

Ключевые слова: *английский язык для специальных целей, техническая документация, навыки чтения, лексическое покрытие, студенты IT-направлений, методика обучения, открытое программное обеспечение.*

Introduction.

For a future software developer, the ability to read documentation in English is a professional necessity: the documentation of almost all programming languages,

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

frameworks and tools is written in English and is updated faster than any translation. Many students, however, face a language barrier: they know general English but get lost in texts full of technical terms, code fragments and instructions. The theory of English for specific purposes holds that teaching should start from the needs of the learner and the texts that he or she will actually use [4]. Research on reading shows that a learner needs to know about 95–98% of the running words of a text for adequate comprehension [3; 5], and specialized word lists, such as the Academic Word List, help to reach this level faster than general vocabulary [1].

The aim of this study is to measure the lexical and structural features of real technical documentation and, on this basis, to propose a staged method of teaching students to read it.

Materials and methods

The material was the documentation of six open-source projects downloaded from their public repositories on GitHub: the Flask web framework, the book «The Rust Programming Language», Visual Studio Code, the Requests library, Git and Docker, in total 1,686 files and 1.32 million words. Code blocks, inline code, links and markup were removed before the analysis. Lexical coverage was calculated as the share of running words that belong to a word list. General vocabulary was represented by the 1,000–5,000 most frequent word forms of the Brown Corpus [2]. A list of technical vocabulary was compiled from three projects (Flask, the Rust book and Visual Studio Code): it included the 300 most frequent words that were not among the 2,000 most frequent general words and occurred in all three projects. The list was then tested on the other three projects. In addition, the mean sentence length, the share of imperative sentences and the frequency of the pronoun *you* were calculated.

Results.

The main results are given in Table 1. The documentation turned out to be written in rather long sentences (19–27 words on average) and addressed directly to the reader: the pronouns *you* and *your* occurred 12.5–25.7 times per 1,000 words, and

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

5–17% of the sentences were instructions beginning with an imperative verb (*install*, *run*, *open*, *add*).

The 2,000 most frequent general words covered only 61–75% of the running words. Adding 3,000 more general words increased the coverage by 9–11 percentage points, whereas adding 300 technical words increased it by 12–14 points. This effect was also observed for the three projects that were not used to compile the list (for example, from 66.8% to 79.3% for Git). The technical list consisted mostly of everyday words of software work, such as *code*, *file*, *settings*, *configuration*, *default*, *folder*, *terminal*, *debug*, *server*, *environment*, *container*, *install*, *repository*, *variable* and *path*.

Table 1

Lexical coverage of technical documentation by general and technical word lists

Project	Words, thousand	Sentence length, words	2,000 general words, %	5,000 general words, %	2,000 general + 300 technical words, %
Flask	50.0	21.5	68.5	77.8	82.8
Rust book	157.7	24.4	74.8	83.8	85.8
Visual Studio Code	470.5	25.1	64.5	76.4	80.5
Requests	8.7	19.1	71.7	80.2	83.9
Git	132.4	27.1	66.8	75.7	79.3
Docker	501.0	24.8	61.4	72.5	74.2

Discussion.

The results explain the language barrier: even a student who knows the 2,000 most common words understands only two thirds of the words in a documentation page, which is far below the 95–98% threshold [3; 5]. At the same time, the gap is closed most efficiently by a small technical lexical minimum rather than by general vocabulary. On this basis, we propose a staged method. At the first stage, students learn a technical minimum of about 300 words with their word families and typical collocations (*run a command*, *set a variable*). At the second stage, they learn the structure of documentation: headings, code examples, notes and links, and the role of imperatives and the pronoun *you*. At the third stage, reading becomes task-based: students install, configure or run a program by following the documentation, so that

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

comprehension is checked by a working result. At the fourth stage, they learn strategies for the remaining unknown words, i.e. guessing from code and context and using glossaries. Finally, students read the documentation of a project of their choice independently and keep reading logs. The innovative character of the method lies in the use of authentic materials, a vocabulary minimum derived from data and practical tasks with a verifiable outcome. The study is limited by the size of the sample and by the use of word forms rather than word families; the effect of the method has to be tested in teaching practice.

Conclusion.

Technical documentation is lexically difficult for students who know only general English, but a technical minimum of about 300 words raises the coverage of running words more than 3,000 additional general words. A staged method that combines this minimum, knowledge of the structure of documentation and task-based reading can help students to move from the language barrier to sustainable professional reading skills.

References

1. Coxhead, A. A new academic word list / A. Coxhead // TESOL Quarterly. – 2000. – Vol. 34, No. 2. – P. 213–238.

2. Francis, W. N. Brown corpus manual / W. N. Francis, H. Kučera. – Revised ed. – Providence (RI) : Brown University, 1979.

3. Hu, M. Unknown vocabulary density and reading comprehension / M. Hu, I. S. P. Nation // Reading in a Foreign Language. – 2000. – Vol. 13, No. 1. – P. 403–430.

4. Hutchinson, T. English for specific purposes: a learning-centred approach / T. Hutchinson, A. Waters. – Cambridge : Cambridge University Press, 1987. – 183 p.

5. Nation, I. S. P. How large a vocabulary is needed for reading and listening? / I. S. P. Nation // The Canadian Modern Language Review. – 2006. – Vol. 63, No. 1. – P. 59–82.

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)