CORPUS-BASED STUDIES OF LINGUISTIC VARIATIONS ACROSS THE ENGINEERING DISCIPLINES

T.B. Maslova

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

Engineering is a varied and diverse field of professional occupation, and has attracted attention of linguists as a realm of the English language of science and technology (EST) for many years. The most recent explorations of corpus data allows getting not only a better understanding of how the language of engineering differs from that of humanities and social sciences, but also what special features are typical of certain engineering sub-disciplines.

Investigating the use of multiple word bundles encountered in academic genres of Biology and Electrical Engineering, representing the applied and pure sciences, and Applied Linguistics and Business Studies, associated with the social sciences, Hyland (2008) calculated that over half of the extended collocations in each discipline did not occur in the other subject areas examined. In addition to different frequencies, the corpora showed that the principal structures of bundles also differed across fields. For example, 4-word bundles like "*as shown in figure*" or "*it can be seen*" appeared to be unique to the Electrical Engineering sub-corpus, which might be explained by the relatively abstract and graphical nature of technical communication. In fact, the rhetoric features of engineering discourse seem to rely on visual and numeric data representation. As arguments in engineering field are based on plausible interpretations of data, they ultimately rest on findings that are often presented in visual form, and in their turn the data or findings are linked in routinely patterned, almost formulaic passive ways. It is interesting, however, that engineers also tended to employ more examples of the anticipatory "*it*" pattern to foreground the writer's evaluation without explicitly identifying its source. Finally, electrical engineering texts were reported to contain the greatest range of bundles, namely 213 different 4-word bundles across 10% of texts, and the highest proportion of words occurring in 4-word bundles.

The results of creating a more field -specific Engineering English Word List, based on 100 textbooks across 20 engineering subject areas, led to believe that there are 5,000 most frequent word families as well as proper nouns, apparent compounds and abbreviations. Civil and mechanical engineering corpora were found to involve 3,500 word families, while marine and biochemical engineering ones were the most demanding in lexis, achieving a vocabulary threshold of 8,500 word families, with the lexical distribution thus being the most diverse. The difference between 3,500 and 8,500 word families widely occurring within the subfields of engineering indicated the different vocabulary demands for engineers majoring in different specialties. In particular, frequent highly-technical vocabulary that is required for some engineers to comprehend specialized texts may be irrelevant for others, and the English textbooks on the engineering sub-fields that entail life- or biology-related content appear to be more difficult to read (Hsu, 2014).

A recent corpus-based investigation of the 8 million word specialized corpus of civil engineering research articles conducted in 2013-2014 by Japanese scientists comprised 11 sub-disciplines of civil engineering and displayed a remarkable heterogeneity of keywords,

proving the fact that despite the existence of the core academic vocabulary, which is common to a wide range of scientific disciplines, some items on academic English word lists can often vary across disciplines in terms of range, frequency, collocation and meaning. It was shown that only 35.9% of specialized word families tend to reoccur in two or more sub-disciplines of civil engineering, this eventually resulting into identifying “softer” and “harder” keyword clusters of civil engineering. The former, related to transport, infrastructure and international projects, was argued to be of a broader nature, which allows using less amounts of specialized vocabulary, and requires more general English language items. “Harder” subject areas, including water resources and construction materials, on the other hand, demonstrated a larger number of the most frequent discipline-specific words, which were divided into five categories, comprising the language means of expressing cause and effect, comparison and contrast; quantification, as well as deictic words, and discourse markers that reveal the writer’s stance (Gilmore & Millar, 2018).

Another corpus-based study dealt with the unique features of English relative clauses used in engineering journal papers of three different sub-fields. In terms of the frequency of relative clauses, life science and biology corpus ranked first, followed by the other two, i.e. electrical and electric engineering (EEE) and chemical engineering corpora, respectively. Across all the corpora, the restrictive type of relative clause that concerns the object of the main clause was found to be more frequent, with “*that*” being typically preferred over “*which*”. At the same time, the language of EEE, which tends to present many unfamiliar, new concepts and findings, and explain these as accurately and extensively as possible, revealed a high ratio of non-restrictive relative clauses, especially of “*which*+ verb” type, which is likely to be effective for delivering detailed and additional information in a clear and simple manner. Also, EEE corpus presented the highest number of “preposition+*which*” type of restrictive relative clauses, used to describe particular situations (Cho & Lee, 2016).

As you can see, there are essential variations in the frequency and types of forms, structures and functions of the English language used in subfields of engineering, caused by specific social and rhetorical practices of academic communities, where interlocutors draw on different resources to develop arguments, establish credibility and persuade the audience. Therefore, corpus-based researches into the linguistic nature of engineering discourse can be extremely helpful for guiding the development of discipline-specific language learning materials, and thus may greatly contribute to effective EST course design practices.

References:

1. Cho, D. W. & Lee, K. (2016). English relative clauses in science and engineering journal papers: A comparative corpus-based study for pedagogical purposes. Ampersand, 3, 61-70.

2. Gilmore, A. & Millar, N. (2018). The language of civil engineering research articles: A corpus-based approach. English for Specific Purposes, 51, 1–17.

3. Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks for EFL undergraduates. English for Specific Purposes, 33, 54–65.

4. Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. English for Specific Purposes, 27, 4-21.