

Corpus-based vocabulary lists in military to learn English for the STANAG 6001

Sayaka Kamio

National Defense Academy of Japan, Japan

National Institute of Technology, Japan

Keywords: curriculum and material development, English proficiency, STANAG 6001, vocabulary learning

Background: Vocabulary learning is one of the most essential factors to improve the reading proficiency. To assess the English vocabulary and reading proficiency in non-native English speaking military personnel, the North Atlantic Treaty Organization defines the STANAG 6001. In the STANAG 6001, three levels (i.e. Level 1, 2 and 3) requires the reader to understand particular sources, which widely range from general English (e.g. advertisement) to genre-specific English (e.g. articles in the Economist). However, it has not fully explored which vocabularies are frequently used in the particular resources of each levels. In addition, further research is needed to identify frequent words commonly appeared in the various texts in all levels.

Purpose: The aim of this study is to develop two kinds of vocabulary list: (1) general

vocabulary list which presents high frequency words covering all reading proficiency levels of the STANAG 6001 (hereafter the General STANAG 6001 vocabulary list) and (2) three sub vocabulary lists which shows discipline-specific vocabularies corresponding to the Level 1, 2, and 3 of the STANAG 6001 (hereafter, the Level 1-, Level 2-, and Level 3 sub list, respectively). Using the lists, this study is also designed to compare the coverage with the previously developed wordlists such as the Academic Word List.

Method: Firstly, the corpus data to extract the vocabulary lists were established. To build the corpus data, the English texts were collected based on the criterion of the STANAG 6001. The data was gathered from various resources of the latest official online sources: webpages of the Ministry of Defense in UK and USA, major broadcasting companies (e.g. BBC) and magazine (e.g., The Economist). After collecting data, a corpus which include various resources of all levels of the STANAG 6001 was established (hereafter the General STANAG corpus). Next, using the same data of the General STANAG corpus, the three types of sub corpus were built which corresponded to the STANAG 6001 level 1, 2, and 3 (hereafter the Level 1- , Level 2-, and Level 3 sub corpus, respectively). With the corpus and Keyword Analysis programs AntConc, the General STANAG 6001 vocabulary list and three sub word lists were

extracted using three methods: vocabulary classification, keyword analysis, and term extraction. Furthermore, the coverages of how the General- and sub STANAG vocabulary lists covers the academic word list (e.g. Academic Word List) and English for specific purposes (ESP) words lists were calculated.

Results and Discussion: This research is now ongoing and the detail of results and further discussion will be presented at the upcoming ISMS conference. Firstly, as for the main purpose of this study, vocabularies that frequently used in the reading resources across all levels of the STANAG 6001, as well as the words in each level, will be presented with both frequency and range as the criteria for selections. Secondly, compared to the Academic Word List and ESP vocabulary lists, the results of how the General- and sub STANAG vocabulary lists cover the other lists. After presenting main findings, the research and pedagogical implications will be also discussed for the researchers, instructors, and curriculum and materials developers in military. As for the research implications, the limitations of the free Keyword Analysis program (e.g. the absence of word class) and the strategy of how we can solve the problem will be explained. The pedagogical implications, which suggest further vocabulary learning with the General- and sub STANAG vocabulary lists, will be also presented. (567 words)