

論文の英文要旨

論文題目	Creating a corpus-based educational word list for learners of Japanese
氏名	Yukari HONDA

The purpose of this research was to create a corpus-based educational word list (hereinafter referred to as, “the List”) for learners of Japanese. This word list is a receptive vocabulary list used to read written Japanese text. Word importance in learning Japanese was quantified by dispersion and utility measurements. The level of the words in the List was determined by the rank of word importance and word-familiarity which indicates the difficulty of each word. Many word lists have been created for Japanese language learners based on some small scale vocabulary surveys, with selection of the words based upon professionals’ subjective opinions. However, the use of the lists in Japanese education and research has been debatable. To overcome the issue of subjective selection of words for word lists, the development of a corpus-based word list was started by researchers after the Balanced Corpus of Contemporary Written Japanese (BCCWJ) was released in 2009 from the National Institute for Japanese Language and Linguistics (NINJAL). However, in the research for corpus-based word lists, whether the corpus balance was appropriate for the purpose of the word list, or not, was not taken into consideration, although it directly affects word frequency. In addition, the ranking of words in corpus-based research was based on professionals’ subjective opinions or references to the ranking of word list which was made by professionals. This practice implies that, even in the corpus-based research, professionals’ subjective opinions were still influential.

Thus, this research helped to make a receptive vocabulary list for Japanese language learners using an adjusted corpus. The words were selected by dispersion and utility measurements. As a result, text coverage of the vocabulary in the List was high for both text used by native speakers and text written for Japanese language learners. In addition, the high text coverage of the List was not only for written Japanese, but also for spoken Japanese.

The following is the outline of this research. The purpose of the research is explained in Chapter One. The words in the List were selected and ranked using

statistical measurements; the ranks were adjusted for Japanese language education. No previous research had been applied to this research.

An overview of the previous studies is presented in Chapter Two. Traditionally, the word lists were made by using professionals' subjective selections which referred to small-scale vocabulary surveys. This type of word list development was started after World War II. Contemporary Japanese vocabulary surveys have been conducted since the 1950s, and they have been used to develop traditional word lists. The word list for the Japanese Language Proficiency Test (JLPT) is the most frequently used list. Although the word list for the JLPT was made specifically for the JLPT, it is also used in research and teaching in ways that are not consistent with its original purpose. Using the JLPT word list for a purpose other than the JLPT has been common, but problematic. Moreover, some previous research has shown that the matching rates of the traditional word lists are not high. It means, the concept of "basic vocabulary" is ambiguous, even among professionals. As long as the word selection for word lists is dependent upon professionals' subjectivity, the matching rates of the word lists will remain low and the concept of "basic vocabulary" will remain ambiguous.

On the other hand, it is not uncommon to make corpus-based word lists for the TESOL. That is why dispersion measurements which shows the stability of word frequency among different sub-corpus groups is well developed for the TESOL. As in the preceding example for the TESOL, corpus-based word lists are also being developed for Japanese language research.

Chapter Three describes the research method. First, the design of the List is shown, including the total number of words, the target learners, and the usage. Second, the process of investigation and optimization of the corpus for the purpose of this research is explained. Third, the process of making the frequency list based on the optimized corpus and quantifying the importance of each word using dispersion measurements and utility measures is explained. The rank adjustment method in light of Japanese language education is also explained. Finally, the method of evaluation of the List is shown. An investigation of the text coverage was done for the evaluation.

The process of developing the List is presented in Chapter Four. First, a corpus was selected and optimized. The BCCWJ consists of 13 different text genre. There are not many high frequency words that are common across the different genres. In the following six genres, namely parliament proceedings, (government) white papers, public relations magazines, Yahoo! "Chiebukuro", school textbooks, and best sellers, many specialized vocabulary were found. Therefore, in the adjusted corpus in this

research, these six genres of text have been reduced for making the word list. As a result, the size of the adjusted corpus became 95 million words. The adjusted corpus is smaller than the original BCCWJ, yet still the biggest Japanese corpus for making word lists. Second, a frequency list was produced, and the dispersion and utility of the vocabulary was calculated. DP (Gries, 2008) was used as a dispersion measurement. The utility measure was calculated by multiplying the dispersion by the frequency. Words were ranked, based on this utility measurement. This word rank was adjusted by word-familiarity which has a numerical value known to be related to word difficulty in learning the Japanese language. The word list was classified into five levels, namely the 2,000 word level, the 4,000 word level, the 6,000 word level, the 8,000 word level and the 10,000 word level.

The evaluation of the List is presented in Chapter Five. A text coverage investigation was made of this research as an evaluation of the List. Text written for Japanese language learners and text written for Japanese native speakers were used for the investigation. As a result, the coverage was high in all the texts. For example, the coverage of the JLPT reading text was 93%, some reading texts written for intermediate level and advanced level Japanese language learners was 91%, and text for native speakers such as from newspapers, novels, web texts was at more than 85%. In addition, compared to the JLPT word list, which is very popular for Japanese language teaching and research, the List marked higher for text coverage. It shows the word selection of the List was appropriate for learning to read Japanese text.

Chapter Six is the conclusion. This research was used to develop a corpus-based word list for Japanese language learners using a new method. Words for the List were selected objectively, utilizing statistical analysis, the resultant List includes words that are appropriate for reading various kinds of Japanese text. This word list can be used for vocabulary research, and making teaching materials and tests. Further analysis is necessary in the areas of compound words, writing systems, adjustment from the point of Japanese language education, and word ranking of middle and low frequency words. Research on the matching rate for the List with other corpus-based Japanese word lists would also be useful.