The Development of a『Dictionary of the Usage Frequency of Basic Korean Vocabulary』 and Its Application to Korean Lexicology

Sang-Kyu Seo**, Jinung Kim***

1. Introduction

The purpose of the paper is to show the methodologies of a quantitative analysis of data applicable to corpus linguistics, lexicography, and Korean language and information by elucidating the process of developing a『Dictionary of Usage Frequency of Basic Korean Vocabulary』 (2014). We carried out the process of extracting information from the corpora and reshaped them into a body of knowledge applicable to linguistic education and research. The number of word forms defined in the dictionary is 7,203(including 58 affixes), which covers 14.9% of 48,281 content words from Korean language education corpus(henceforth Kec). Interestingly, the number of word tokens of only 7,203 words is 803,820 which cover 76.4%. In other words, 76.4 % of Kec is annotated with word senses. In particular, this paper deals with the following topics.

1) A detailed introduction of the background, structure, and characteristics of a『Dictionary of Usage Frequency of Basic Korean Vocabulary』, which has been compiled from 1998 to 2014
2) Methodologies in selecting basic vocabulary and research on usage frequency

* This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2009-361-A00027).
** Yonsei University, first author
*** ILIS, Yonsei University, corresponding author
The Development of a "Dictionary of the Usage Frequency of Basic Korean Vocabulary"

3) Challenges faced while compiling the dictionary and their solutions
4) Questions that arose from the compilation of the Dictionary

2. What is the need for a "Dictionary of Usage Frequency of Basic Korean Vocabulary"?

There are several words which have a singular meaning, but in most cases words are used with multiple meanings. Such polysemic characteristics are more prominent in words which are frequently used in everyday life (or, with very high usage frequencies). The meanings of a word can be defined in a dictionary, and through this we can understand the various meanings of a word. However, it is difficult to know which meaning is more important or more common. In such situations, we can only rely on experiential judgment or instinct. Generally, when defining a dictionary meaning, the central (or etymological) meaning is placed first, after which the scope of the meaning is widened to peripheral (expanded) meanings. Below is an example from the Korean Standard Dictionary.

(1) Sichimi
   ① quadrilateral horn tied inside the tail feathers of a falcon with written address to declare the owner of the falcon = danjangpan
   ② the attitude which pretends not to have done something when they did, or feigning ignorance (Korean Standard Dictionary, 1999)

---

1 To examine the aspects of the homograph entries and senses of words, 300 words with the highest frequencies in the Vocabulary Frequency of Modern Korean Language (1, 2) (Seo, 1998a) as investigated from the Yonsei corpus (approximately 40,000,000 mati) were studied to determine how many homograph entries and senses were defined in the Yonsei Korean Dictionary (1998)(henceforth YKD). The results showed that one word had average of approximately 1.7 homograph entries of the same PoS (approximately 3.5 when including different PoSs), and an average of 7.2 senses. Moreover, when the entries in a "Dictionary of Usage Frequency of Basic Korean Vocabulary" were analyzed, 7,203 homograph entries were included in the total of 5,162 entries, resulting in an average of 1.4 homograph entries.