Variation across Korean text registers

Beom-mo Kang and Hung-gyu Kim
Korea University

Our paper describes the results of an adoption of Biber's (1988, 1995) multivariate statistical analysis of text registers to the analyses of Korea text registers and styles, comparing our results with Kim's (1990). Here, registers are conceived of as kinds or types of texts with respect to text/discourse situations and are not distinguished from genres in a broad sense. The text corpus on which our study is based is composed of 334 text samples of 36 registers/genres of Korean texts. Each text sample is a computerised one and consists of about 1,000 words, the total size of the corpus being about 370,000 words. Among 13 spoken text registers, we included not only transcriptions of real conversations but also some texts written to be spoken--transcripts of TV dramas, plays, and movies. Among 23 written text registers, we included various kinds of informative and imaginative ones--newspaper reports, editorials, essays, encyclopaedia, academic, informative books, etc. For the multivariate statistical analyses, we counted the occurrences of 82 linguistics features of each text samples and performed factor analysis, cluster analysis, and canonical discriminant analysis. Before we counted linguistic features occurring in texts, we had to perform morphological tagging because many of the linguistic features could only be counted from morphologically analysed texts. We used automatic morphological analyser first, and then checked and corrected the tags manually.

By multidimensional factor analysis using SAS, we found that there are six dimensions of Korean texts: D1. Informal interaction vs. Elaboration; D2. On-line/situated production vs. Informative content; D3. Narrative vs. Abstract concerns; D4. Formal statement of opinions; D5. Public report; D6. Public mention of modern things. By cluster analysis, we found that there are eight text types which are solely based on linguistic characteristics and which are different from conventional genres or registers. These eight text types are: T1. Abstract content; T2. Modelled informal interaction; T3. Event description; T4. Opinion; T5. Planned presentation of information; T6. On-line production; T7. Explanation of facts and procedures; T8. Public report. Our analysis is an application of Biber's multivariate statistical analysis to Korean texts. The results show that this method is a fruitful one in investigating aspects Korean text registers and styles, and reveal similar as well as different textual characteristics of English and Korean.

References

Biber D 1988 *Variation across Speech and Writing*. Cambridge, CUP.
Biber D 1995 *Dimensions of Register Variation: A Cross-linguistic Comparison*. Cambridge, CUP.
Kim Y 1990 Register Variation in Korean: A Corpus-Based Study, Ph.D. thesis, USC.