A Study of Text Types and Register Variation in the British National Corpus

TAKAHASHI, Kaoru

Toyota National College of Technology

takahasi@toyota-ct.ac.jp

The paper is concerned with text typology. The British National Corpus (World Edition), which is a one hundred million-word collection of spoken and written British English, will be used to study the characterization of text types by identifying their linguistic characteristics. By means of multivariate analysis, the variation of the occurrence of selected linguistic features among genre categories will be classified. A multivariate analysis of this sort holds out the promise of being able to systematize the genre categories in the corpus while also revealing the characteristic linguistic features of the groups classified. The criteria for classification will be based exclusively on the dimensions revealed as significant by the multivariate analysis. The groupings will then be interpreted linguistically.

In the research I also wish to focus on what kind of role language variation takes in the study of linguistics. Therefore, this paper deals with the study of sociolinguistic variables which are fairly systematically associated with 'region', 'class', 'ethnic group', 'age', and 'gender' and so on. The aim of an analysis of these variables is to yield a more sophisticated classification and systematization of textual sub-categorization in the British National Corpus. The groupings will provide sociolinguistic interpretations.