

Developing a Corpus for Learners of Korean as a Foreign Language in Sri Lanka

Nisansala Pathirana¹
Jin-Ryang Kim²

INTRODUCTION

This study aims to establish a detailed plan and suggest guidelines for developing an error corpus of Sinhala native speakers learning Korean as a foreign language. Previous studies and research conducted on learner corpus in Korea were analyzed and a key point that needs to be considered in each stages of the corpus construction was studied at length.

BACKGROUND

A Korean learner corpus can be used as a resource to support language education policy and to throw light on areas of difficulty in learning Korean as a foreign language by Sinhala native speakers. Also, the corpus can be utilized as teaching materials for students, and have a significant impact as a fundamental resource on research on Korean language education in Sri Lanka. Korean as a Foreign Language was introduced to University of Kelaniya in 1997 and the number of students offering the language as a main subject for the Bachelor Degree has continuously increased over the years. Hence, it is timely that steps are taken to develop a learner corpus.

APPROACH

Successful language learning could be achieved by error identification and correction. Correcting errors collected in scientific method from extended cases could facilitate the improvement of the learners' skill. Initially, common errors need to be identified. Thereafter, efforts should be exercised to identify the situations leading to errors and to ascertain the causes for errors. Gathering of such data is not an easy task. This difficulty emerges especially in collecting data regarding errors of spoken discourse.

RECOMMENDATIONS

This paper suggests the following on developing a learner corpus based on the results of the research conducted: First, in the designing stage, providing guidelines for the corpus based on the studies on secondary data is pivotal. Thereafter, data collection for the corpus will be implemented where the sample will be taken from students of Korean as a foreign language at the university. In this research the target words and sentences for voice recording was suggested by after careful consideration of selected seriously considering phonological variations in Korean.

Second, in the constructing stage, a multifaceted corpus that applies multimedia including as videos, spoken and written discourse will be pursued. The corpus planned in this research includes various types of data from text to video images. Usually 25% of spoken language data is recommended for 75% of written language data.

Third, taking the reliability of resource into account is also an important point. Usually the appropriate size of a corpus should be at least four million words. Considering the number of Korean language learners in Sri Lanka, several years will be required to achieve such a target. Therefore, the corpus should be updated every year, and an integrated management system should be prepared for maintaining construction of a consistent corpus.

CONCLUSION

The efforts made for the first time at the Centre for Korean Studies of the University of Kelaniya with the objective of building up of error corpus of Sri Lanka students learning Korean language will be completed as a long term plan. However, as a pilot study a corpus comprising 100,000 words will be conducted within a period of one year.

KEYWORDS

Korean learner corpus, error corpus, Korean language education, Korean as a Foreign Language in Sri Lanka, error analysis

¹Korean Studies, Department of Modern Languages, University of Kelaniya, Sri Lanka. pathi.0348@gmail.com

²Korean Studies, Department of Modern Languages, University of Kelaniya, Sri Lanka. julian_lanka@yahoo.com