A Corpus Output Error Analysis of Chinese L2 Learners From America, Myanmar, and Singapore

Authors : Qiao-Yu Warren Cai

Abstract : Due to the rise of big data, building corpora and using them to analyze ChineseL2 learners' language output has become a trend. Various empirical research has been conducted using Chinese corpora built by different academic institutes. However, most of the research analyzed the data in the Chinese corpora usingcorpus-based qualitative content analysis with descriptive statistics. Descriptive statistics can be used to make summations about the subjects or samples that research has actually measured to describe the numerical data, but the collected data cannot be generalized to the population. Comte, a Frenchpositivist, has argued since the 19th century that human beings' knowledge, whether the discipline is humanistic and social science or natural science, should be verified in a scientific way to construct a universal theory to explain the truth and human beings behaviors. Inferential statistics, able to make judgments of the probability of a difference observed between groups being dependable or caused by chance (Free Geography Notes, 2015)and to infer from the subjects or examples what the population might think or behave, is just the right method to support Comte's argument in the field of TCSOL. Also, inferential statistics is a core of quantitative research, but little research has been conducted by combing corpora with inferential statistics. Little research analyzes the differences in Chinese L2 learners' language corpus output errors by using theOne-way ANOVA so that the findings of previous research are limited to inferring the population's Chinese errors according to the given samples' Chinese corpora. To fill this knowledge gap in the professional development of Taiwanese TCSOL, the present study aims to utilize the One-way ANOVA to analyze corpus output errors of Chinese L2 learners from America, Myanmar, and Singapore. The results show that no significant difference exists in 'shì (是) sentence' and word order errors, but compared with Americans and Singaporeans, it is significantly easier for Myanmar to have 'sentence blends.' Based on the above results, the present study provides an instructional approach and contributes to further exploration of how Chinese L2 learners can have (and use) learning strategies to lower errors.
Keywords : Chinese corpus, error analysis, one-way analysis of variance, Chinese L2 learners, Americans, myanmar, Singaporeans