

This study replicates Linlin Sun & David C. Saavedra's research on determining the grammatical status of units in Chinese using quantitative methods. Following their work, we use a binary logistic regression model to calculate the grammatical status scores of selected lexical units. As a source of data, the Lancaster Corpus of Chinese (LCMC), which contains contemporary standard Chinese texts, is used.

In this paper, we detail the metrics and modeling approach used to derive the grammatical status scores and compare our results with those of the authors of the original study.

In our analysis, we further assess the appropriateness of the selected metrics, evaluate the accuracy of the binary logistic regression model's prediction, and analyze where units are on the lexico-grammatical scale after grammatical status assignment, and whether this corresponds to their "traditional" placement in word categories.

Key words: grammatical status, corpora, Chinese, quantitative linguistics, replication study