# Rendaku “enthusiasts” and rendaku “indifferents”: Classification of compound nouns based on the frequency of rendaku

Shinri Ohta and Satoshi Ohta (The University of Tokyo/Yamaguchi Univeristy)

Rendaku, or sequential-voicing, is one of the most famous morphophonological phenomena in Japanese, where an initial voiceless obstruent of a non-initial word is voiced in compounds. Regarding frequency of the rendaku in native compound nouns, Rosen (2001) proposed the categories “rendaku lovers” and “rendaku haters”, in which rendaku occurs more than 67% and less than 33% of compounds, respectively. He also claimed that very few nouns appeared to be neither haters nor lovers. On the other hand, Irwin (2012) showed that about 10% of the non-initial elements (NIEs) were actually “rendaku waverers”, in which rendaku occurs between 33 and 66%. Since these studies used arbitrary criteria, i.e., 33% and 67%, it is crucial to examine the plausibility of such criteria for classifying compound nouns based on the rendaku frequency. In this study, we examined an optimal number of clusters, as well as boundary criteria, by using a model-based clustering method.

Based on Rosen (2001) and Irwin (2012), we calculated “rendaku rates” for each NIE, using the Rendaku Database (v.1.1) (Irwin & Miyashita, 2013). Here we took into account compound nouns whose NIEs are native words. To obtain accurate rendaku rate, we excluded compounds with right-branching, coordination, bare numerals, /m/ originated from /b/, and proper names. The NIEs fewer than 10 compounds were also excluded, resulted in 311 different NIEs (average rendaku rate = 77.6%). To determine the number, as well as boundary criteria, of clusters, we applied clustering method, in which data were divided into normal distributions. In these analyses, we considered both model complexity and model accuracy, and determined the optimal model.

The cluster analyses clarified that the model with two clusters were optimal to classify the compound nouns. The average rendaku rates for the cluster 1 (“rendaku enthusiasts”) and cluster 2 (“rendaku indifferents”) were 97.0 ± 3 % (162 NIEs) and 58.5 ± 32 % (149 NIEs), respectively. The boundary value for these clusters was 0.89, which was higher than those proposed by Rosen (2001) and Irwin (2012). Our bounding value was consistent with the fact that the rendaku appears in most of compound nouns. In contrast to Irwin (2012), our results demonstrated that the rendaku haters and rendaku waverers in fact composed a single cluster of “rendaku indifferents”, leading to a more general model for these compounds with the low rendaku rate.

References

Rosen, Eric Robert (2001), PhD dissertation, University of British Columbia.
Irwin, Mark (2012), Rendaku Haters and the Logistic Curve, 22nd Japanese/Korean Linguistics Conference.
Irwin, Mark & Miyashita, Mizuki (2013), The Rendaku Database v1.1.