

How Informative is One Stroke?

A Contrastive Study of Korean Hangul and Japanese Katakana

Jaewoo Song (jsong.st@gmail.com)

Abstract

Although graphemes are by definition the smallest units of writing systems, they consist of even smaller building blocks: strokes, curves, dots and circles. However, there has not been much empirical study on how much information those smaller units (called ‘mini-graphemes’ in the context) deliver. In this work, the relative amount of information one stroke conveys in Korean Hangul and Japanese Katakana is analyzed and compared extensively with 1676 most frequently used vocabulary pairs. Korean Hangul and Japanese Katakana have good characteristics to be the research subjects. First, they are almost entirely made with clearly distinguishable strokes which make it easy to count their numbers. Circles in Hangul and Katakana, the only mini-graphemes which are not a stroke, can be considered as one stroke without loss of generality. Second, it is easy to find one-to-one mappings of Korean and Japanese words. Because words of similar meanings but with different nuances can make the analysis result of this work less precise, it is important to select two target languages like Korean and Japanese which can be translated into each other with much of their subtle meanings kept intact. The analysis has shown that one stroke in Hangul is 21.20% less informative than its counterpart in Katakana: on average 0.7880 (which is $1 - 21.20\%$) Katakana strokes are needed to express the meaning conveyed in 1 Hangul stroke. In detail, the average ratio of the number of strokes of Katakana to Hangul for each word pair is 0.7880 and the standard deviation is 0.3538. This research is by far the first research which addressed the amount of information conveyed by strokes of writing systems. It will be easy to broaden the research to many other writing systems such as both traditional and simplified Chinese characters, Japanese Hiragana and Latin and Greek alphabets.