# Quantitative evidence for a grammatical analysis of Chinese character component size

**James Myers**
*National Chung Cheng University*
Lngmyers@ccu.edu.tw

In the grammatical terms of the Chinese character analysis of Myers (2019), the most productive "morphological" operation affixes a semantic component to one related to pronunciation, as in 接 *jiē* 'connect' (扌 'hand' + 妾 *qiè*). Far less productive is semantic compounding, as in 掃 *sǎo* 'sweep' (= 扌 'hand' + 帚 'broom'). Character "prosody", motivated by left-to-right writing order, adds a left-small/right-large asymmetry in stroke count ("segmental phonology") and physical size ("phonetics"), as in the left-side reduction of 手 *shǒu* 'hand' to 扌.

This study presents a quantitative test of this grammatical analysis in a database of 5,084 traditional Chinese characters with left-right structure, using component sizes from the Wenlin Character Description Language database (Bishop & Cook 2007). Mixed-effects linear regression models predicted component stroke count or physical size from character type (affixed/compounded), component type (affix/non-affix), component position (left/right), and log token frequency.

There were three main findings. (1) In both character types, the left component had significantly fewer strokes and was significantly narrower than the right one. (2) In affixed characters, stroke count and size were significantly influenced both by position and by component type, which did not interact. (3) There was a stronger effect of position on stroke count in rarer compounds and in more common affixed characters, while component size was mostly unaffected by frequency.

All three findings are consistent with the grammatical analysis. (1) Position (prosody) has independent effects on stroke count (segmental phonology) and component size (phonetics). (2) The affix/non-affix contrast (morphology) has effects independent of position (prosody). (3) The distinct mechanisms underlying morphology, prosody, and phonetics lead them to be associated with distinct frequency effects (Bybee 2006). Rarer semantic compounds are regularized, that is, reanalyzed as if derived via the more productive operation of affixation, with the left component reduced as if it were an affix. Meanwhile, prosodic regularities affect more common characters through the automatization of articulation (i.e., repeatedly writing the same components left to right). Finally, component size is not much influenced by lexical frequency because it is non-contrastive phonetics.

**References:** • Bishop, T. & R. Cook (2007). A character description language for CJK. *Multilingual*, 62–68. • Bybee, J. (2006). *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press. • Myers, J. (2019). *The Grammar of Chinese Characters*. Abingdon: Routledge.