Prosody and word length in Chinese

San Duanmu, University of Michigan

The eminent Swedish sinologist Bernhard Karlgren (1918) made an insightful observation of an unusual property of Chinese: Many Chinese words have two synonymous forms, one short (monosyllabic) and one long (disyllabic), such as *jian* 'see' and *kan-jian* 'look-see'. The long form is made by adding another morpheme to the short one, although the extra morpheme contributes little new meaning. Such short-long pairs have also been called 'elastic words'. Karlgren further offers an explanation for the abundance of elastic words in Chinese: Because Chinese words are mostly monosyllabic, yet Chinese has a small syllable inventory, there are too many homophones. To avoid ambiguity, disyllabic words (or what Karlgren calls 'elucidative compounds') are created as alternatives to monosyllabic ones, giving rise to elastic words.

Karlgren's explanation is accepted widely, from the West to the East and from the beginning to the present day. In addition, his explanation is so simple and so intuitive that little evidence is thought to be needed, and little evidence has ever been offered. However, his theory of homophone avoidance makes explicit predictions, which can be tested. For example, some Chinese dialects have twice as many syllables as others. If Karlgren is correct, we would expect the former to have fewer elastic words than the latter. Similarly, within the same dialect, (i) some syllables have no homophones, (ii) some have a few, and (iii) some have many. If Karlgren is correct, we would expect syllables of case (i) to have a low percentage of elastic words, syllables of case (ii) to have a higher percentage, and syllables of case (iii) to have the highest.

In this talk, I review previous attempts to confirm Karlgren's theory (e.g. Ke 2006 and Jin 2011). Then I report a quantitative study of elastic words in Chinese, though a complete manual annotation of *Xiandai Hanyu Cidian* (Modern Chinese Dictionary 2005). In particular, I report percentages of elastic words in each POS (part of speech) category, and percentages of elastic words for syllables of each degree of homophony. The results show no evidence for the theory of homophone avoidance, despite its apparent simplicity and intuitive naturalness, and despite its nearly universal acceptance.

I also offer an alternative theory, according to which elastic words are motivated by prosody. In prosodically strong positions, foot binarity requires a disyllabic word, whereas in other positions monosyllabic words remain viable. The prosodic approach offers a better explanation of the distribution of elastic words in the lexicon. In addition, it accounts for word length preferences in language use, such as [N N] compounds and [V N] phrases. I conclude that prosody plays a far greater role in language, than it appears at first sight, leaving a footprint on a wide range of apparently unrelated places.