OCP Avoidance in Classical Chinese: Implications for Tonogenesis

Recent reconstructions of Old Chinese lack tones, instead reconstructing so-called 'post-coda' segments at the end of syllables, which caused tonal contours and were subsequently lost as a process of tonogenesis between Old Chinese and Middle Chinese (Baxter 1992; Baxter and Sagart 2014). Although Shen (Under Review) notes that the original identities of these post-codas can be reconstructed through comparisons with genetic relatives and transcriptions of Sanskrit, the exact dating of tonogenesis and the loss of post-codas is difficult to determine due to the ideographic nature of the Chinese script. Through a corpus study of five pre-Qin texts, this paper argues that post-codas were present in Chinese as late as the Warring States Period (475 – 221 BCE). This paper shows that the poetry of the Book of Odes (詩經 *shījīng*) and Songs of the South (楚辭 *chǔcí*) has significantly fewer ($p < 0.005$) OCP violations on word boundaries than predicted by randomization of characters in each text when accounting for the existence of post-codas. This is much in line with Shih and Zuraw (2017)'s analysis of phonologically conditioned word order variation, which would cause such a discrepancy. Five works in Classical Chinese make up the corpus of this study: The Book of Odes (詩經 *shījīng*), Songs of the South (楚辭 *chǔcí*), Zuo Zhuan (左傳 *zuǒzhuàn*), Shan Hai Jing (山海經 *shānhǎi jīng*), and the Art of War (孫子兵法 *sūnzi bīngfǎ*), all of which were provided by Sturgeon (2019)'s Chinese Text Project.

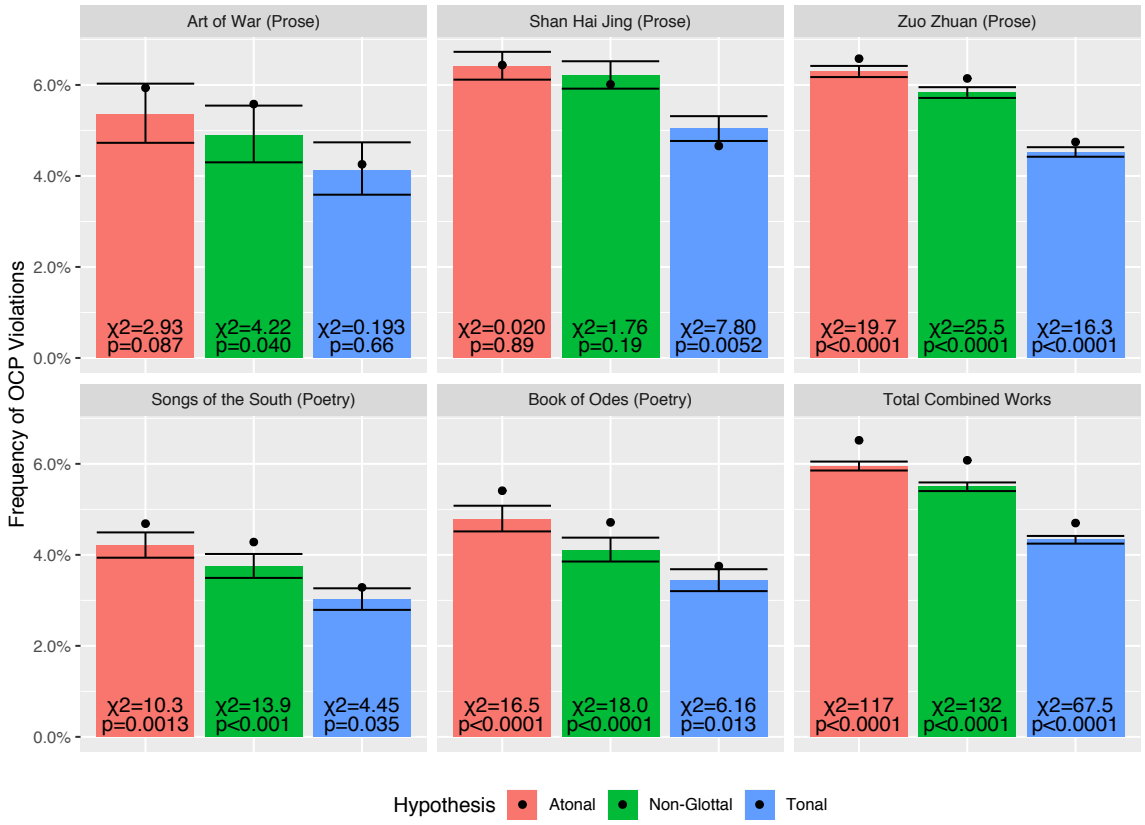
Gunkel and Ryan (2015) shows that the avoidance of haitus motivates marked word order in the Rigveda, and argue that both choice of vocabulary and change in syntax may be utilized to avoid OCP violations, particularly in poetic form, where diction and syntax are less restricted and sensitivity to phonological constraints is heightened. This study finds that Classical Chinese poetry utilizes these strategies as well; OCP violations along syllable boundaries are significantly less frequent than predicted by randomization of characters ($p < 0.01$ for Songs of the South, $p < 0.001$ for Book of Odes). Initial tests of OCP violations are done only for bigrams whose first syllable is *píng* or *rù* tone in Middle Chinese, as these tones are reconstructed in Old Chinese without post-coda segments, and thus provide a theory-neutral groundwork, where OCP violations among these syllable boundaries will be the same regardless of the presence of post-codas.

This paper considers three hypotheses regarding the state of tonogenesis for each of the analyzed works: 1) the Atonal Hypothesis, where post-codas were present during the time that these works were written and thus influence OCP restrictions, 2) the Tonal Hypothesis, where post-codas were not present, and 3) the Non-Glottal Hypothesis, where post-codas were present but where glottal stop post-codas do not effect OCP restrictions.

OCP violations are calculated under each hypothesis by grouping characters into initial and final OCP categories based on their reconstruction and the hypothesis. Bigrams where the OCP category of the first characters final matches that of the second character's initial are considered violations. For example, the character 瞰 *kàn* *k^hɑm(s) has a -m- coda and *-s post-coda, thus ending in a -m# for the tonal hypothesis, and ending in a *-s# for the other two hypotheses. In Bēi Huí Fēng (悲回風), a poem of Songs of the South, 瞰 *k^hɑm(s) is followed by 霧 (*wù*) *mɨo(s), and thus forms a OCP violation in the tonal hypothesis (m#m), but not in the atonal and non-glottal hypotheses (s#m).

In the following figure, each point represents the predicted rate of OCP violations. While producing more OCP violations than the tonal hypothesis, OCP effects driven by post-codas

are significant, particularly for the works of poetry (first two panels of the second row), where the OCP effects are only borderline under the tonal hypothesis, but are are strongly significant with the atonal and non-glottal hypotheses, which consider post-codas.



These findings suggest that OCP violations were avoided in Classical Chinese, and supports the existence of post-codas as late as the Warring States Period (475 – 221 BCE), when Songs of the South was written. The findings of this study reinforces Shih (2017)'s theory that phonology has synchronic effects on the syntax of a language, and suggests both that the OCP carries its effect across word boundaries and that these effects extend beyond the avoidance of pseudo-geminates, including non-identical phones with shared features.

Bibliography

Baxter, William Hubbard. 1992. *A handbook of old Chinese phonology*. Berlin ; New York: Mouton de Gruyter.

Baxter, William Hubbard, and Laurent Sagart. 2014. *Old Chinese: a new reconstruction*. Oxford ; New York: Oxford University Press.

Gunkel, Dieter, and Kevin M Ryan. 2015. Investigating Rigvedic Word Order in Metrically Neutral Contexts. Vienna: University of Vienna.

Shen, Zhongwei. Under Review. *A Phonological History of Chinese*. Cambridge ; New York, NY: Cambridge University Press.

Shih, Stephanie S. 2017. Phonological Influences in Syntactic Alternations. In *The Morphosyntax-Phonology Connection*, ed. Vera Gribanova and Stephanie S. Shih, 223–252. Oxford University Press.

Shih, Stephanie S., and Kie Zuraw. 2017. Phonological conditions on variable adjective and noun word order in Tagalog. *Language* 93:317–352.

Sturgeon, Donald. 2019. Chinese Text Project: A dynamic digital library of premodern Chinese. *Digital Scholarship in the Humanities* .