OCP Avoidance in Classical Chinese: Implications for Tonogenesis

Jack Rabinovitch
Harvard University

1. Introduction

Tonogenesis occured some time between the stages of Old and Middle Chinese.

(1)

Old Chinese	Middle Chinese
$\sim$ 1200–200 BCE	$\sim$ 300–1300 CE
Post-codas	Tones

Chinese is logographic: It is hard to pinpoint when tonogenesis happened.

Obligatory Contour Principle: Avoid adjacent segments which are too similar.

Gunkel and Ryan (2015): High ranking OCP constraints causes marked word order.

(2) a. The presence or absence of post-codas affects cross-word OCP violations, thus:
b. If OCP avoidance happens across word boundaries, then:
c. Texts reflect presence/absence of post-codas by a lack of bigrams with OCP violations.

1. Introduction

Corpus: Five classical texts in Chinese, taken from Sturgeon (2019). Approximate dates come from Tz'ù and Hawkes (1959), Lust (1996), and Durrant et al. (2016).

(3)

Abbr.	English	Chinese	Pinyin	Date	Genre
SJ	The Book of Odes	詩經	*shījīng*	1600–771 BCE	Poetry
AoW	Art of War	孫子兵法	*sūnzi bīngfǎ*	500–400 BCE	Prose
ZZ	The Commentary of Zuo	左傳	*zuǒzhuàn*	∼300 BCE	Prose
SHJ	Classic of Mountains and Seas	山海經	*shānhǎi jīng*	475–221 BCE	Prose
CC	Songs of the South	楚辭	*chǔcí*	475 BCE–220 CE	Poetry

A given Chinese text may have OCP constraints reflecting one of the three:

(4)
a. Atonal Hypothesis: Post-codas are present and can cause OCP violations
b. Non-Glottal Hypothesis: Atonal but glottal stops are transparent to OCP effects
c. Tonal Hypothesis: Post-codas are not present, OCP effects are based entirely on coda

This study finds that OCP effects in Classical Chinese are significant in two ways:

(5)
a. For bigrams lacking an intervening post-coda segment, cross-word OCP effects are significant for poetry ($p<0.01$) and for the Commentary of Zuo ($p<0.0001$).
b. When considering bigrams with post-coda segments, cross-word OCP effects in poetry are significant for the Atonal and Non-Glottal hypotheses ($p<0.001$), and borderline for the Tonal hypothesis ($0.01<p<0.05$).

Baxter and Sagart 2014 (B&S) for Old Chinese and Zhèngzhāng 2003 (ZZSF) for Middle Chinese.

2. Meet the Post-Codas

(6) Exaggeration and Transfer Theory (Kingston 2011)

a. **Exaggeration**: Some segments causes tonal contour to appear in a language.

b. **Transfer**: Said segments undergo deletion or are otherwise merged such that the tonal contour becomes phonemic.

Post-codas: segments (/s/ and /ʔ/) in Old Chinese, can co-occur with nasal or glide codas.

(7)

Old Chinese Segment	Exaggeration		Transfer		Middle Chinese Tone
	segm	tone	segm	tone	
m]$_\sigma$	m	level	m	level	píng (平) Ø
n]$_\sigma$	n	level	n	level	
ŋ]$_\sigma$	ŋ	level	ŋ	level	
V]$_\sigma$	V	level	V	level	
j]$_\sigma$	j	level	j	level	
w]$_\sigma$	w	level	w	level	
ʔ]$_\sigma$	ʔ	rising		rising	shǎng (上) X
s]$_\sigma$	s	falling		falling	qù (去) H
p]$_\sigma$	p	short	p	short	rù (入) Ø
t]$_\sigma$	t	short	t	short	
k]$_\sigma$	k	short	k	short	

For example, take *rén* 壬 'ninth' and *rèn* 妊 'pregnant', only distinguishable by tone.

(8)

Char.	Old Chinese (B&S)	Exaggerate		Transfer	Middle Chinese (ZZSF)	Mandarin (Pinyin)
壬	/*n[ə]m/	/*n[ə]m^{Ø}/	→	/*n[ə]m^{Ø}/	ȵiɪm^{Ø}	*rén*
妊	/*n[ə]ms/	/*n[ə]ms^{H}/	→	/*n[ə]m^{H}/	ȵiɪm^{H}	*rèn*

3. The OCP and Capturing Sound Change

Optimality Theory: OCP appears as **violable constraints**, e.g. *PseudoGem assigns violations to identical adjacent segments.

(9)

Pinyin	Char.	*PseudoGem	Old Chinese		*PseudoGem	Middle Chinese
fàngxīn	放心	*	paŋs səm	→		pʉɐŋ siim
fàngyuè	放月		paŋs ŋʷat	→	*	pʉɐŋ ŋʉɐt˥

At different stages of Chinese, different sequences would violate OCP constraints.

Avoid OCP by non-canonical word order / word choice:

(10)

暗星 /qˤums stsʰˤeŋ/	*PseudoGem	*NounAdj	Faith[Word]
a. 暗星 [qˤums stsʰˤeŋ]	*!		
b. ☞ 星暗 [stsʰˤeŋ qˤums]		*	
c. ☞ 黑星 [m̥ˤək stsʰˤeŋ]			*

Old Chinese readings: preference to use unfaithful word order/word choice.

(11)

暗星 /ʔʌmᴴ seŋ/	*PseudoGem	*NounAdj	Faith[Word]
a. ☞ 暗星 [ʔʌmᴴ seŋ]			
b. 星暗 [seŋ ʔʌmᴴ]		*!	
c. 黑星 [hək seŋ]			*!

Middle Chinese readings: no OCP violation, preference is given to the faithful candidate.

4. First Steps

Zhèngzhāng (2003): Guǎngyùn 廣韻 maintains Sui Dynasty (581–618 CE) pronunciation.

Glide codas: difficult to distinguish from vowels; considered non-codas for this study.

(12) a.

Char	Readings	Initials	Codas
參	/ʃʏiɪm/, /tsʰʌm/, /tʃʰʏiɪm/	ʃ, tsʰ, tʃʰ	m∅
差	/tʃʰʏɛ/, /tʃʰʏɛᴴ/	tʃʰ	∅∅, ∅s
荇	NA	NA	NA
菜	/tsʰʌiᴴ/	tsʰ	∅s

b.

Bigram	First Char. Codas	Second Char. Initials
參差	m∅	tʃʰ
差荇	∅∅, ∅s	NA
荇菜	NA	tsʰ

Expected frequency of a given bigram type (coda # initial combination) in a given text = frequency of a given coda type multiplied by the frequency of a given initial type.

(13)

Initial	Coda	Initial Count	Coda Count	Product Count	Frequency
b	m∅	105	49	5145	0.0456
b	ms	105	7	735	0.00651
m	m∅	191	49	9359	0.0829
m	ms	191	7	1337	0.0118

5. Finding OCP Violations

Initials and codas are grouped: OCP violations occur between segments within the group (place and manner of articulation).

(14)

Category	Initials	IPA reconstructions
P	幫滂並	p, p^h, b
T	端透定知徹澄 精清從章昌常 莊初崇	t, t^h, d, ʈ, $ʈ^h$, ɖ, ts, ts^h, dz, tɕ, $tɕ^h$, dʑ, tʃ, $tʃ^h$, dʒ
K	見溪群	k, k^h, g
M	明	m
N	泥日娘	n, ɳ, ɳ
Ŋ	疑	ŋ
S	心書生邪船俟	s, ɕ, ʃ, z, ʑ, ʒ
ʔ	影	ʔ
other	曉云匣以來	h, ɦ, ɦ, j, l

Which group a coda occurs under depends on hypothesis.

(15)

Coda	Atonal	Non-Glottal	Tonal	Tone
∅∅	∅	∅	∅	píng (平)
m∅	M	M	M	
n∅	N	N	N	
ŋ∅	Ŋ	Ŋ	Ŋ	
∅ʔ	ʔ	∅	∅	shǎng (上)
mʔ	ʔ	M	M	
nʔ	ʔ	N	N	
ŋʔ	ʔ	Ŋ	Ŋ	
∅s	S	S	∅	qù (去)
ms	S	S	M	
ns	S	S	N	
ŋs	S	S	Ŋ	
p∅	P	P	P	rù (入)
t∅	T	T	T	
k∅	K	K	K	

6. Results

Characters with only *píng* and *rù* tone are in the same group across hypotheses. In the "predicted" database and in the corpus:

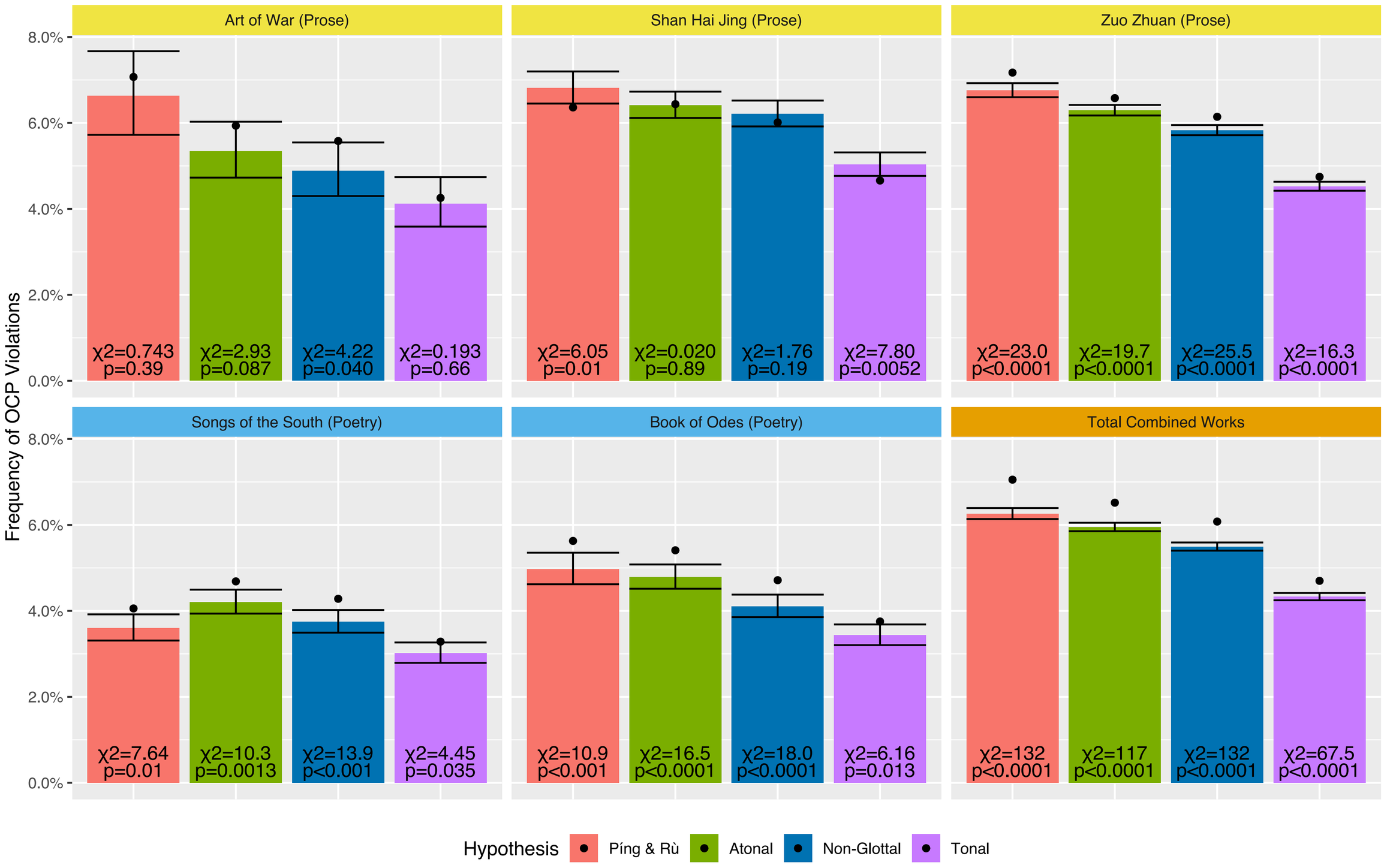
(16) a. Bigrams whose first-character have a *shǎng* or *qù* tone reading are ignored.
b. Bigrams with first character mixed-coda, second character mixed-initials, or with NA segments are ignored.

χ^2 test performed on each book against the predicted rate of violations for that book.

(17) In Hypothesis Neutral Contexts:
a. Cross-word OCP effects significant: ZZ, CC, SJ
b. Poetry more significant than prose
c. Borderline negative results for SHJ

(18) In All Contexts:
a. The same patterns show: ZZ, CC, SJ significant.
b. AoW dates earlier than ZZ; is less significant.
c. Poetry: Atonal, Non-glottal are significant; Tonal borderline.

χ^2 test including bigrams with first-character have a *shǎng* or *qù* tone reading added, across all hypotheses:



Bibliography

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

Durrant, Stephen W., Wai-yee Li, and David Schaberg, ed. 2016. *Zuo Tradition/Zuozhuan: Commentary on the "Spring and Autumn Annals."*. Classics of Chinese Thought. Seattle: University of Washington Press.

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

Kingston, John. 2011. Tonogenesis. *The Blackwell companion to phonology* 1–30.

Lust, John. 1996. *Chinese popular prints*, volume 4 of *Handbuch Der Orientalistik. Vierte Abteilung, China, Handbook of Oriental Studies, China*. Leiden ; New York: Brill.

Shih, Stephanie S., and Kie Zuraw. 2017. Phonological conditions on variable adjective and noun word order in Tagalog. *Language* 93:e317–e352.

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

Tz'ù, Ch'u, and David Hawkes. 1959. *The Songs of the South: An Ancient Chinese Anthology*. Columbia UP.

Zhèngzhāng, Shàngfāng. 2003. *Old Chinese Phonology*. Chinese Modern Linguistics Series. Shanghai.