Towards a Time Course of Chinese Tonogenesis: Evidence from Phonologically Conditioned Syntax

Jack Rabinovitch

Harvard

April 30, 2021

Background

Periodization of Chinese: Old Chinese (circa 1250 – 221 BCE) and Middle Chinese (circa 386 – 1279 CE).

Old Chinese: Not a tonal language; 'post-coda' segments at ends of syllables.

Middle Chinese: Tonal language; 'post-coda' segments elided and formed tonal contours.

Chinese has a **logographic** script, so sound change (tonogenesis) is not represented by orthography.

If Classical Chinese, like many languages, has **phonologically conditioned syntax**, then this should reflect stages of tonogenesis in different works written at different times.

Overview

In this talk I will go over:

(1)
a. An introduction to Old Chinese and Middle Chinese phonotactics and sound change
b. An overview on what we know about **phonologically conditioned syntax** in other languages
c. How we can use **phonologically conditioned syntax** to find what stage of tonogenesis a text is likely in
d. The corpus and methodology of data collection
e. Some interesting results

Old Chinese reconstructions are based on Baxter and Sagart (2014) (abbreviated B&S), Middle Chinese Zhèngzhāng (2003) (abbreviated ZZSF). As shorthand, I use 'falling' to refer to what is classically known as 'departing' tone.

Exaggeration and Transfer

(2) Kingston (2011)'s 'Exaggeration and Transfer'

a. Stage I: A language has some set of segments X with no influence on tone.

b. Stage II: The set of segments X cause a predictable tonal contour on their syllables.

c. Stage III: The segments X are no longer contrastive; their resulting tonal contours are contrastive.

We can define tonogenesis as a process with these three stages. Stage II exists currently in Yabem (Ross 1993) and Korean (Cho 2017).

(3)

Stage I	Exag.	Stage II	Trans.	Stage III
ABC	→	ABC + Contour	→	AB + Tone 1
AB		AB + No Contour		AB + Tone 2

The Old Chinese Coda

The Old Chinese coda consists of two parts, the coda proper, or 'pre-coda', and the 'post-coda'.

(4)
a. Optional 'pre-coda': /w, j, m, n, ŋ, p, t, k/
b. Optional 'post-coda': /s, ʔ/
c. Banned combinations: /pʔ, tʔ, kʔ/

Stop 'pre-codas' with an /-s/ 'post-coda' underwent elision early on: /ps#, ts#, ks#/ > /s#/

(5) Ignoring glide codas:

/m/	/mʔ/	/ms/	/p/	/s/
/n/	/nʔ/	/ns/	/t/	/ʔ/
/ŋ/	/ŋʔ/	/ŋs/	/k/	∅

Phonological reconstructions come from related Sino-Tibetan languages, which have these correspondences to Chinese tones.

The Middle Chinese Coda and Tone

Middle Chinese is represented by the *Qièyùn* (601 CE) rime dictionary. 'Post-coda' segments are gone, but now there are four tones.

(6)

OC Segment (Stage I)	Stage II		Stage III	
	Segment	Tone	Segment	Tone
ʔ]$_\sigma$	ʔ	rising		rising
s]$_\sigma$	s	falling		falling
p]$_\sigma$	p	checked	p	checked
t]$_\sigma$	t	checked	t	checked
k]$_\sigma$	k	checked	k	checked
Otherwise Z]$_\sigma$	Z	level	Z	level

The Middle Chinese Coda and Tone (Cont.)

Stop endings /p,t,k/ form checked tone syllables; other codas form level, rising, or falling tone syllables depending on OC 'post-coda'.

(7)

Char.	Stage I (B&S)	Stage II	Stage III (ZZSF)	Mandarin (Pinyin)
壬	*n[ə]m	*n[ə]m$^{\text{Level}}$	ȵiɪm$^{\text{Level}}$	*rén*
妊	*n[ə]mʔ	*n[ə]mʔ$^{\text{Rising}}$	ȵiɪm$^{\text{Rising}}$	*rěn*
荏	*n[ə]ms	*n[ə]ms$^{\text{Falling}}$	ȵiɪm$^{\text{Falling}}$	*rèn*

Comparison of *rén* 壬 'ninth celestial stem', *rěn* 妊 'pregnant', and *rèn* 荏 'perilla frutescens'.

Phonologically Conditioned Syntax and Word Choice

Recent work in synchronic phonology: variation in word order and word choice is influenced by the phonology of the language.

Breiss and Hayes (2020): bigrams which cause marked phonology at word boundaries are avoided:
*Iambic Clash, *NC, OCP[Sibilant]

(8) a. Genitive construction in English (Shih et al. 2015)
b. Noun-adjective order in Tagalog (Shih and Zuraw 2017)
c. Noun-modifier order in Rigvedic Sanskrit (Gunkel and Ryan 2011, 2015).

This might be the same for Classical Chinese. (In fact it is.)

Pre-tonogenesis and Post-tonogenesis bigrams will have different markedness, and thus should have difference avoidance patterns.

Different Violations at Different Times

Depending on the stage of tonogenesis, a bigram may or may not violate certain phonological constraints.

(9)

		Reconstruction	OCP[Sib]
放心	Old Chinese	*paŋ-s səm	*
fàng xīn	Middle Chinese	pʉɐŋ$^{\text{Fall}}$ siɪm$^{\text{Level}}$	OK

This goes both ways, some bigrams become more or less marked over time.

(10)

		Reconstruction	OCP[Dor]
放眼	Old Chinese	*paŋ-s [ŋ]$^{\text{ʕ}}$<r>ə[n]ʔ	OK
fàng yǎn	Middle Chinese	pʉɐŋ$^{\text{Fall}}$ ŋɣɛn$^{\text{Rise}}$	*

Phonologically Conditioned Syntax and Word Choice

(11) Old Chinese:

暗星 /qˤums stsʰˤeŋ/			OCP[SIB]	*N-A	F[WD]
a.		暗星 [qˤums stsʰˤeŋ]	*!		
b.	☞	星暗 [stsʰˤeŋ qˤums]		*	
c.	☞	黑星 [m̥ˤək stsʰˤeŋ]			*

Changes in diction or word order may be motivated at certain stages of tonogenesis but not others.

(12) Middle Chinese:

暗星 /ʔʌm$^{\text{Fall}}$ seŋ$^{\text{Level}}$/			OCP[SIB]	*N-A	F[WD]
a.	☞	暗星 [ʔʌm$^{\text{Fall}}$ seŋ$^{\text{Level}}$]			
b.		星暗 [seŋ$^{\text{Level}}$ ʔʌm$^{\text{Fall}}$]		!*	
c.		黑星 [hək$^{\text{Check}}$ seŋ$^{\text{Level}}$]			!*

The Hypotheses

For a given text, it may have been written in a version of Chinese in which post-codas were either present or absent.

(13) a. Hypothesis 1: No Tonogenesis Occurred, post-codas affect constraint violations
b. Hypothesis 2: /-ʔ/ post-codas Elided
c. Hypothesis 3: /-s/ post-codas Elided
d. Hypothesis 4: Both /-ʔ/ and /-s/ post-codas Elided

Which hypothesis is true for a given text should be reflected by which bigrams are avoided.

The Corpus

Corpus consists of 27 texts provided by Sturgeon (2019), all created in a range from (a low estimate of) 1046 BCE to (a high estimate of) 540 CE.

Shījīng (1046 BCE–771 BCE), *Zhōuyì* (1046 BCE–771 BCE), *Shàngshū* (772 BCE–476 BCE), *Sūnzi Bīngfǎ* (515 BCE–512 BCE), *Lúnyǔ* (480 BCE–350 BCE), *Chǔcí* (475 BCE–221 BCE), *Dàodéjīng* (475 BCE–221 BCE), *Shānhǎi Jīng* (475 BCE–221 BCE), *Yílǐ* (475 BCE–221 BCE), *Zuǒzhuàn* (468 BCE–300 BCE), *Zhuāngzi* (350 BCE–250 BCE), *Mèngzǐ* (340 BCE–250 BCE), *Zhōulǐ* (300 BCE–100 BCE), *Chūnqiū Fánlù* (206 BCE–9 CE), *Shuōyuàn* (206 BCE–9 CE), *Xīnshū* (206 BCE–9 CE), *Jiāoshì Yìlín* (206 BCE–220 CE), *Jīngshì Yìzhuàn* (57 BCE–37 BCE), *Tàixuánjīng* (33 BCE–18 CE), *Wúyuè Chūnqiū* (50 CE–100 CE), *Lùnhéng* (80 CE–80 CE), *Qiánfūlún* (102 CE–167 CE), *Jīnguì Yàolüè* (170 CE–219 CE), *Rénwùzhì* (190 CE–249 CE), *Sānguózhì* (265 CE–300 CE), *Bàopǔzǐ* (300 CE–343 CE), *Wénxīn Diāolóng* (485 CE–540 CE)

Methodology

For a given bigram type (such as [+LAB][+LAB]) and for each text under each hypothesis:

Actual Rate of a bigram is found for each kind of hypothesis in each text.

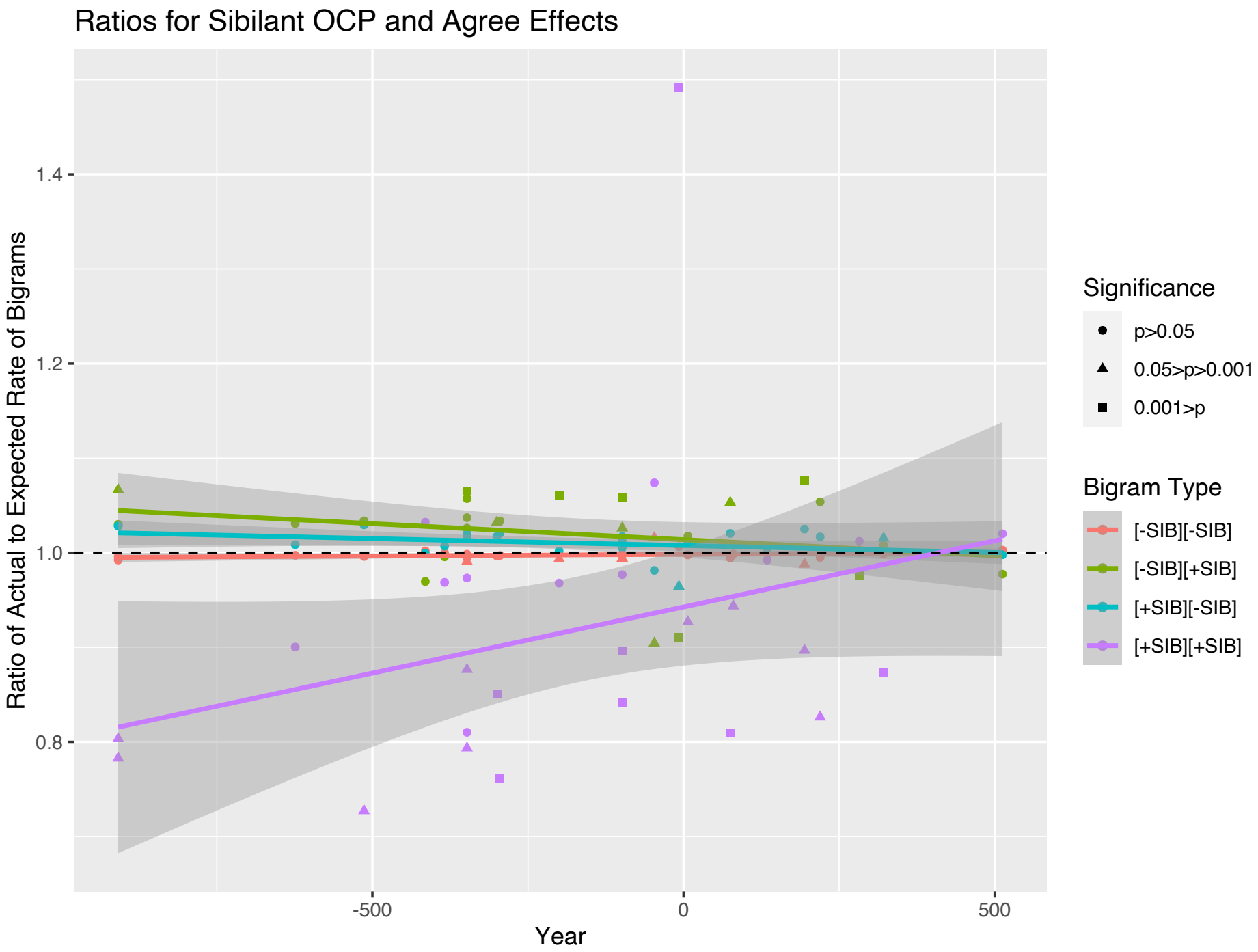
Expected Rate of a given bigram type is found through randomization.

Ratio of Actual to Expected Rate: Above 1 = more likely than chance.
Below 1 = less likely than chance.

Each text is plotted with respect to its average estimated time of creation.

SIBILANTS

Avoidance of $\mathrm{OCP}[\textsc{Sib}]$ violations at word boundaries


Ratios for Sibilant OCP and Agree Effects
Ratio of Actual to Expected Rate of Bigrams
1.4
1.2
1.0
0.8
-500
0
500
Year
Significance
p>0.05
0.05>p>0.001
0.001>p
Bigram Type
[-SIB][-SIB]
[-SIB][+SIB]
[+SIB][-SIB]
[+SIB][+SIB]

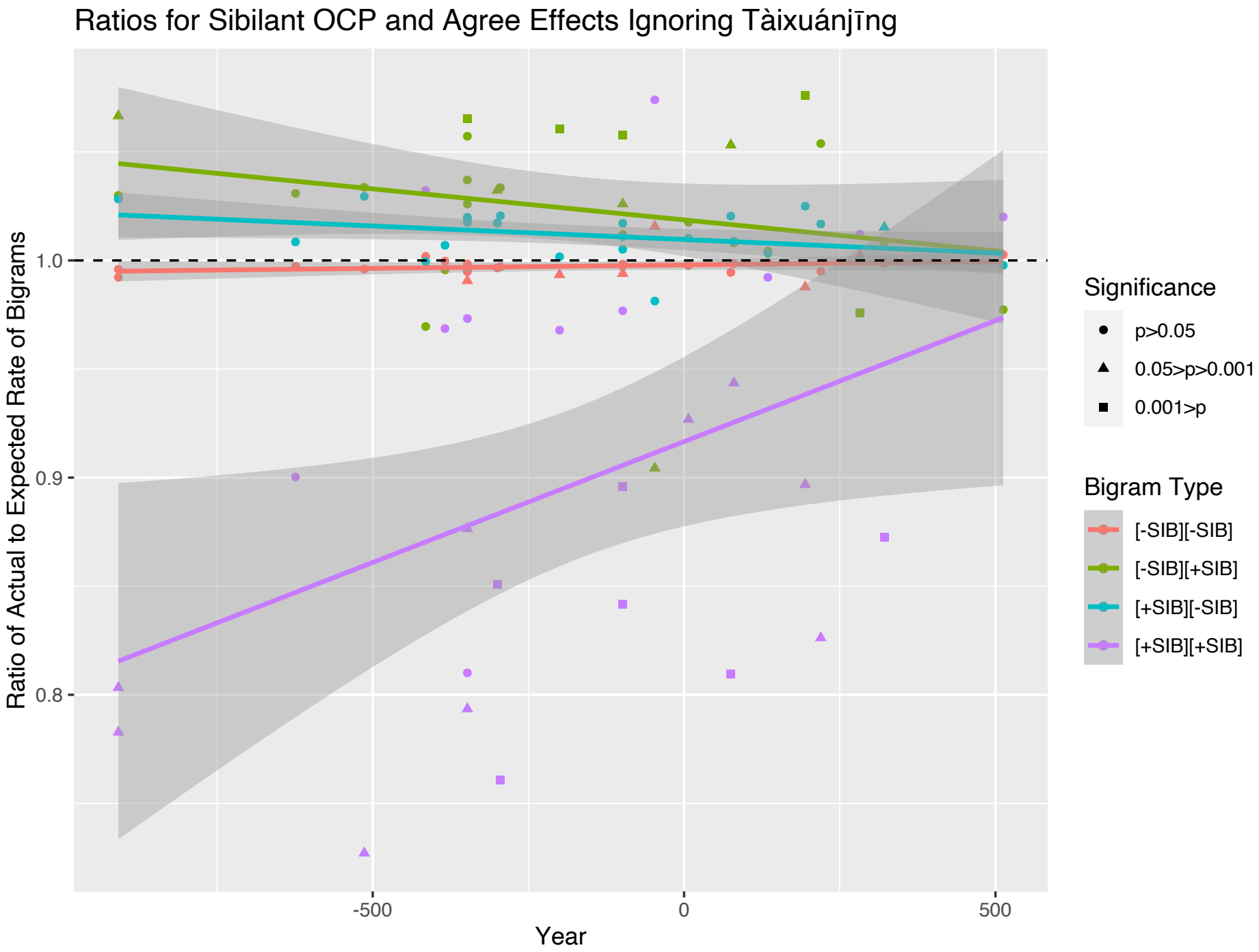
Explaining OCP[SIB]

Assuming no tonogenesis occured, Bigrams which violate OCP[SIB] are avoided more in older texts than later texts.

Explanation 1: The weight of the OCP[SIB] falls, [+SIB][+SIB] bigrams become less marked and less corrected.

Explanation 2: Later texts have /-s/ post-coda elision, [+SIB][+SIB] bigrams become [−SIB][+SIB]; no longer marked or avoided.

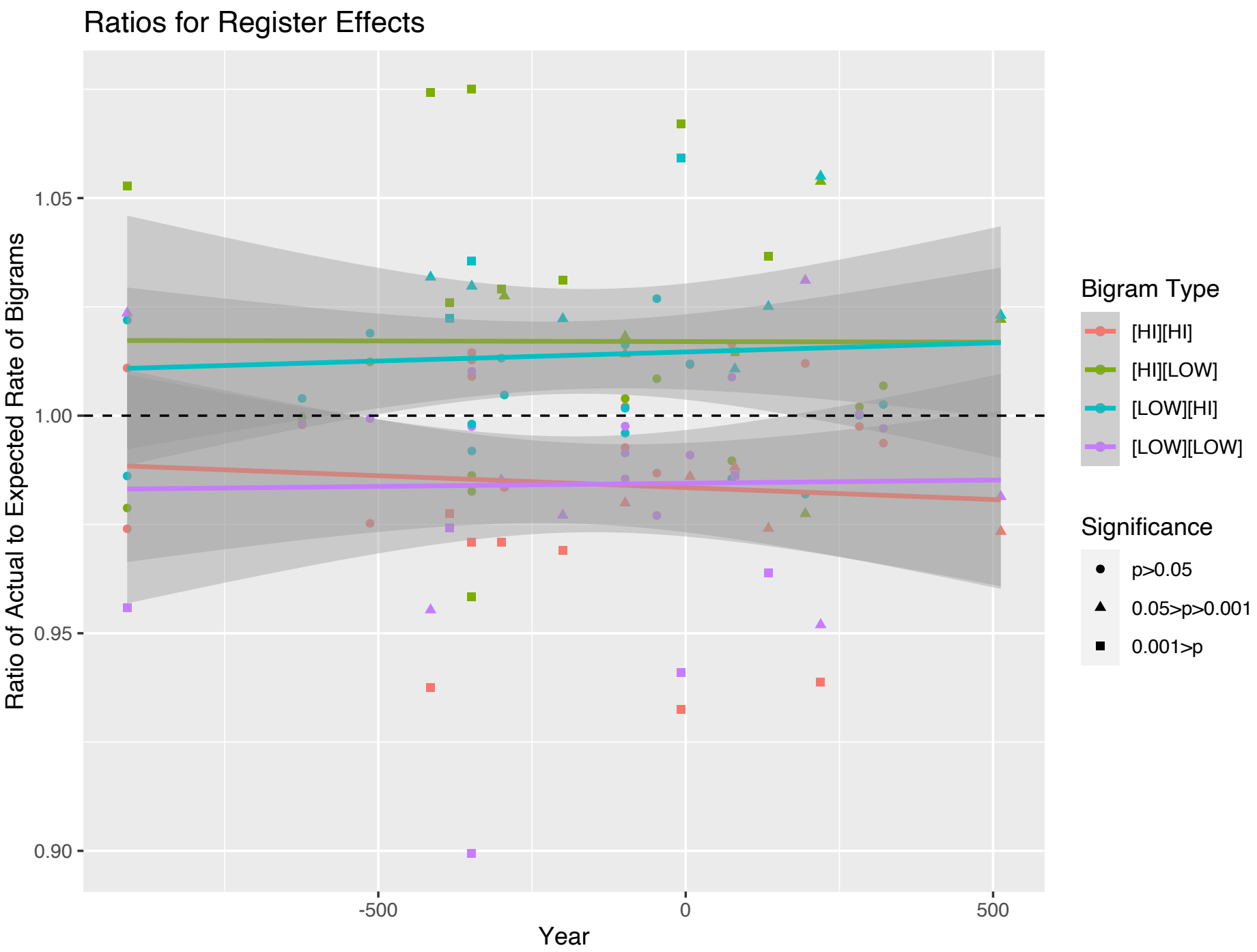
Tàixuánjīng (33 BCE–18 CE) seems like an outlier with a ratio of 1.49. Ignoring it, the ratio of rates still significantly rises over time.


Ratios for Sibilant OCP and Agree Effects Ignoring Tàixuánjīng
Ratio of Actual to Expected Rate of Bigrams
1.0
0.9
0.8
-500
0
500
Year
Significance
p>0.05
0.05>p>0.001
0.001>p
Bigram Type
[-SIB][-SIB]
[-SIB][+SIB]
[+SIB][-SIB]
[+SIB][+SIB]

REGISTER

Middle Chinese is often reconstructed with register, which is predictable based on onset voicing:

$$[+\textsc{Vce}] \text{ Onset} \longleftrightarrow \text{Low Register}$$


Ratios for Register Effects
Ratio of Actual to Expected Rate of Bigrams
1.05
1.00
0.95
0.90
-500
0
500
Year
Bigram Type
[HI][HI]
[HI][LOW]
[LOW][HI]
[LOW][LOW]
Significance
p>0.05
0.05>p>0.001
0.001>p

Explaining Register Effects

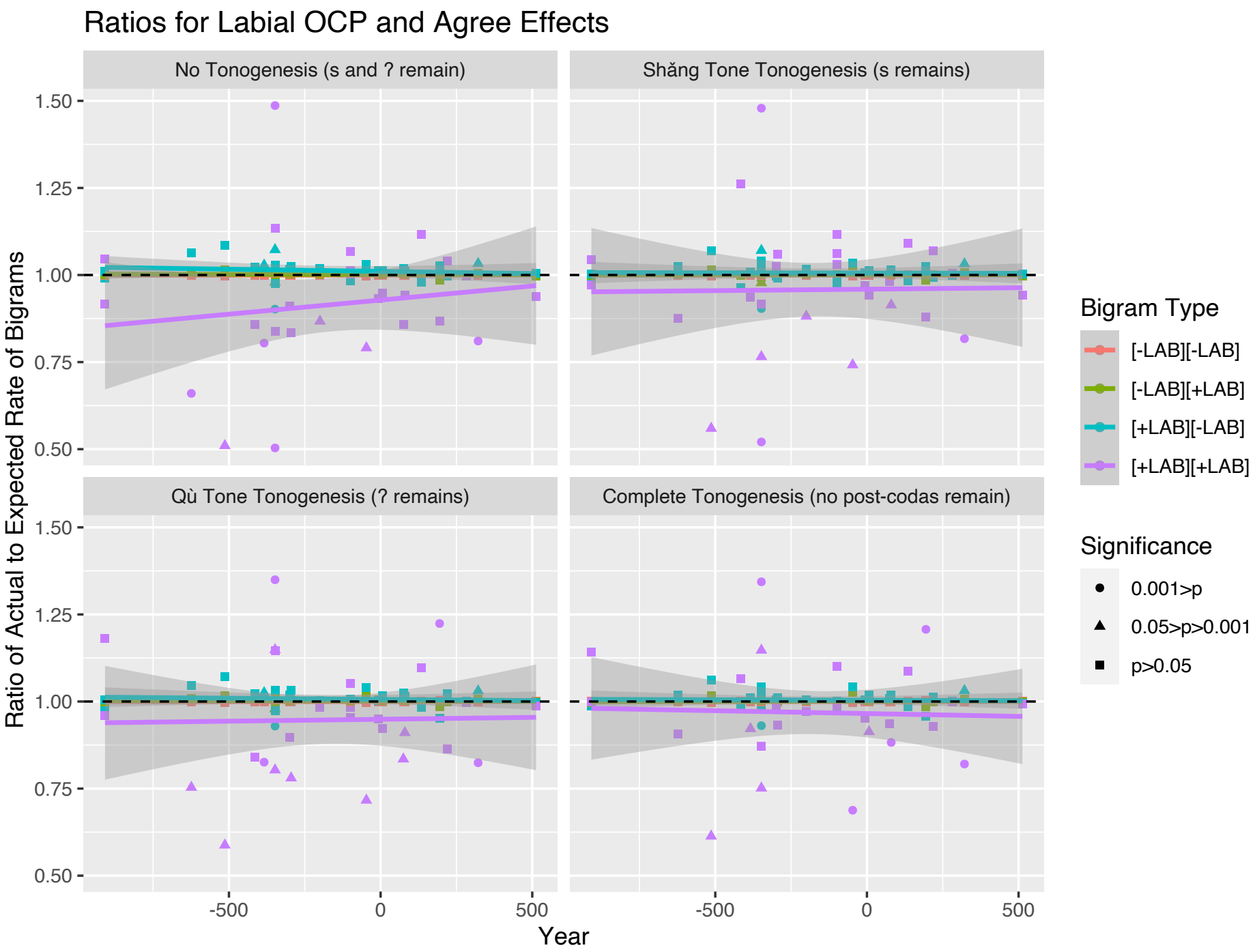
Consistently, both two adjacent high register syllables and two adjacent low register syllables are avoided.

In order for this not to be register based, there would need to be voicing OCP avoidance which targets only onsets.

Suggests that tonal register was present throughout the time period researched here.

LABIAL CLUSTERS

Avoidance of $\mathrm{OCP}[\textsc{Lab}]$ violations at word boundaries


Ratios for Labial OCP and Agree Effects
No Tonogenesis (s and ʔ remain)
Shǎng Tone Tonogenesis (s remains)
Qù Tone Tonogenesis (ʔ remains)
Complete Tonogenesis (no post-codas remain)
Ratio of Actual to Expected Rate of Bigrams
1.50
1.25
1.00
0.75
0.50
-500
0
500
Year
Bigram Type
[-LAB][-LAB]
[-LAB][+LAB]
[+LAB][-LAB]
[+LAB][+LAB]
Significance
0.001>p
0.05>p>0.001
p>0.05

Explaining $\mathrm{OCP}[\textsc{Lab}]$

Regardless of the hypothesis, bigrams which violate $\mathrm{OCP}[\textsc{Lab}]$ are avoided.

Under the 'No Tonogenesis' hypothesis, the slope is significant ($p < 0.05$) in the positive direction.

No significant negative slope for another hypothesis to suggest that said hypothesis becomes a more accurate model for later texts.

Conclusion

Sibilant avoidance is a clear show of /-s/ post-coda elision by 500 CE.

Register seems to be active as far back as 500 BCE.

Change seems to not be categorical; heterogeneity of texts.

The avoidance of specific constraints may be tied to the salience of the concerned features; as salience of post-codas falls, so does avoidance.

Need for wider time range; larger corpus; more targeted tests?

Bibliography I

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

Breiss, Canaan, and Bruce Hayes. 2020. Phonological markedness effects in sentence formation. *Language* 96:338–370.

Cho, Sunghye. 2017. Development Of Pitch Contrast And Seoul Korean Intonation. PhD dissertation, University of Pennsylvania, Philadelphia.

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. 2011. Hiatus avoidance and metrification in the Rigveda. *Proceedings of the 22nd Annual UCLA Indo-European Conference* 53–68.

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

Bibliography II

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.

Ross, Malcolm D. 1993. Tonogenesis in the North Huon Gulf Chain. *Oceanic Linguistics Special Publications* 24:133–153.

Shen, Zhongwei. 2020. *A Phonological History of Chinese*. Cambridge ; New York, NY: Cambridge University Press.

Shih, Stephanie, Jason Grafmiller, Richard Futrell, and Joan Bresnan. 2015. Rhythm's role in genitive construction choice in spoken English. In *Rhythm in Cognition and Grammar*, ed. Ralf Vogel and Ruben van de Vijver, 207–234. De Gruyter Mouton.

Bibliography III

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.

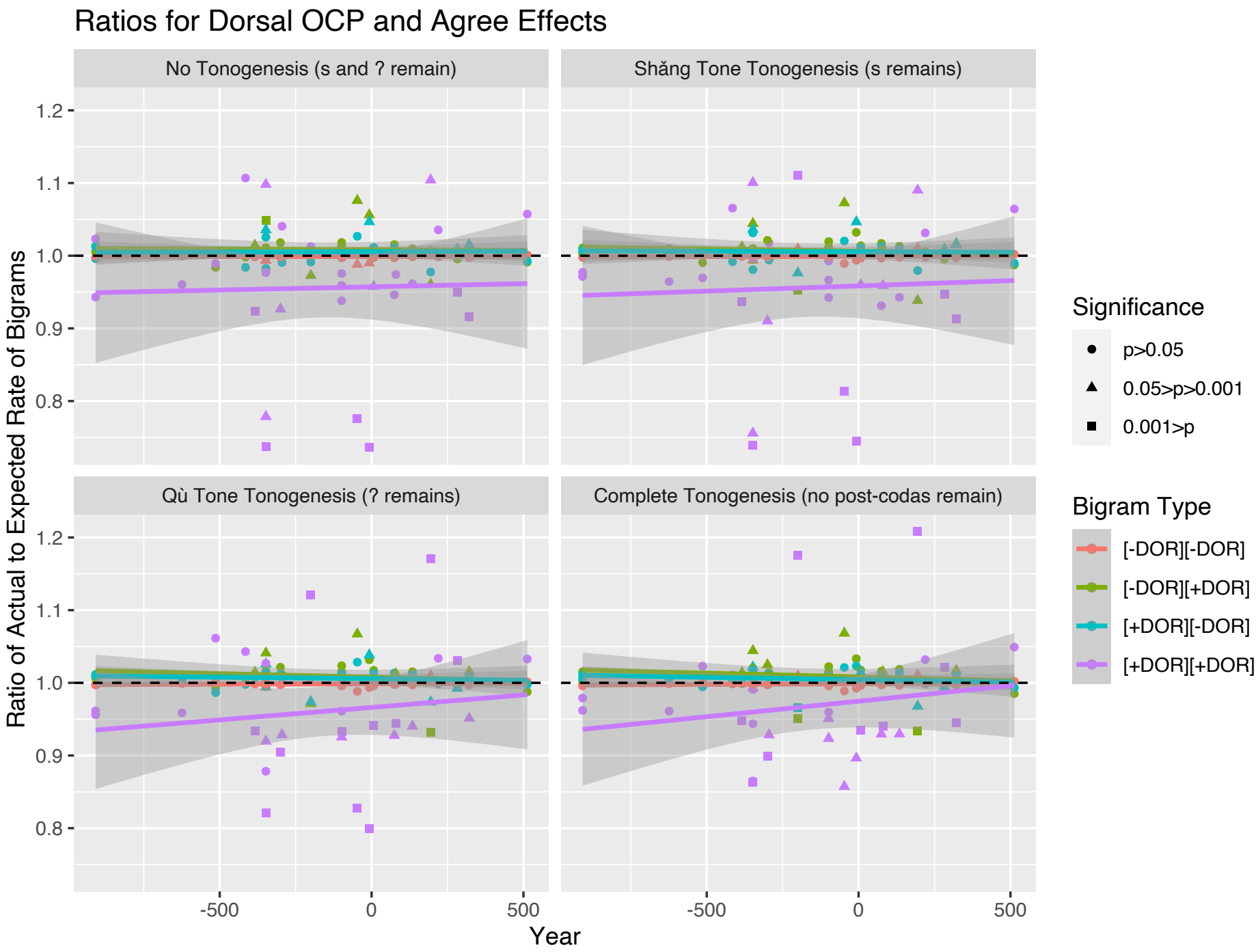
Shih, Stephanie S., and Kie Zuraw. 2017. Phonological conditions on variable adjective and noun word order in Tagalog. *Language* 93:e317–e352.

Shih, Stephanie S., and Kie Zuraw. 2018. The nature of the phonology-syntax interface, from variable adjective and noun word order in Tagalog. *Unpublished manuscript, USC and UCLA. lingbuzz/004296* .

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 University Press.

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

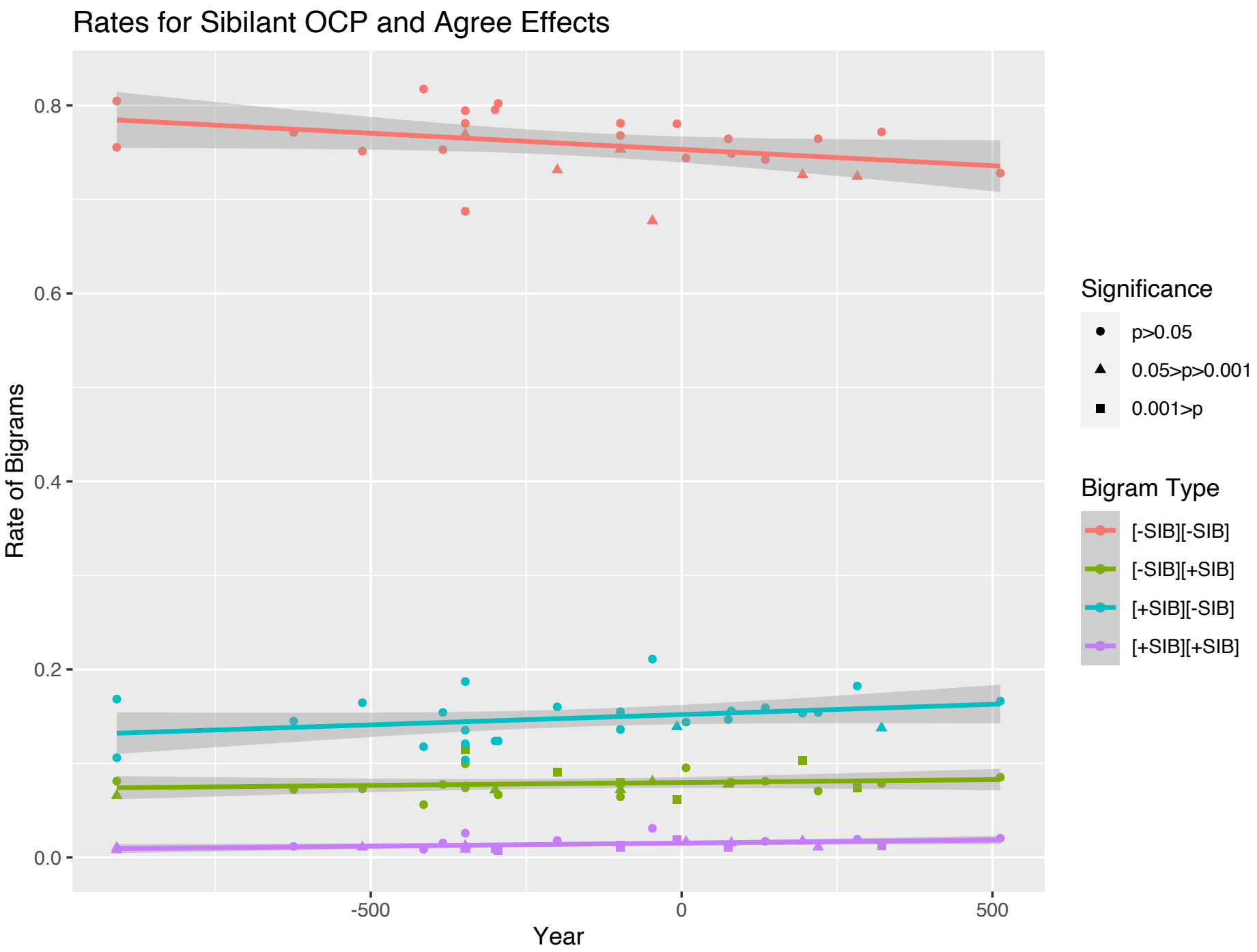

Ratios for Dorsal OCP and Agree Effects
No Tonogenesis (s and ʔ remain)
Shǎng Tone Tonogenesis (s remains)
Qù Tone Tonogenesis (ʔ remains)
Complete Tonogenesis (no post-codas remain)
Ratio of Actual to Expected Rate of Bigrams
1.2
1.1
1.0
0.9
0.8
-500
0
500
Year
Significance
p>0.05
0.05>p>0.001
0.001>p
Bigram Type
[-DOR][-DOR]
[-DOR][+DOR]
[+DOR][-DOR]
[+DOR][+DOR]

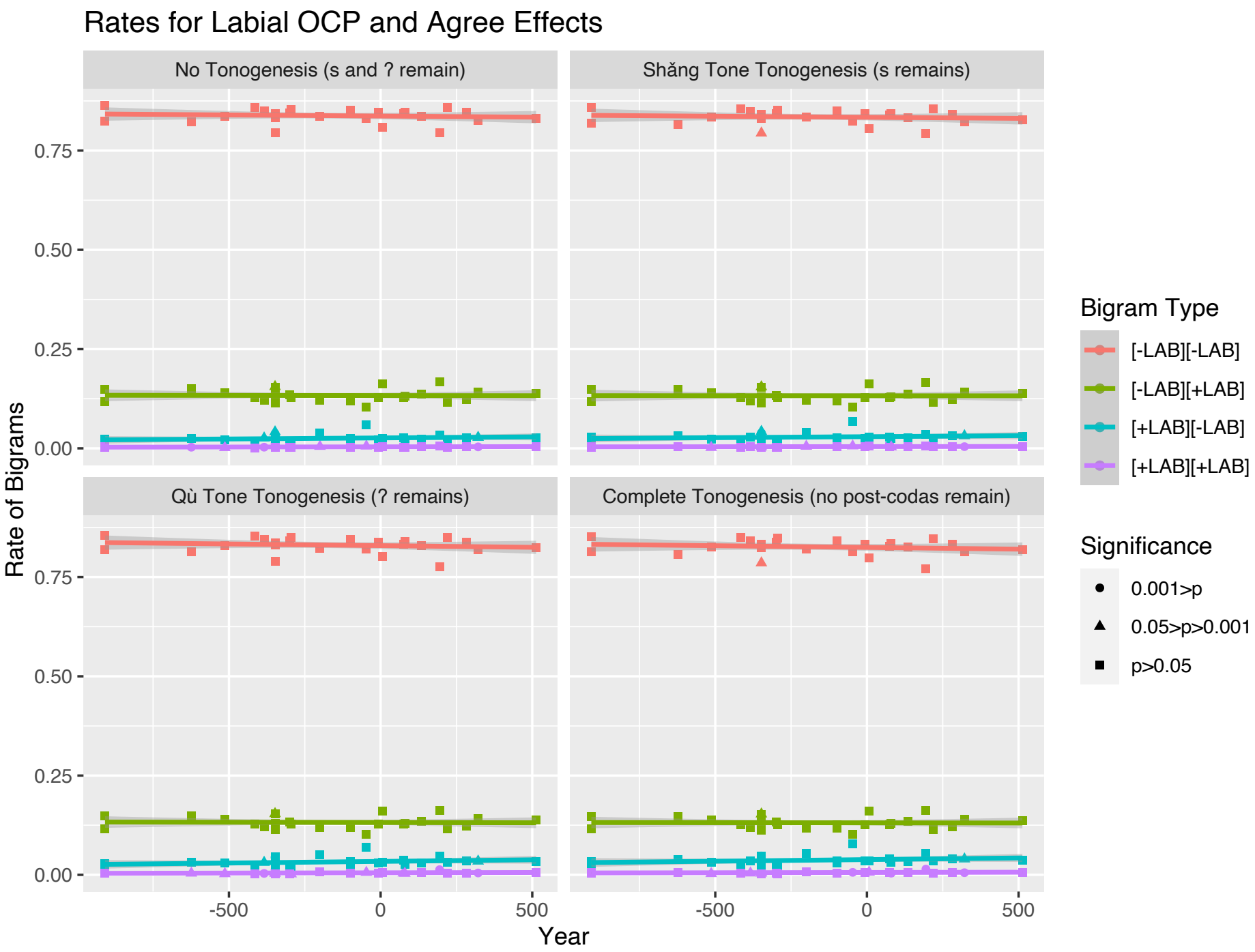
Describing OCP[Dor]

Regardless of hypothesis, OCP[Dor] bigrams are avoided. Under hypotheses where /-s/ is elided, positive slopes show that this markedness becomes less pronounced over time.

Zhōulǐ (300 BCE–100 BCE), *Jīnguì Yàolüè* (170 CE–219 CE) both seem like outliers assuming elision of /-s/ post-coda.

Bottom quadrangle: *Dàodéjīng* (475 BCE–221 BCE), *Yílǐ* (475 BCE–221 BCE), *Jīngshì Yìzhuàn* (57 BCE–37 BCE), *Tàixuánjīng* (33 BCE–18 CE)


Rates for Sibilant OCP and Agree Effects
Rate of Bigrams
0.0
0.2
0.4
0.6
0.8
-500
0
500
Year
Significance
p>0.05
0.05>p>0.001
0.001>p
Bigram Type
[-SIB][-SIB]
[-SIB][+SIB]
[+SIB][-SIB]
[+SIB][+SIB]


Rates for Labial OCP and Agree Effects
No Tonogenesis (s and ʔ remain)
Shǎng Tone Tonogenesis (s remains)
Qù Tone Tonogenesis (ʔ remains)
Complete Tonogenesis (no post-codas remain)
Rate of Bigrams
0.00
0.25
0.50
0.75
-500
0
500
Year
Bigram Type
[-LAB][-LAB]
[-LAB][+LAB]
[+LAB][-LAB]
[+LAB][+LAB]
Significance
0.001>p
0.05>p>0.001
p>0.05