Historical tone change from Middle Chinese to modern Beijing Mandarin: Usage-based phonology and modeling

Haowen Zhang

June 18, 2019

Abstract

The Middle Chinese *ru* 'entering' tone split completely into the 4 tones in modern Beijing Mandarin. Sun (2018) tried to provide explanations to this seemingly random phenomenon in his recent paper. His proposals are in fact in line with the existing theoretical framework of usage-based phonology (cf. Bybee (2003)), which assumed the effects of frequency and has been used to explain other phenomena like English schwa deletion, Spanish /s/-reduction and French liaison. We'll try to provide statistical data to test Sun (2018)'s proposals, and furthermore, we hope to build a mathematical model, which could possibly clarify this historical tone change, based the exemplar dynamics model of Pierrehumbert (2001).

1 Introduction to the problem

Middle Chinese (MC) distinguished between 4 tones, which we'll denote by *T1, *T2, *T3 and *T4 respectively.[1] Almost all the morphemes were monosyllabic, and each morpheme was associated with one tone, according to rime dictionaries like *Qieyun* (601) which allowed scholars to construct the phonology of MC. From these 4 MC tones, Beijing has developed 4 tones: T1A, T1B, T2 and T3, and the splitting and merging conditions are largely determined by the manner of articulation of the initial consonants of the syllables, as shown in Table 1 (cf. Chen (1976). The "Env" refers to environment in MC with respect to initials, and we regard ∅ as voiced and sonorant). However, we can see from the table that these conditions are inadequate to explain the distribution of *T4 syllables with voiceless initials, which split completely into the 4 tones in Beijing[2]. In fact, even a lot of homophonous pairs split (cf. Aleshire and Streeter (1970)), and one same morpheme which used to be in *T4 with voiceless initials can now have more than one variants of tones in Beijing (cf. Sun (2018)). How to account for this historical change is thus a subtle question.

Now we follow Chao (1965)'s descriptions to give some more background information about Beijing Mandarin. We use a five point scale to represent tonal shapes: point 1 represents low, 2 half-low, 3 mid, 4 half-high, and 5 high. T1A, 55, is a high-level tone; T1B, 35, high rising; T2, 214, low-dipping; T3, 51, high falling. There is another neutral tone, which is proclitic, short and

[1] In Chinese literature, they are traditionally named as *ping* 'even', *shang* 'ascending', *qu* 'departing' and *ru* 'entering' respectively.

[2] cf. Appendix for the corresponding numbers.

MC Tones \ Env	Voicelss	Voiced	
		Sonorant	Obstruent
*T1	T1A	T1B	
*T2	T2		T3
*T3	T3		
*T4	T1A, 1B, 2, 3	T3	T1B

Table 1: Tonal correspondence between MC and Beijing

the pitch depends on the preceding syllable. One morpheme in one of the four regular tones can be in the neutral tone under certain conditions. A small number of morphemes, such as suffixes and particles, are always in the neutral tone and do not belong to any one of the four regular tones. In terms of tone sandhi, the tone of a syllable is affected much more by the tone of the following syllable than by a preceding one. There are two noticeable patterns of tone sandhi: When a T2 is followed by another T2, the first T2 is realized as 35 (identical to T1B); when a T2 is followed by any other tone, this T2 is realized as 21, which is the full T2 minus its terminal rising part. There is no voicing contrast of initials in modern Beijing. The former *T1 morphemes with voiced obstruent initials now have the corresponding voiceless and aspirated initials; The former *T2, *T3 and *T4 morphemes with voiced obstruent initials now have the corresponding voiceless and unaspirated initials.

2 Review of Sun (2018)

First, we summarize Sun (2018)'s points:

1. The rime dictionary *Zhongyuan Yinyun* (1324) recorded that *T4 with voiced initials split into T1B and T3 depending on whether the initial was obstruent or sonorant, and *T4 with voiceless initials merged into T2. In fact, unlike the established splitting of *T4 with voiced initials, *T4 with voiceless initials didn't really merge into T2 at that time but formed a new tone category (which we'll call "voiceless *T4" from now on) whose tone shape was similar to that of T2. Those voiceless *T4 morphemes that are now in the T2 category are residues that haven't been subject to the other processes below but have undergone the neutralization process afterward.

2. Some formal and low-frequency voiceless *T4 morphemes changed to T3, and it is caused by the influence of the authoritative Jianghuai Mandarin: the dialect of the former Chinese capital Nanjing before the capital was moved to Beijing in 1421. In modern Jianghuai Mandarin, former *T4 morphemes now have a high tone. Furthermore, they all end with the glottal stop and thus the duration of such syllable is short. The corresponding tone in Beijing which satisfies the feature of being high and short is T3.

3. The distribution of high-frequency *T4 morphemes is caused by the lexicalization of tone sandhi. By Obligatory Contour Principle, it is plausible to suppose that there was tone sandhi concerning voiceless *T4 which preferred patterns of tonal polarity, like rising-falling and long-short.

   (a) Those high-frequency voiceless *T4 morphemes that are now in T1B are frequently followed by T2 or T3 syllables. T1B-T3 sequences show rising-falling tonal polarity.

T1B-T2 is the result of T2-T2 sandhi in modern Beijing and note that voiceless *T4 used to have a tone shape close to T2 (cf. Point 1).

(b) Those high-frequency voiceless *T4 morphemes that are now in T1A are frequently followed by neutral tone syllables. These include monosyllabic/monomorphemic verbs, which are often followed by aspectual particles: *le* (perfective), *zhe* (durative) and *guo* (experiential); nominal morphemes that often combine with the nominal suffix *zi*. These T1A-neutral tone sequences show long-short tonal polarity.

These are hints suggesting that the distribution of high-frequency voiceless *T4 into T1A and T1B is the result of lexicalized tone sandhi.

4. Some other voiceless *T4 morphemes changed to T1B by analogy. All the unaspirated voiceless *T4 are now in T1B. This could be explained by the effect of analogy: The former *T4 syllables with voiced obstruent initials first changed to T1B with unaspirated voiceless initials, and they differed from their corresponding unaspirated voiceless *T4 syllables only in tones, and the latter (unstable) changed to the former (established) by analogy. Also, changes caused by analogy happened between homophonous pairs which contained a high-frequency morpheme as described in 3a that already underwent the change to T1B.

5. The fact that one same voiceless *T4 morpheme can now have more than one tonal variants in Beijing is due to the mixed effect of the different forces above depending on its usage. The T1A~T1B variation of one such morpheme is often conditioned by whether the following syllable has neutral tone or T3, supporting the mixed effect of 3a and 3b. Another example is the T3~else variation, which is often conditioned by whether the usage is formal or colloquial.

Though the author didn't mention, his claims actually based on the theoretical hypothesis of usage-based phonology, which has already been used to explain other phenomena. This framework can also give a more unified motivation for the above processes proposed by Sun (2018). Furthermore, if Sun (2018)'s findings are true, they provide more supporting evidence for the framework of usage-based phonology. We'll go over the hypotheses and existing findings concerning this framework, and compare it with Sun (2018), trying to complement one with the other. This will be one part of our work in this paper.

Another part will be the plan of a quantitative study of this problem. Sun (2018) used many examples to justify each of his claims. However, even though he emphasized the effect of frequency, he didn't provide statistical data to justify the frequency of the concerned morphemes, words or environments, which could have been done with a corpus study. This lack of data makes his examples less convincing when he claimed that some environments were more frequent than the others. Thus we hope to complete this part by a corpus study to test and hopefully quantitatively justify his examples. Also, we want to build a mathematical model to capture this historical tone change based on the exemplar dynamics model of Pierrehumbert (2001), and see if the mixed effect of the factors mentioned in Sun (2018) can give us the desired result. If we succeed, this will be a strong proof of Sun (2018)'s claims and also give supporting evidence for usage-based phonology.

3 Usage-based phonology

Wang (1969) introduced the notion of lexical diffusion: Words change their pronunciations by discrete, perceptible increments (i.e. phonetically abrupt), but severally at a time (i.e. lexically gradual) rather than always in a homogeneous block. He and his colleagues provided supporting examples of on-going sound changes in several Chinese dialects and some other languages in Wang and Cheng (1977), Chen and Wang (1975) etc., challenging the Neogrammarians' point of view that sound change has to do with the change of types (phonemes), not of tokens (occurrences of phonemes in individual words or morphemes). In our case of *T4 distribution, the problem per se is in favor of lexically gradual change, otherwise homophonous pairs wouldn't wildly split as they appear to be now.

Later on, the effect of frequency is found to be closely related with lexical diffusion: Hooper (1976) noted that sound change seems to affect high-frequency words first, but analogical change affects low-frequency words first. Also, instead of the original assumption of lexical diffusion that a change must be phonetically abrupt, Hooper (1981) and Bybee et al. (2000) argued that sound change can be both phonetically gradual and lexically gradual. All these brought forth an exemplar model of phonological representation which predicts that the frequency with which words are used in the contexts for change will affect how readily the word undergoes a change in progress (cf. Bybee (2002)).

A complete and self-contained introduction to usage-based phonology is too long to be included in this paper. We'll rather use two examples taken from the book of Bybee (2003) to illustrate some of its claims which are related with Sun (2018).

Spanish /s/-reduction In many dialects of Spanish, an /s/ before another consonant reduces to [h], or deletes entirely. There is a tendency that this reduction of final /s/ is for the /s/ to eventually disappear in all contexts, leaving no phonetic variants of the word that depend upon the initial segment of the next word. The Argentinian dialect can be considered to be at an early stage in the implementation of this change, where the maintenance rate of /s/ is at 11-12% before a consonant, and 88% at the end of a word before a vowel. In Cuban Spanish, these two rates are 2-3% and 18% respectively, indicating that this dialect is in a more advanced stage of the change. To account for such a phonetic change which occurred outside of its phonetic environment, Bybee (2003) adopts an exemplar model of phonological representation, in which specific tokens of use are stored and categorized phonetically with reference to variables in the context. But if such memory representations are strong and stable, no further change could be expected to occur. Thus the representation is supposed to organize the exemplars such that infrequent or marginal instances are lost, but frequent ones are kept and strengthened. Then for words of medium to low frequency, more frequent exemplar types might replace less frequent ones.

French liaison The French liaison alterations originated from a phonetically conditioned consonant deletion process, but evolved to the present state as a highly lexically and morphologically governed process. In fact, this can also be explained by the frequency effect of the exemplar-based representation where all exemplars are not equally accessible: While high-frequency contexts maintain liaison longer, the loss of liaison in certain contexts corresponds to morphological regularization. Furthermore, French liaison, as a case of lexicalized external sandhi, provides evidence for the existence and nature of storage units beyond the traditional word: Frequent fixed phrases are storage and processing units, as are constructions containing grammatical morphemes.

Pierrehumbert (2001) built the exemplar dynamics model in order to capture the main findings of usage-based phonology and see quantitative predictions. Her model showed the relationship of word frequency to the progress of a leniting historical change, as well as how an unstable category collides and merges with a stable one in a situation where there is a neutralizing pressure on the system. Some more details of this model will be discussed in the next section.

Up to this point, it is already clear that Sun (2018)'s claims are very much in line with usage-based phonology. With an exemplar-based phonological representation, the splitting of voiceless *T4 to four different directions with four completely different processes described by Sun (2018) could be unified with a single motivation. Point 3 describes precisely a process of lexicalized external sandhi, and like French liaison, provides evidence of storage units of exemplars beyond the word level (e.g. verb + particle, cf. 3b), and a fortiori the morpheme level, which in return renders Point 5 unsurprising. A phonetically-conditioned regularization tendency for voiceless *T4 is described together by Point 1 and Point 4: change to T1B if unaspirated, or to T2 if aspirated. Frequency effect then determines which process would be the dominating one, just as in all the other typical examples in Bybee (2003).

In short, if we briefly rephrase Sun (2018)'s claims as follows, they then become pretty much like a typical explainable example under the framework of usage-based phonology: Except for the dialectal influence for some formal and low-frequency words/phrases, the distribution of voiceless *T4 is the result of two competing processes; while high-frequency contexts maintained and lexicalized tone sandhi, low-frequency ones were replaced through regularization.

4 Quantitative study: a plan

Our study will be corpus-based. First, as we already mentioned in Section 2, Sun (2018) didn't provide statistical data to justify his examples, even though frequency effect was the crucial factor. With a corpus study of the texts from the concerned period, we can test, for a given morpheme, if its distribution corresponds to the assumed frequency ranking of the different processes in Sun (2018). For example, we can compare the frequency of different morphemes to see if it matches the choice between lexicalized tone sandhi and regularization; for lexicalized tone sandhi, we can ask if the choice between T1A/T1B matches the more frequent environment of the two described in Point 3. These tests seem to be indispensable for Sun (2018)'s arguments.

Then, to further test if these processes can generate the change we see, we hope to build a mathematical model. Our model will be based on the exemplar dynamic model of Pierrehumbert (2001) which was applied to leniting historical change.

In her model, a category is represented by a list of remembered tokens of that category. Memories are assumed to decay so each exemplar has an associated strength which represents a resting activation level. The exemplars encoding frequent and recent experiences have higher resting activation levels than exemplars encoding infrequent and temporally remote experiences. Perception corresponds to labeling a token with a certain category. Production corresponds to selecting a particular exemplar in this category with the likelihood proportionate to its strength. There is systematic bias in production, and in her case of leniting historical change, it is the tendency to undershoot articulatory target in order to save effort and speed up communication. The effect of entrenchment is introduced so that the systematic production bias won't steadily increase the variance for a given category, and it is mathematically realized in her model by activation-weighted averaging over a group of exemplars. Then the historical change is modeled by the perception-production loop (suppose that one speaker talks to himself over time). Her model successfully predicts leniting historical change as observed by Bybee (2003).

Now we want to model the distribution of *T4 from the period of *Zhongyuan Yinyun* (1324) when all the *T1, *T2, *T3 syllables and *T4 syllables with voiced initials are supposed to have completed their changes, so our system has five tone labels: T1A, T1B, T2, T3 and *T4. A morpheme category is represented by the exemplars of words/phrases containing this morpheme. For perception, the tone label of a former voiceless *T4 morpheme is decided by sandhi rules and dialectal influence: We label it T3 if it appears in a formal context[3], otherwise T1B if followed by T2/T3; T1A if followed by the neutral tone; *T4 elsewhere. The production bias corresponds to the regularization tendency, and is mathematically represented by the increased likelihood[4] to produce a voiceless *T4 as T1B if unaspirated, or T2 if aspirated[5]. Then we can read in the corpus to introduce frequency and combine it with the perception-production loop to see if they would generate the desired result after a certain number of iterations.

If we can never get the historical change we want, or if the statistical data doesn't support Sun (2018)'s claims, we can replace the sandhi rules and/or dialectal-influence rule by unspecified parameters (e.g. a combination of some possible sandhi patterns with undetermined coefficients) and take a training set, use supervised learning to find out these parameters (i.e. phonological rules).

Besides all the parameter-tuning problems, there are still many technical problems to be considered. To name just a few:

(i) How to decide a list of formal expressions?

(ii) How to choose the natural storage units? For a parsed corpus, it is easy to choose words. But since we already have evidence that storage goes beyond the word level, one possible choice would be trigrams with the concerned morpheme in the middle, for example.

(iii) Since a Chinese corpus is usually represented by Chinese characters, and a few characters represent different morphemes with different tones, we should then consider how to automatically decide the correct morpheme from contexts. Tools from machine learning may help.

All these should be addressed in our future work.

References

Aleshire, S. and M. Streeter (1970). Homonyms from Middle Chinese. *MIM (Monthly Internal Memorandum, Phonology Laboratory, University of California, Berkeley), July issue*, 1–26.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language variation and change 14*(3), 261–290.

Bybee, J. (2003). *Phonology and language use*, Volume 94. Cambridge University Press.

Bybee, J. et al. (2000). The phonology of the lexicon: Evidence from lexical diffusion. *Usage-based models of language*, 65–85.

Chao, Y. R. (1965). A grammar of spoken Chinese.

Chen, M. Y. (1976). From middle Chinese to modern Peking. *Journal of Chinese Linguistics*, 113–277.

Chen, M. Y. and W. S.-Y. Wang (1975). Sound change: actuation and implementation. *Language*, 255–281.

[3] This dialectal-influence rule can interact with the sandhi rules so the order matters. We can also rank them differently or give different weights of likelihood for a random choice when two of the rules apply simultaneously.

[4] The number of this increase is a parameter that should be tuned, for example by a training set.

[5] Note that we actually supposed the change to be phonetically abrupt, since we don't have enough evidence concerning the phonetic details of the realization of tones at that time.

Hooper, J. B. (1976). Word frequency in lexical diffusion and the source of morphophonological change. *Current progress in historical linguistics 96*, 105.

Hooper, J. B. (1981). The empirical determination of phonological representations. In *Advances in Psychology*, Volume 7, pp. 347–357. Elsevier.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. *Typological studies in language 45*, 137–158.

Sun, J. (2018). The secondary distribution of rusheng syllables with voiceless obstruent initials in Beijing dialect. *Zhongguo Yuwen 2018, (5)*, 515–537.

Wang, W. S.-Y. (1969). Competing changes as a cause of residue. *Language*, 9–25.

Wang, W. S.-Y. and C.-C. Cheng (1977). Implementation of phonological change: The Shaung-feng Chinese case. *The lexicon in phonological change 148*, 58.

Appendix

The following table shows the number of voiceless *T4 morphemes with respect to their tones in modern Beijing Mandarin (taken from Sun (2018)).

T1A	87
T1B	47
T2	41
T3	84
T1A~T1B	22
T1A~T2	10
T1A~T3	18
T1B~T2	7
T1B~T3	5
T2~T3	9
T1A~T1B~T3	2
T1A~T1B~T2	1
T1A~T2~T3	1
T1B~T2~T3	2
Total	315