Sandhi with Chao Letters and MC Categories
(Using Rules and FSTs)

Chris Oakden

July 21, 2018

This is the basic format we're going to work with here:

- The entire inventory of lexical tones
- A rule (or rules) using tone letters, and then using traditional categories
- Consider the inverse of the rule as part of the logical possibilities (in discussion?)
- FSTs (also using letters and then traditional categories?)

In a separate section at the end, discuss generalizations

1 Rizhao (Mandarin)

Rizhao is a Mandarin dialect spoken in Shandong province.[7]

(1) Tonal Inventory

Category	Chao Letters
yinping	214
yangping	53
shang	55
qu	31

Attested disyllabic sandhi patterns are below in Chao tone letters and in MC categories. Note that many sandhi tones do not have a MC lexical equivalent, so they are designated with prime or double-prime where appropriate.

(2)

Chao Letters	MC Categories
214 → 24 / __ 214, 31	*yinp* → *yinp′* / __ *yinp*, *qu*
214 → 213 / __ 53, 55	*yinp* → *yinp″* / __ *yangp*, *shang*
53 → 54 / __ 53	*yangp* → *yangp′* / __ *yangp*
53 → 24 / __ 31	*yangp* → *yangp″* / __ *qu*
55 → 53 / __ 55	*shang* → *shang′* / __ *shang*
31 → 24 / __ 31	*qu* → *qu′* / __ *qu*

The 'elsewhere' case for *yinping* is accounted for. Surprisingly, its citation form never occurs in initial position in a disyllabic sandhi environment. The other 'elsewhere' cases are outlined below:

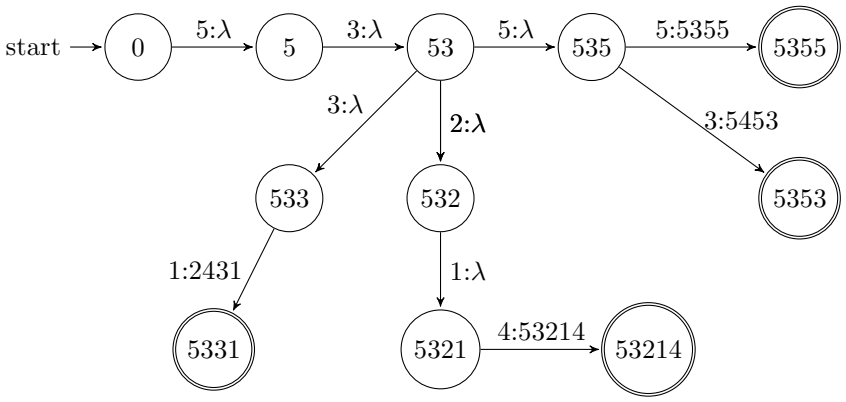
(3) a. *yangping*
53 → 53 / __ 214, 55

b. *shang*
55 → 55 / __ 214, 53, 31

c. *qu*
31 → 31 / __ 214, 53, 55

This is a preliminary attempt at constructing an FST for just the 53.T patterns.

(4)



2 Hefei (Mandarin)

Hefei is a Mandarin dialect spoken in Anhui province.[5]

(5) Tonal Inventory

Category	Chao Letters
yinping	21
yangping	55
shang	24
qu	53
ru	5q

Attested disyllabic sandhi patterns are below in Chao tone letters and in MC categories.

(6)

Chao Letters	MC Categories
21 → 24 / __ 21, 53	*yinp* → *shang* / __ *yinp, qu*
5q → 2q / __ 55, 24, 5q	*ru* → *ru′* / __ *yangp, shang, ru*

Elsewhere cases are outlined below for each tone which undergoes sandhi.

(7) a. *yinping*
21 → 21 / __ 55, 24, 5q

b. *ru*
5q → 5q / __ 21, 53

3 Yinchuan (Mandarin)

Yinchuan is a Mandarin dialect spoken in Ningxia province.[10] Strangely, the author splits *shang* into two categories based on sandhi: *shangjia* and *shangyi*. Their pronunciation in isolation is the same, but differs in sandhi contexts.

(8) Tonal Inventory

Category	Chao Letters
ping	44
shangjia	53
shangyi	53
qu	14

One sandhi alternation is reported.

(9)

Chao Letters	MC Categories
53 → 35 / __ 14	*shangyi* → *shangyi′* / __ *qu*

The elsewhere cases for *shangyi* include all other environments, including at a word boundary.

(10) a. *shangyi*
53 → 53 / __ 44, 53, #

4 Gao'an Laowu Zhoujia (Gan)

Gao'an is a Gan dialect spoken in Jiangxi province.[9]

(11) Tonal Inventory

Category	Chao Letters
yinping	55
yangping	24
shang	42
yinqu	33
yangqu	11
yinru	3q
yangru	1q

One tone sandhi alternation is reported for Gao'an:

(12)

Chao Letters	MC Categories
55 → 53 / __ 33, 31, 3q, 1q	*yinp* → *yinp′* / __ *yinqu, yangqu, yinru, yangru*

Elsewhere cases are outlined below for *yinping*:

(13) a. *yinping*
55 → 55 / __ 55, 42, 24

5 Zhangping (Min)

Zhangping is a Min dialect spoken in Fujian province.[11, 2]

(14) Tonal Inventory

Category	Chao Letters
yinping	24
yangping	11
yinshang	31
yinqu	21
yangqu	53
yinru	5q
yangru	53q

Tone sandhi in the dialect is marked by widespread neutralization.

(15)

Chao Letters
24, 11, 21, 5q → 33 / __ 24, 11, 5q, 53, 53q
24, 11, 21, 5q → 55 / __ 31, 21
31, 53, 53q → 21 / __ 24, 11, 5q, 53, 53q

(16)

MC Categories
yinp, yangp, yinqu, yinru → 33 / __ *yinp, yangp, yinru, yangqu, yangru*
yinp, yangp, yinqu, yinru → 55 / __ *yinshang, yinqu*
yinshang, yangqu, yangru → 21 / __ *yinp, yangp, yinru, yangqu, yangru*

All environments for 24, 11, 21, and 5q are accounted for. Elsewhere cases for 31, 53, and 5q include:

(17) a. *yinshang*
31 → 31 / __ 31, 21

b. *yangqu*
53 → 53 / __ 31, 21

c. *yangqu*
53q → 53q / __ 31, 21

6 Pingyao (Jin)

Pingyao is a Jin dialect spoken in Shanxi province [4, 2]. Note that *yinping* and *yangping* have the same citation form (but they are differentiated in sandhi environments).

(18) Tonal Inventory

Category	Chao Letters
yinping	13
yangping	13
shang	53
qu	35
yinru	23q
yangru	54q

Pingyao tone sandhi is complex and shows evidence of structure sensitivity. I will focus on *yangping* patterns to simplify. Note that this includes some context free patterns.

(19)

Chao Letters	MC Categories
13 → 31 / __ 35	*yangp* → *yangp′* / __ *qu*
13.53 → 35513	*yangp.shang* → *yangp′.shang′*
13.54q → 35513q	*yangp.yangr* → *yangp′.yangr′*

Yangping does not change in other environments, including:

(20) a. *yangping*
13 → 13 / __ 13, 23q

7 Yudu (Hakka)

Yudu is a Hakka dialect spoken in Shanxi province [8, 2].

(21) Tonal Inventory

Category	Chao Letters
yinping	31
yangping	44
shang	35
yinqu	22
yangqu	42
ru	5q

I will simplify the *yinqu* patterns in describing sandhi, describing only the rule in which it surfaces as 44 at word edges. In other positions, its distribution is structure sensitive, with a subset of its morphemes surfacing as *rusheng* in certain contexts.

(22)

Chao Letters	MC Categories
44 → 42 / __ 44	*yangp* → *yangp′* / __ *yangp*
35 → 31 / __ T	*shang* → *shang′* / __ T
5q → 42q / __ T	*ru* → *ru′* / __ T
22 → 44 / T __	*yinqu* → *yinqu′* / T __

Elsewhere cases are provided below (except for *yinqu*):

(23) a. *yangping*
44 → 44 / __ 31, 35, 22, 42, 5q

b. *shang*
35 → 35 / __ #

c. *ru*
5q → 5q / __ #

8 Changsha (Xiang)

Changsha is a Xiang dialect spoken in Hunan; experimental studies claim that the language has no level citation tones. [6]

(24) Tonal Inventory

Category	Chao Letters
yinping	23
yangping	13
shang	42
yinqu	45
yangqu	42
ru	24q

Three sandhi processes are reported.

(25)

Chao Letters	MC Categories
21 → 22 / __ T	*yinqu* → *yinqu′* / __ T
42 → 33 / T __ T	*shang* → *shang′* / T __ T
24q → 3q / __ T	*ru* → *ru′* / __ T

Elsewhere cases for these three tones are thus:

(26) a. *yinqu*
21 → 21 / __ #
b. *shang*
42 → 42 / # __ #
c. *ru*
24q → 3q / __ #

9 Wuyi (Wu)

Wuyi is an Wu dialect spoken in Zhejiang province. [3]

(27) Tonal Inventory

Category	Chao Letters
yinping	24
yangping	(2)13
yinshang	55
yangshang	13
yinqu	53
yangqu	31
yinru	5q
yangru	212q

Tone sandhi in Wuyi, much like Zhangping, is characterized by widespread neutralization on the initial syllable. Some sandhi forms are structurally determined, and in some cases multiple patterns are observed for syllables of the same category. What is presented below is a simplified version of Wuyi tone sandhi.

(28)

Chao Letters	MC Categories
24 → 55 / __ T	*yinp* → *yinp′* / __ T
13 → 55 / __ T	*yangp* → *yangp′* / __ T
53 → 55 / __ T	*yinqu* → *yinqu′* / __ T
31 → 11 / __ T	*yangqu* → *yangqu′* / __ T
55 → 11 / __ T	*yinshang* → *yinshang′* / __ T
13 → 11 / __ T	*yinp* → *yinp′* / __ T
212q → 5q / __ T	*yangru* → *yangru′* / __ T

Elsewhere cases are irrelevant here; all environments are accounted for, and citation forms appear in non-initial position.

10 Ningbo (Wu)

Ningbo is an Wu dialect spoken in Zhejiang province. [1]

(29) Tonal Inventory

Category	Chao Letters
yinping	53
yangping	35
yinshang	424
yangshang	313
yinqu	33
yangqu	213
yinru	5q
yangru	34q

Tone sandhi in Ningbo is rather complex, and varies based on a variety of factors, including onset voicing in some cases. Below is just an example of disyllabic forms which begin with *yangping*:

(30)

Chao Letters
35T → 2244

I'm not yet sure how to formalize this spreading rule in terms of an FST.

11 Discussion

Restricting ourselves to Chao tone letters, permutations of citation forms in sandhi processes are broadly divisible into two types: right-edge modification and total modification.

(31)

Total Modification		Right-edge Modification	
Pattern	Dialect	Pattern	Dialect
53 → 24	Rizhao	214 → 213	Rizhao
31 → 24	Rizhao	55 → 53	Gao'an
24 → 33	Zhangping	53 → 54	Rizhao
24 → 11	Zhangping	21 → 24	Hefei
5q → 42q	Yudu	35 → 31	Yudu
22 → 44	Changsha	44 → 42	Yudu
42 → 33	Changsha	21 → 22	Changsha
24q → 3q	Changsha	13 → 11	Wuyi
24 → 55	Wuyi	53 → 55	Wuyi
13 → 55	Wuyi		
55 → 11	Wuyi		

With few exceptions (like 21 → 24, an analog of the Mandarin 3rd tone sandhi rule, also analyzable as L → LH), the right-edge modification patterns can be characterized as either contour simplification or contour formation. In other words: a level tone becomes a contour, or a contour tone becomes level by modifying the right edge of the tone. This type of pattern is pervasive across dialects.

Three patterns do not (necessarily) fit into either generalization. Two are "inversion"-like patterns on contours (or "contour metathesis" in Chen's terminology): 53 → 35 (Yinchuan) and 13 → 31 (Pingyao). The pattern 31 → 11 (Wuyi) is ostensibly a "left-modification" pattern, the only one that emerged from the survey of these dialects. I would like to suggest that all three are cases of total modification.

The first two inversion patterns are cases of contour switching, so, falling-to-rising and rising-to-falling respectively. In the total modification patterns, the same situation obtains, albeit with non-symmetric phonetic values. Examples include 53 → 24 and 31 → 24 (falling-to-rising) in Rizhao. In fact, all logical possibilities are observed in the total modification alternations: contour to level, level to contour, level to level, and contour to contour. The "inversion" patterns fit into this schema.

To determine where 31 → 11 fits into this pattern, it is worthwhile to consider what *types* of alternations total vs right modification patterns constitute. All the (arguably) syllable-level OCP sandhi patterns are of the right-edge modification type:

(32)

Modification	Environment
214 → 213	__ 214
53 → 54	__ 53
21 → 24	__ 21
44 → 42	__ 44

The majority of total modification patterns, by contrast, are cases of extensive neutralization in non-final position or other non-OCP motivated alternations. These are examples like Zhangping and Wuyi, where multiple tonal categories converge on the same phonetic output in these positions. Viewed in isolation, the

pattern 31 → 11 appears to be a case of left-edge modification; however, when viewed in its context as a Wuyi neutralization pattern (31, 55, 13 *all* map to 11), a total modification is more felicitous.

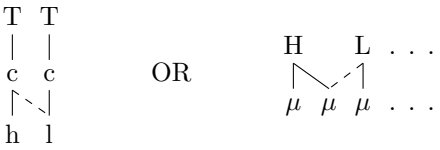
Where does this leave us in terms of possible vs impossible sandhi patterns? One generalization that seems to emerge from the data is: there is no such thing as a "left-edge modification" in right-dominant sandhi patterns, at least not for the syllable-level OCP cases. That is to say, the locus of alternation is the immediate environment between the first and second syllable (note that this is also true of Tianjin and Nanjing, not discussed here). There is, then, an interesting asymmetry in the sandhi patterns under our current assumptions. In terms of Chao tone letters, modification of citation tones either targets the entire tone (usually in neutralization contexts, but not always) or it targets the right edge, but never the left edge. Why?

To determine whether this is a locality effect, it will be necessary to analyze left-dominant (non-spreading) patterns and look for the inverse: left-modification but no right-modification. But there are more general questions to ponder about what we mean by *locality*, as well. As a string of symbols, why should 55 → 35 / __ 55 be non-local (or somehow less local) in comparison to 55 → 53? It is clear that, as ISL functions, their k-value is the same (in this case k = 4). Additionally, representation in terms of single units (H, L, R, F) obscures any distinction whatsoever:

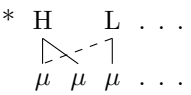
(33) a. 55 → 53 __ 55 ≡ H → F __ H
b. 55 → 35 __ 55 ≡ H → R __ H

With this in mind, the crucial difference lies not in the available window the function needs to map an input to an output, but rather *where* in the string the two mappings share a discrepancy.

How would other theories or representations account for this discrepancy? Autosegmental theories can account for local *spreading* patterns via the No-Crossing Constraint. The Gao'an pattern 55 → 53 / __ 33, 31, ... is simply a case of local regressive spreading:

(34)


In principle, 55 → 35 as a case of spreading would be ill-formed.

(35)


But what about dissimilation, for example Mandarin 3rd tone sandhi (11 → 24 / __ 11) in which we could argue tonal epenthesis (L → L**H**) obtains? As far as I am aware, ARs have nothing to tell us about why L → LH and not *L → HL.

In an optimization framework, there are a variety of approaches to privilege right-edge modification over left-edge modification. In my thesis I used a high-ranking ANCHOR constraint to prefer L → LH.

(36) L-ANCHORT-[σ]- Any tonal element at the left edge of a syllable in the input has a correspondent at the left edge of the syllable in the output

The problem with the above analysis is that, thinking of ANCHOR constraints in terms of a broad typology, we actually predict its inverse; the existence of L-ANCHORT-[σ] entails another constraint R-ANCHORT-[σ] which under a number of viable total orders would produce the unattested *L → HL. Our goal here is not to argue the specifics of an OT analysis of tone, though. What is important to consider is that this approach crucially uses *faithfulness* to derive the attested patterns. In a computational analysis, is there a way to model this?

Below are a few miscellaneous points regarding the survey of tone sandhi.

- Also take a look at trisyllabic sandhi in Wuyi, because it looks a lot like NJH (ISL)
- One of the generalizations you lose from just using tone categories is that there are often phonetically identical sandhi forms which are the same across different categories, so calling an output form xxx′ is opaque in a sense. just because citation form of category x is phonetically identical to the sandhi form of category y, do we want to give them the same representation?
- Underlying forms are important! It is often the case that the UR of a citation form only shows up in isolation. so then how do we represent it?
- there is evidence suggesting that underlying representation is arbitrary, especially in neutralization. but then there are others which suggest some sort of local spreading or dissimilation
- phonetic representation explodes the value of k

References

[1] Ning-Ping Chan. "Ningbo fangyan de biandiao xianxiang [Tone sandhi phenomenon in Ningponese]". In: *Fangyan* 1985.1 (1985), pp. 15–27.

[2] Matthew Y Chen. *Tone sandhi: Patterns across Chinese dialects*. Vol. 92. Cambridge University Press, 2000.

[3] Guotong Fu. "Wuyi fangyan de liandu biandiao [Tone sandhi in the Wuyi dialect]". In: *Fangyan* 1984.2 (1984), pp. 109–127.

[4] Jingyi Hou. "Pingyao fangyan de liandu biandiao [Tone sandhi in the Yinchuan dialect]". In: *Fangyan* 1980.1 (1980), pp. 1–14.

[5] Huifang Kong. "Hefei fangyan shengdiao de youxuanlun fenxi [An OT analysis of tones in the Hefei dialect]". In: *Hefei Xueyuan Xuebao* 25.2 (2008), pp. 30–32.

[6] Bing Li and Yanni Liu. "Changsha fangyan danzidiao ji biandiao de shiyan yuyinxue baogao[An acoustic study of citation tones and tone sandhi in Changsha Chinese]". In: *Hunan Daxue Xuebao (Shehui kexue ban* 4 (2006), p. 19.

[7] Xiuling Wu and Dongbing Liu. "Shandong Rizhao fangyan zhong de liandu biandiao [Tone sandhi in the Rizhao dialect of Shandong]". In: *Xiandai Yuwen* 2010.1 (1981), pp. 106–107.

[8] L. W. Xie. "Jiangxisheng Yudu fangyan de liangzizu liandu biandiao [Disyllabic tone sandhi in the Yudu dialect of Jiangxi province]". In: *Fangyan* 1992.3 (1992), pp. 222–227.

[9] Sen Yan. "Gaoan (laowu zhoujia) fangyan de yuyin xitong [The phonetic system of the Gao'an (Laowu Zhoujia) dialect". In: *Fangyan* 1981.2 (1981), pp. 104–121.

[10] Shengyu Zhang. "Yinchuan fangyan de shengdiao [Tones in the Yinchuan dialect]". In: *Fangyan* 1984.1 (1984), pp. 19–26.

[11] Zhenxing Zhang. "Zhangping (Yongfu) fangyan de liandu biandiao [Tone sandhi in the Zhangping dialect]". In: *Fangyan* 1982.3 (1983), pp. 175–196.