A Unified Account of Mandarin Tone 3 and Tone 4 Sandhi

Hui Yin
Faculty of International Languages, Xiamen University of Technology, China
yinhui@xmut.edu.cn

Abstract
This paper aims to provide a unified account of tone 3 and 4 sandhi. Some linguists have noticed there are obvious phonetic differences between sandhi tone 3 and tone 2 and that they are not neutralized. Traditional analyses viewing tone 3 sandhi as categorical change fail to account for differences between sandhi tone 3 and tone 2. In this paper a constraints-based analysis within the Optimality Theory (OT) framework is developed to look into Mandarin tone sandhi and to explain differences between sandhi tone 3 and tone 2. A new analysis claiming that tone 3 sandhi is a tonal reduction is proposed and justified within the OT framework. The OT device of constraint conflicts permits the present analysis of tone 3 sandhi as a resolution of conflicts between a general principle of markedness and a general principle of tonal faithfulness. This approach provides straightforward descriptions and motivations for tone 3 sandhi. This paper also shows tone 4 sandhi can be accounted for within the OT framework. In the present analysis both tone 3 sandhi and tone 4 sandhi are viewed as tonal reductions and the same basic set of constraints are used to accommodate both tone sandhi phenomena.

Index Terms: tone, sandhi, OT, constraint

1. Introduction
There are four basic tone categories in Mandarin Chinese. Every stressed syllable belongs to one of the four, although actual realizations may vary according to context (Ch’en et al., 1994). When the four tones are applied to the same syllable, four different lexical items may result (Wang & Norval, 1997). The following is the classic example of tonal contrasts:

(1) syllable | tone | gloss
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ma</td>
<td>1:</td>
<td>‘mother’</td>
</tr>
<tr>
<td>ma</td>
<td>2:</td>
<td>‘hemp’</td>
</tr>
<tr>
<td>ma</td>
<td>3:</td>
<td>‘horse’</td>
</tr>
<tr>
<td>ma</td>
<td>4:</td>
<td>‘scold’</td>
</tr>
</tbody>
</table>

In analyses of the Mandarin tonal system, the voice range is divided into five levels: upper, upper middle, middle, lower middle and lower. The numbers from 1 to 5 is used to designate these levels, where 1 represents the lowest comfortable pitch of the vocal range and 5 represents the highest. The first tone (55) is high and level. It is pitched near the top of the comfortable voice range. The second tone (35) starts around the middle of the voice range 3 and rises straight towards the level of the first tone 5. The third tone (214) begins near the bottom of the comfortable voice 2, proceeds to the bottom 1, and then upward to end above the middle 4. The fourth tone (51) begins at the top of the comfortable range 5 and falls quickly to the bottom 1 (Ch’en et al., 1994).

A well-known phenomenon in Mandarin phonology is tonal variations according to different phrasal contexts, among which tone 3 sandhi is the most noticeable and most complicated. There is another tone sandhi which is called tone 4 sandhi. In my analysis, both tone 3 sandhi and tone 4 sandhi can be regarded as a kind of tonal reduction. I will show that the same basic set of constraints can be employed to accommodate both tone sandhi to provide a unified account. In the next section, I will discuss both tone 3 sandhi and tone 4 sandhi.

2. Tone 3 and tone 4 sandhi

In rule-based analyses the basic rule of Mandarin tone 3 sandhi can be simply stated as in (2):

(2) 3---->[2]/___3

When one third tone syllable is immediately followed by another, the first of the two changes from a “falling-rising tone [214] into high rising [35]” (Chen, 2000:20) i.e. tone 2, which is indicated by brackets in rule-based analyses throughout the paper. However, some linguists (e.g. Hockett, 1950; Martin, 1963) claim that the changed third tone did not really become a tone 2 but rather a new tone (tone 5), which will be dealt with in detail in the OT analysis part of this paper. Phonetically, there are some differences between the changed tone 3 and the underlying tone 2 though it sounds like an underlying tone 2 (e. g. Wang & Li, 1967). There are a number of linguists who have demonstrated this kind of phonetic differences (their data mainly come from Beijing Mandarin or mainland Mandarin). Kratochvil (1984) claims that T3 + T3 seems to be quite distinct from T2 + T3 phonetically and his experiment (1987) shows that the first T3 in a two tone 3 sequence is more often classified as T3 rather than T2. Zee (1980) reports that the mean F0 for sandhi tone 3 is 17.5 Hz, lower than that of T2 and he concludes that sandhi tone 3 is not identical to underlying tone 2. Xu (1997) also has data which clearly indicate that the third sandhi results in a tone lower than tone 2. However, these phonetic differences have received little attention in traditional third tone sandhi analyses.

Most, if not all, of the rule-based approaches fail to answer some fundamental questions regarding tone 3 sandhi such as the following:

(i) How can we account for the differences between the so-called derived tone 2 from tone 3 sandhi and the lexical tone 2?
(ii) Why does it happen that the first tone 3 undergoes tone 3 sandhi instead of the last one in a 33 ton sequence?

In the recent literature (e. g. Zhang, 1997; Duanmu, 1999; Lin, 2002; Yip, 2002; Wee, 2008 & 2010), there are some OT analyses regarding tone 3 sandhi, and these analyses mainly deal with tone 3 sandhi domains and/or the role of speech rate in tone 3 sandhi. However, the afore-mentioned basic questions have yet to be answered.
There is another sandhi rule which has not received as much attention as the third tone sandhi.

(3) 4-----[4Half]---->4

This rule is very similar to that of third tone sandhi: when a fourth tone follows another fourth tone, it retains its first half (53) but loses its second half (31).

3. OT account of tone 3 and 4 sandhi

In the present analysis, both third tone sandhi and fourth tone sandhi are regarded as tonal reduction phenomena, which can be accommodated by the results of conflicts of faithfulness and markedness constraints. In this paper, a set of basic constraints are proposed to give a unified account for both tone 3 sandhi and tone 4 sandhi phenomena, which has never been done before in Mandarin phonology.

3.1. OT account of tone 3 sandhi

In this section, I will propose an alternative analysis to address the third tone sandhi in Mandarin Chinese. With markedness and faithfulness constraints within the OT framework, this analysis provides a straightforward account for the third tone sandhi.

As mentioned previously, some linguists believe that the changed third tone in a two third tone sequence does not really become a tone 2 but rather a new tone. Some rule-based analyses incorrectly predict that a tone 3 which undergoes such kind of sandhi is phonetically indistinguishable from an underlying tone 2. Hockett (1950) claims that phonetically, the first third tone in such a sequence has the terminal rise in pitch characteristic of tone 3. It has been realized by some linguists (e.g., Hockett, 1950; Martin, 1963; Xu, 1997) that there are noticeable differences between the changed third tone and the underlying second tone though there are some perceptual confusions (e.g., Li, 1967; Shih, 1997) between them. A number of linguists (e.g., Zee 1980; Shen 1990; Xu 1993) have data that indicate that the third tone sandhi results in a tone lower that tone 2. In fact, the sandhi tone 3 does not really neutralize to tone 2 and instead it keeps its rising part (the second half). Thus, this phenomenon could be viewed as a kind of tonal reduction.

The present analysis is based on the OT framework. An argument in favor of OT could be constructed. The OT device of constraint conflict permits an analysis of tone 3 sandhi as a resolution of conflict between a general principle of effort minimization and a general principle of tonal faithfulness. This could largely explain the loss of the falling portion of tone 3 in a 33 sequence and the loss of its rising portion in other contexts. As mentioned previously, the full third tone contour is 214. Thus, I will use the notations of [214] for a full tone 3, [14] for tone 3 with loss of the falling portion and [21] for tone 3 with loss of the rising portion in the OT analysis in this paper.

The basic idea of OT is that “each linguistic output is optimal, in the sense that it incurs the least serious violations of a set of conflicting constraints” (Kager, 1999:8). Outputs are results of conflicts between markedness and faithfulness. Surface variations could result from rankings of relevant constraints. The OT framework is based on markedness constraints and faithfulness constraints. Such constraints are violable, but violation must be minimal. “Markedness constraints require that output forms meet some criterion of structural well-formedness” (Kager, 1999: 9). In third tone sandhi in Mandarin, one of the important markedness constraints is final lengthening/non-final shortening in a phonological phrase.

(4) Final-L/N-Final-S

In a phonological phrase, the final syllable should be lengthened and the non-final syllable(s) should be shortened.

This markedness constraint is well motivated cross-linguistically (e.g., Nooteboom, 1997, Duanmu, 2000). In the case of Mandarin Chinese the final syllable in a phonological phrase usually indicates a major boundary.

Another markedness constraint active in the third tone sandhi is the effort minimization constraint:

(5) *ComCT/ShortS

Complex contour tones should be prohibited in shortened syllables.

In Mandarin tone 3 consists of two parts: a falling portion and a rising part. Thus, it can be regarded as a complex contour tone. The tone 3 sandhi process could be characterized as a tonal reduction (loss of the first part—the falling portion of tone 3 in a 33 sequence) in contexts where the syllable has a shortened or non-lengthened duration. In other contexts, tone 3 will lose its rising portion (second part) due to a markedness constraint: the T-POLARITY constraint, which requires an initial tone to be followed by an opposite tone (Duanmu 1999). The phenomenon for a tone 3 to lose its second half will be dealt with in detail later. Duanmu (1999) claims that the duration of the third tone is longer than other tones. If enough duration cannot be maintained, the full third tone as a complex contour tone will not be realized. This is a consequence of the effort minimization principle – the shortened duration increases the effort cost of achieving all 3 tonal targets. So it is reasonable to assume that the tone 3 sandhi is a kind of tonal reduction process.

Faithfulness constraints require that “outputs preserve the properties of their basic (lexical) forms, requiring some kind of similarity between the output and the input” (Kager, 1999). The relevant faithfulness constraint in third tone sandhi is one of the correspondence constraints.

(6) Max(contour)

Correspondent segments in input and output have identical values for tone contours. In the case of third tone, the contour refers to both the falling part (I use contour numbers 21 to represent it) and the rising part (I use contour numbers 14 for this part).

Thus we can derive two sub-constraints of correspondence regarding tone 3:

(a) Max(21)

Input segments must have output correspondents in terms of the falling part of tone 3.

(b) Max(14)

Input segments must have output correspondents in terms of the rising part of tone 3.

In this paper, the numbers [14] are used to represent the underlying representation of the rising portion of tone 3 but besides [14] its actual surface realization could be [24] or [34] as allographs due to some factors such as further reductions in non-final contexts or individual speakers’ differences. In any
way, all the three allophones maintain the basic characteristics of its rising portion of the third tone. For the sake of simplicity, I use \([14]\) to characterize the rising portion of the third tone in this paper. (Another way to solve the problem is to simply use \([R]\) instead of \([14]\) to represent the rising portion of tone 3 to cover all the allophones.)

In Mandarin, when Final-L/Non-Final-S and *ComCT/ShortS rank higher than Max(contour), third tone sandhi rule will apply as the following example hao3 jiu3 ‘good wine’ illustrates. (In the input, I will specify which tone it is and in the output I will use contour numbers to represent it, e. g. \([214]\) for a full tone). Following traditions of OT, relevant constraints are listed on the top row horizontally in a descending ranking from left to right. However, the dashed line shows that there is no crucial ranking between the constraint that the input is shown in the top left cell. The output candidates are indicated in the first column. An asterisk mark ‘*’ indicates a violation of a constraint heading the column and an exclamation mark ‘!’ indicates a violation that rules out a candidate. The optimal candidate is marked by the index ‘☞’. Shaded cells mean that they are no longer relevant for the evaluation.

### Table 1. Evaluation of candidates for /hao3 jiu3/

<table>
<thead>
<tr>
<th>Input /hao3 jiu3/</th>
<th>Final-L/N-Final-S</th>
<th>*ComCT/ShortS</th>
<th>Max (21)</th>
<th>Max (14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>☞ a. (hao[14]) jiu([214])</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☞ b. (hao[14]) jiu([14])</td>
<td>!</td>
<td></td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>☞ c. (hao[214]) jiu([214])</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>☞ d. (hao[214]) jiu([214])</td>
<td>!</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In tables in this paper, a parenthesis represents a phonological phrase boundary and bold letters represent the lengthening. From Table 1, we can see that the optimal candidate is (a): hao[14] jiu[214] which violates the lower ranked Max(contour) but satisfies Final-L/Non-Final-S and *ComCT/ShortS. Here we can discern one of the essential assumptions made by OT: constraints are violable when higher constraints need to be satisfied.

However, when Max(contour) dominates both Final-L/Non-Final-S and *ComCT/ShortS the third tone sandhi rule would be blocked and thus the optimal candidate in that case would be (c): hao[214] jiu[214] instead of (a): hao[14] jiu[214] though such a ranking does not happen in Mandarin third tone sandhi. The above example shows that different rankings could result in different outputs.

Then how can we account for the fact that the rising part rather than the falling part drops when tone 3 precedes any other tone? Here still another markedness constraint is at work. I adopt Duanmu’s constraint: T-POLARITY (1999) to explain this phenomenon (and also explain the loss of the falling part rather than the rising part of tone 3 in 33 sequences):

(7) T-Pol

A tonal polarity constraint requires an initial tone to be followed by an opposite tone, that is, a low tone should be followed by a high tone and vice versa.

This constraint is similar to Yip’s OCP solution to tone 3 sandhi (Yip, 2002).

In Mandarin tonal system, tone 3 is a complex tone which could be viewed as a low tone (falling part) plus a high tone (rising part) whereas all the three other tones basically can be regarded as high tones. Duanmu (1999:23) points out that a monosyllabic tone 3 in Mandarin usually surfaces as two parts, which ends in a H (high tone). Chao (1933:132) also claims that a monosyllabic tone 3 “often breaks into two parts” with a glottal stop in between. As for the other tones, tone 1 (55) and tone 2 (35) as a whole can be viewed as high tones. In tone 4 (51), the first portion beginning with (5) is a high tone. So when tone 4 follows tone 3 the most adjacent portion to tone 3 is a high tone in a tone 3 and tone 4 sequence.

The following table illustrates that tone 3 retains its falling part rather than its rising part in the example of hao3 shu1 ‘good book’:

### Table 2. Evaluation of candidates for /hao3 shu1/

<table>
<thead>
<tr>
<th>Input /hao3 shu1/</th>
<th>Final-L/N-Final-S</th>
<th>*ComCT/ShortS</th>
<th>T-Pol</th>
<th>Max (21)</th>
<th>Max (14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>☞ a. (hao[21] shu[55])</td>
<td></td>
<td>!</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>☞ b. (hao[14] shu[55])</td>
<td></td>
<td></td>
<td>!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Table 2, candidate (b) violates the constraint T-POLARITY while (a) abides by T-POLARITY constraint. It can be seen from the table that candidates (a) incurs the least serious violations though it violates the lower ranked Max(14) constraint and accordingly, it is the best candidate.

Then how can we account for the phenomenon that in a two tone 3 sequence the first tone 3 should retain its rising portion instead of its falling portion? With the same ranking as in Table 2, we can derive its ideal output straightforwardly.

### Table 3. Evaluation of candidates for /hao3 ma3/

<table>
<thead>
<tr>
<th>Input /hao3 ma3/ ‘good horse’</th>
<th>Final-L/N-Final-S</th>
<th>*ComCT/ShortS</th>
<th>T-Pol</th>
<th>Max (21)</th>
<th>Max (14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>☞ a. (hao[14] ma[214])</td>
<td></td>
<td>!</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>☞ b. (hao[21] ma[214])</td>
<td></td>
<td></td>
<td>!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As illustrated in Table 3, in the example of hao ma ‘good horse’, the second syllable ma ‘horse’ begins with a low tone (21) since it gets lengthened in the final position and thus, it should be preceded by an opposite tone, a high tone to satisfy the T-POLARITY constraint. As a result, the first tone 3 in a 33 sequence should keep its rising portion rather than its falling portion to be a winning candidate.
3.2. OT account of tone 4 sandhi

There is another sandhi rule which receives some attention (e. g. Chao 1968; Ch’en et al. 1994) but not as much as third tone sandhi does. This rule looks very similar to tone 3 sandhi rule. In a sequence of two fourth tones the first fourth tone begins where the full fourth tone begins and falls only about half as far as the full fourth tone (Ch’en et al. 1994). Like third tone sandhi, fourth tone sandhi can be regarded as a tonal reduction process since the sandhi tone retains its first half (53) but loses its second half (31) as in the following example:

(8) kan(51) bao(51) ----> kan(53) bao(51)
read newspapers

Fourth tone sandhi can also be accommodated by the results of conflicts of faithfulness and markedness constraints. There are two basic constraints: the Final-L/Non-Final-S constraint and the MAX(contour) constraint, which work in both third tone sandhi and fourth tone sandhi. In the case of the fourth tone, the contour can be analyzed to consist of two parts and as such it can refer to both the first half (I use contour numbers 53 to represent it) and the second half (I use contour numbers 31 for this part). Thus we can derive two sub-constraints of correspondence regarding tone 4:

(9) (a) \[ \text{MAX(53)} \]
Input segments must have output correspondents in terms of the first part of tone 4.

(b) \[ \text{MAX(31)} \]
Input segments must have output correspondents in terms of the second part of tone 4.

In addition to the Final-L/Non-Final-S constraint and the MAX(contour) constraint, there is another markedness constraint, which works in both third tone sandhi and fourth tone sandhi as well. This constraint is *ComCT/ShortS, which does not allow complex contour tones in a non-final position in a phonological phrase.

Actually this markedness constraint, which does not allow a full fourth tone in a non-final position in a sequence of two fourth tones, is phonetically motivated. Among all the four tones, the fourth tone has the highest pitch at its initial portion. Xu (1997) has demonstrated that when tones are produced in isolation, tone 4 has the highest \( f_0 \) value (140 Hz) of the four tones at the start point and tone one is the second highest (a little below 130 Hz) at its initial portion. Therefore, if tone 1 is viewed as a high tone the initial portion of tone 4 could be regarded as the super high. As a result, if we use the number 6 to indicate the super high pitch, it could be justified in using 61 instead of 51 to describe tone 4.

In a two fourth tone sequence, the distance of pitch between the end of first tone 4 and the beginning of the second tone 4 is the greatest. In order to reach the super high pitch of the second tone 4, the first tone 4 which is in a non-final position doesn’t fall all the way to the bottom (1) instead it falls half way to (3). In order to ease articulation as in the case of Canadian raising (Myers 1997), the half fourth tone comes into being in this way.

Now with the two markedness constraints (Final-L/Non-Final-S and *ComCT/ShortS) and one faithfulness constraint (MAXcontour) and with their proper ranking, we can derive correct outputs regarding the fourth tone sandhi.

Table 4 shows that candidate (b) violates the markedness constraint *ComCT/ShortS which is ranked higher than the faithfulness constraint MAX(31). As a result, candidate (b) is worse than candidate (a) since this candidate only violates the lower ranked MAX(31) constraint and it survives to be the ideal candidate.

Some linguists (e. g. Moira Yip p.c.) point out that when a tone 4 is followed by a tone other than a tone 4 it also gets reduced somehow rather than falls to the bottom (1). However, the reduction in this case is so small that it often could not be noticeably perceived as such and thus the tone with such small reductions are still often treated as a full tone. Therefore, when speaking of tone 4 sandhi we usually refer to the situation in a two tone 4 context. However, the reduction phenomenon in either case renders supports to the Final-L/Non-Final-S constraint because the final tone 4 does not undergo reductions since it gets lengthened but the first (non-final) tone 4 undergoes reductions since it gets shortened so that it might not have enough duration to achieve all its underlying targets.

4. Conclusions

This paper investigates the most complicated tonal change in Mandarin: the third tone sandhi. A constraints-based analysis within the OT framework is developed in this paper to look into third tone sandhi in Mandarin. With proper rankings of markedness constraints and faithfulness constraints, this approach provides straightforward descriptions for third tone sandhi and can cover all the data including those which traditional approaches fail to account for. Thus this paper demonstrated that with the constraints-based analysis within the OT framework, third tone sandhi in Mandarin can be fully accounted for.

This paper also addressed fourth tone sandhi, which has not received much attention. It has been shown in this paper that fourth tone sandhi can also be accounted for within the OT framework. In the present analysis, both tone 3 sandhi and tone 4 sandhi are viewed as tonal reductions instead of categorical changes and thus, the same basic set of constraints are demonstrated to accommodate both tone sandhi phenomena. Since the applications of the two kinds of tone sandhi both lead to their half tones it would be reasonable to analyze the sandhied tones as phonetic rather than phonological variants of the unsandhied forms. Tone 3 which has undergone the sandhi retains its rising portion (second half) rather than neutralizes to tone 2, and thus, it is naturally different from underlying tone 2 phonetically.

In a word, according to OT, constraints are violable and the optimal candidate should incur least serious violations of these constraints. In fact, outputs are results of conflicts between markedness constraints and faithfulness constraints. Surface variations could result from proper rankings of relevant constraints.

<table>
<thead>
<tr>
<th>Input</th>
<th>Final-L/Non-Final-S</th>
<th>*ComCT/ShortS</th>
<th>MAX(53)</th>
<th>MAX(31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) [ kan(53) bao(51) ]</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>(b) [ kan(51) bao(51) ]</td>
<td></td>
<td></td>
<td></td>
<td>!</td>
</tr>
</tbody>
</table>

Table 4. Evaluation of candidates for /kan51 bao51/
5. References