

# John Benjamins Publishing Company



This is a contribution from *Linguistics in the Netherlands 2014*  
© 2014. Algemene Vereniging voor Taalwetenschap  
All rights reserved

This electronic file may not be altered in any way.  
The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.  
Permission is granted by the publishers to post this file on a closed server which is accessible only to members (students and faculty) of the author's/s' institute. It is not permitted to post this PDF on the internet, or to share it on sites such as Mendeley, ResearchGate, Academia.edu. Please see our rights policy on <https://benjamins.com/#authors/rightspolicy>  
For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: [www.copyright.com](http://www.copyright.com)).  
Please contact [rights@benjamins.nl](mailto:rights@benjamins.nl) or consult our website: [www.benjamins.com](http://www.benjamins.com)

# The four tones of Mandarin Chinese

## Representation and acquisition\*

Jeroen van de Weijer<sup>1</sup> and Marjoleine Sloos<sup>2</sup>

<sup>1</sup>Shanghai International Studies University / <sup>2</sup>Aarhus University

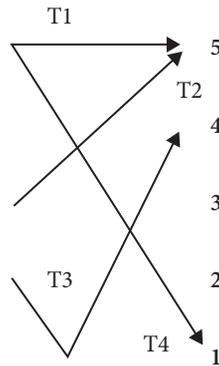
In this paper we discuss the four tones of Standard (Beijing) Mandarin Chinese. First, we will suggest a proposal for their phonological representation. Then, we discuss the order in which they are acquired in first language acquisition, relating this both to the representations we propose and to the relative frequency with which these tones appear in the most frequently used Chinese words. It turns out that the former predicts the order of acquisition more closely than the latter: We provide an explanation for this based on the early stage at which the tones are acquired.

**Keywords:** Tone, Mandarin Chinese, L1 acquisition, phonological representation, frequency

### 1. Introduction

As is well known, Mandarin Chinese has four lexical tones, abbreviated as T1-T4, and also a fifth, 'neutral' tone (T0 or T5), which we will regard as toneless. The realisation of T1-T4 is generally agreed to be as in Figure 1 (overleaf) (see also Figure 1 in Wong 2012), where the tone levels are represented using the numbers 1-5, the so-called 'Chao digits' (Chao 1930).

In this paper, we discuss the acquisition of these tones, with particular attention paid to the order in which they are acquired. We first discuss our expectations as to order of acquisition, both on the basis of (proposed) phonological representations (and phonetic realisations) of the tones, and on the basis of the frequency with which they occur in Mandarin. Then we discuss what is known from the literature about the order of acquisition of the tones. Although frequency of occurrence of a particular sound, or pattern, predicts order of acquisition in many cases, this does not apply in this case. Their phonological representations, on the other hand, predict the order of acquisition of the four tones quite well. We suspect



**Figure 1.** The four tones of Mandarin Chinese

the reason for the fact that frequency turns out to be largely irrelevant is that tone is acquired very early in the acquisition process — when the lexicon is still quite small — so that no meaningful generalisations can be made on the basis of the set of acquired words.

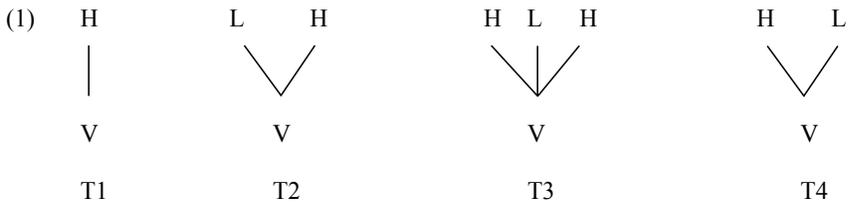
This paper is organised as follows: In Section 2, we propose underlying representations of the four tones, on the basis of their citation realisations. We discuss the merits and drawbacks of these representations, and present a brief analysis of the tone sandhi rules that apply in Mandarin. Section 3 seeks to establish the frequency of the tones in Mandarin on the basis of a corpus. Section 4 briefly sums up the predictions for the order of acquisition, for both a representation-based account as well as a frequency-based account. Section 5 then discusses the data on acquisition, as reported in a variety of studies. Section 6 discusses the findings and compares them to other studies on the relation between representation and lexical frequency on order of acquisition.

## 2. Representation

The representation of tone in terms of features is a topic of some controversy (see e.g. Yip 2002 and Hyman 2011 for issues and proposals). At stake are both the exact set of tone features (e.g. with respect to the number of levels needed), the question if one or more separate register features should be included, and the question if tone features behave like other distinctive features (see e.g. Clements, Michaud & Patin 2011). Tone features are often not included in specific feature geometry proposals (e.g. Sagey 1986), although the behaviour of tone is often adduced as evidence for the autosegmentalisation of distinctive features (Clements & Hume's 1995 seminal paper illustrates both points). One of the main challenges

with respect to tone features is to unify tonal representations for languages with level-tone systems (e.g. many Bantu languages) with representations of tone in languages with contour-tone systems (e.g. many East-Asian languages). The issue is important in the light of the phonological behaviour of tone and their acquisition: If tones have a featural representation, then certain interactions are expected (in the sense of ‘can be formalised’), and we may also derive predictions for their acquisition from their representation. Similar to the acquisition of segmental phonology, in which more complex structures are acquired after less complex ones (e.g. simple syllable structure before complex syllable structure, and simple segments before complex ones, see e.g. Levelt, Schiller & Levelt 1999), we expect that more complex tonal structures are acquired after less complex ones.

What might the phonological representation of the four tones in Figure 1 be? Here we will adopt representations with H(igh) and L(ow), familiar from work on tonology in autosegmental phonology (e.g. Goldsmith 1979). Although register is important in many dialects of Chinese (see e.g. Bao 1990), it does not seem to play a role in the standard Mandarin variety discussed here; so here we will not commit ourselves on the representation of register. Based on these assumptions, T1 is the only level tone, and should be represented as a single H. T2 is a rising tone, representable as a LH contour. It would be possible to represent the tone as MH (since it starts with mid-level 3), but a mid-level tone M seems phonologically superfluous for Mandarin. In other words, based on this view, the realisation of T2 as 35 (in the Chao notation) is the phonetic implementation of a phonological LH contour. T3 (the ‘dipping’ tone) is a concave tone, which would be represented as HLH. Finally, T4, the falling tone, is represented as HL. Again, the fact that L in T2 is a mid-level tone and L in T4 is genuinely low is left to the phonetic implementation. Note also that the first H in T3 (HLH, 214) is in fact *lower* than the L in T2 (LH, 35), consistent with the fact that tone is a relative notion. The proposed representations for T1-T4 are given in (1):



These representations are slightly different from those proposed by Duanmu (2000: 236ff), who represents T3 as single L in non-final full syllables and as trimoraic LH in full syllables (whereas T2 is bimoraic LH), referring also to Yip (1980). At first glance, this seems an odd proposal, because, of course, most syllables in



the rule is needed to exclude T3 from triggering this rule, which makes it slightly more complex, but the assimilation process in itself seems quite natural if the representations in (1) are assumed.

Thus, the tonal representations advanced in (1) offer a possible perspective on complexity and therefore on order of acquisition. We will make our proposal as to the latter explicit in Section 4.1 below. First, we will examine the relative frequency of the four Mandarin lexical tones.

### 3. Frequency of the four tones

Are the four tones equally common in Mandarin? If they are not, the more common ones may be acquired more easily than the less common. It is well known that, in general, acquisition of common structures precedes acquisition of less common structures (see Saffran 2001 and other related work). If a particular segment, or syllable pattern, or phonological phenomenon, occurs often, it is acquired more quickly than a segment (or pattern), which is less frequent. The acquisition of tones in Chinese is expected to be subject to the same tendency, i.e. that frequent tones should be acquired more rapidly (and/or accurately) than infrequent ones.

To find out the frequency of the four tones, we selected the 500 most frequent words in Mandarin from Cai & Brysbaert (2010). This is a corpus of Mainland Chinese subtitles, downloaded from the internet. According to Cai & Brysbaert (2010), “these word frequencies are a good estimate of daily language exposure and capture much of the variance in word processing efficiency”. One may think that it would be better to investigate tonal exposure on the basis of a corpus of child-directed speech, but we assume that the number of tones, like other aspects of phonological structure, is roughly the same for child-directed and adult-directed speech, and the number of tokens that overlap in both types of corpora is likely to be very high as well (see also Sloos & van de Weijer (submitted) for discussion and a test for two French corpora). We ranked the words in this corpus in order of token frequency, and extracted the first, most frequent, 500 words for analysis. Subsequently, we counted the occurrence of each tone in these words (resulting in 759 tones) and multiplied them by the token frequency of the words on which they occurred. For disyllabic and trisyllabic words, all tones were counted and multiplied by the token frequency of the word. For sequences of T3 tones, we took T3 sandhi into account (see the discussion in Section 2) and counted the first tone as T2, which occurred in 11 cases (T2 sandhi did not occur in the data). Since the lexicon grows continuously through acquisition, the relative frequencies of the tones may shift during the acquisition process. We therefore computed the relative frequency of the tones as a function of the growing lexicon (where lexicon

size is determined for each following word in the frequency ranking). The relative frequency of the tones is expressed as a percentage of all tones in the lexicon of a particular size. The results are shown in Figure 2 (including T5).

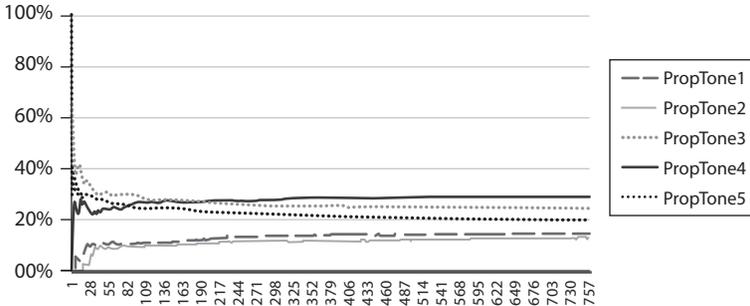


Figure 2. Percentages of Chinese tones per n syllables, from 1 to 500 words

Figure 2 shows that T1 and T2 have similar frequencies throughout the acquisition process: Their initial relative frequency is low, but quickly increases and remains stable at 12–14% (T1) and 10–12% (T2), respectively, from roughly 200 syllables onwards (which corresponds to a lexicon size of about 150 words (see Figure 3)). T3 initially has a high frequency, but rapidly declines and stabilizes at a level of about 25% around 350 syllables (which corresponds to a lexicon size of a little over 300 words). Initially, T4 is not very frequent, but its frequency increases faster than T1 and T2, surpasses T3 and becomes the tone with the highest frequency from roughly 135 syllables onwards, at a level of around 28% (which corresponds to a lexicon size of a little over 100 words).

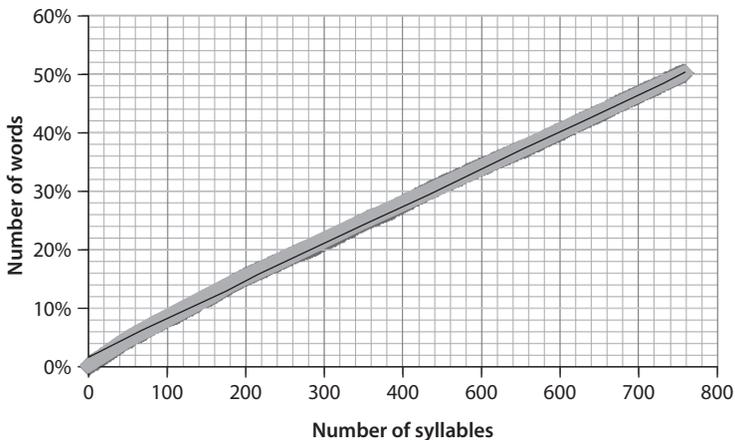


Figure 3. The number of words as a function of the lexicon size

In the next section, we will translate these findings into a specific proposal for the order of acquisition of the four tones.

#### 4. Predictions

In this section, we summarize the predictions made by the representation-based account of acquisition (Subsection 4.1) and the frequency-based account (Subsection 4.2).

##### 4.1 Representation-based predictions

What would the tonal representations in (1) predict for the order of tonal acquisition? T1 (level) is expected to be easiest, having only one tonal feature. Tones 2 and 4 (both contours) might be expected to be intermediate. T3 seems to have the most complex representation and might therefore be expected to be hardest to acquire. These representations accord with the phonetic facts quite well: Level tones are assumed to be articulatorily easier than contour tones, which accounts for the contrast between T1 on the one hand and the other tones on the other hand (Zhang 2001, among others).<sup>1</sup> Tone systems with level tones are typologically also more common (and hence less ‘marked’) than tone systems that involve both level and complex tones (see Map 13 in Haspelmath, Dryer, Gil & Comrie 2005). T3 involves a ‘double contour’ (both falling and rising) and is therefore, also phonetically, expected to be most difficult. Strictly representationally speaking, there is no way of distinguishing between the complexity of the rising (T2) and falling contour (T4), which are both complex tones. Here we must therefore turn to the phonetics to complete our prediction. Phonetically, it has been argued that rising tones are more difficult or require more phonetic effort than falling ones (see Ohala & Ewan 1973; Sagart, Hallé, de Boysson-Bardies & Arabia-Guidet 1986; Zhang 2001), so that we might expect T4 to be acquired more easily than T2. The order of acquisition that is predicted by the phonological representations is relatively straightforward and can be summarized as follows:

- (4) T1 > T4 > T2 > T3 (where ‘>’ means ‘is acquired before’)

These predictions are tentative since there are many factors which are not equal for the four tones, and which may therefore also independently interfere with order of acquisition. For instance, T3 is longer than the other tones and often accompanied by creaky voice (‘murmur’, see Duanmu 2000: 227), so that it may even be more difficult to produce than suggested by its ‘double contour’ representation (and therefore does not change our predictions). T4 (or rather, the vowel on which it is

realised) is much shorter than the other tones (see e.g. Hallé 1994). Thus, acquisition of phonation and the length contrast may also have to be taken into account in a full analysis of the acquisition of tones.

#### 4.2 Frequency-based predictions

Recall that, in general, frequent segments, syllable structures, and other linguistic patterns are acquired before less frequent ones. Figure 2 above shows that there are differences in frequency between the four tones. In the earliest stages of acquisition, up to a lexicon size of just over a 100 words, T3 is the most frequent. T4 is less frequent, followed by T1 and T2. In the second stage, T3 becomes a bit less frequent than T4. The prediction for order of acquisition would therefore be as in (5):

$$(5) \quad T3 > T4 > T1, T2$$

### 5. The facts of acquisition

In this section, we present the findings on tonal acquisition, based on previous work by Zhu (2002), Wong (2012) and references cited there, in order to verify whether the representation-based prediction (4) or the frequency-based prediction (5) accounts better for the facts.

Zhu (2002) makes a useful contrast between the first appearance of tone and stabilization of a tone. The first appearance of T1-T4 is captured by her Table 4.5 and shows that T1 and T4 were earliest for all children. Rising tone (T2) was already present in two children when recordings began and arose later than T1 and T4 for two other children. T3 arose latest for all children. A tone was considered to become stabilized (Zhu's Table 4.6, Zhu 2002: 90) only when its accuracy rating in spontaneous speech reached the 66.7% level and afterward remained (minimally) at that level (Zhu 2002: 89). Based on longitudinal data from four children, T1 was the first to be stabilized, followed by T4. There were variations in the order of stabilization of T2 and T3, differences between which were small. This is consistent with the finding by Li & Thompson (1976), summarized by Yip (2002): "Syllables with [T2 and T3] [...] are either avoided or changed to level or falling. When these last two tones are acquired, at first they are quite often confused, and this confusion continues on into the two- to three-word stage. The confusion takes place mainly in final position, where both have a rising portion" (Yip 2002: 302).

Wong (2012) introduces the concept of 'adult-like' production of tones by first-language learners and finds that children produce T4 most adult-like earliest, followed by (in that order) T1, T2 and T3. It is a little hard to interpret this finding:

It seems to indicate that first-language learners know how to use (something that sounds like) T1 first contrastively, but they know how to produce T4 completely correctly (i.e. adult-like) first.

To conclude, acquisition of the four tones in Mandarin seems to take place roughly as follows:

$$(6) \quad T1 > T4 > T2 > T3$$

## 6. Discussion and conclusion

First, let us repeat the predictions and the actual finding on order of acquisition of Chinese tones:

(7) Phonological representation:	$T1 > T4 > T2 > T3$	(4)
Frequency-based:	$T3 > T4 > T1, T2$	(5)
Actual:	$T1 > T4 > T2 > T3$	(6)

What we see is that the representational account, together with the phonetic insight that falling tones are easier than rising ones, makes exactly the correct predictions. Lexical frequency hardly plays any role, it seems. Recall that frequency often predicts order of acquisition quite well (Saffran 2001, and related work; see also van de Weijer & Sloos 2013), so here a dilemma arises: Why is it apparently irrelevant to the acquisition of Mandarin tone?

We think the answer to this riddle is related to the age at which tone is acquired by Mandarin-speaking children. All authors in the literature comment on the ease and speed with which the tones appear in L1 speakers. E.g. Zhu (2002: 88) notes that some tones have already emerged at the time of her first recordings at age 1;2 and that the other tones appear rather quickly afterward (see her Table 4.5). Li and Thompson's (1976: 185) first observation is that "tone acquisition is accomplished within a relatively short period of time". In Jeng (1979), cited by Zhu (2002: 79), one of the subjects was able to produce tones correctly by the age of 1;7,15 and another by the age of 1;5. In Zhu's own study, all tones are stably acquired by age 1;9,15 (Zhu, Table 4.6, p. 90). What is the lexicon of Mandarin speakers at this age? Zhu does not provide information on this point, but Tardif (1996) deals with the acquisition of nouns and verbs in a group of children of similar age (20–22 months: Tardif 1996: 497) and finds that at that age the number of acquired lexical items ranges between 50 and 115 words (Tardif 1996: Figure 1). This could explain why the statistical patterns of the tones in actual words would only have a very small role to play in the differentiation between acquisition of the different tones: Simply put, too few words have been acquired to make any meaningful generalizations on

the basis of their lexical frequency. In other words, we suggest that for properties that appear relatively ‘early’ in acquisition, word frequency is not expected to play a role, while for features that are acquired later, its role is expected to be larger. This prediction, which is completely in line with models of grammar that combine phonological representation with usage-based factors such as frequency (e.g. van de Weijer 2012; Sloos 2013), could be further investigated in future research.

To conclude, we have proposed representations of the four lexical tones of Mandarin which correspond well to the phonetic facts, and which enable us to express the tone sandhi rules that occur in Mandarin. The tones have different representational complexity, on the basis of which we predicted that T1 would be easiest to acquire, followed by T4, T2 and T3 in that order. This approach turned out to be nearly correct, although T4 appears to be produced most adult-like earliest (Wong 2012). We suggested that the early nature of tonal acquisition is responsible for this: Tones are acquired when only a small part of the lexicon has emerged. We also found that lexical frequency is not (or less) relevant for early acquisition.

## Notes

\* We thank Liu Zenghui for comments after the presentation of this paper at the TIN-dag 2014, as well as the audience of the Shanghai International Studies University Linguistics Salon. We also thank two anonymous LIN reviewers, as well as Luo Mingqiong, Wang Xingwei and Wu Minghui for their helpful comments on earlier drafts. The usual disclaimers apply.

1. As an anonymous reviewer points out, the fact that T1 is least marked is also supported by data from aphasia: This is the only tone that remains after language loss (Liang & van Heuven 2004).

## References

- Bao, Zhiming. 1990. *On the nature of tone*. Cambridge, MA: MIT dissertation.
- Cai, Qing & Marc Brysbaert. 2010. SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *Plos ONE* 5(6), e10729. DOI: 10.1371/journal.pone.0010729
- Chao, Yuen-Ren. 1930. A system of tone-letters. *Le Maître Phonétique* 45. 24–27.
- Chen, Matthew Y. 2000. *Tone sandhi: Patterns across Chinese dialects*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511486364
- Clements, G. Nick & Elisabeth V. Hume. 1995. The internal organization of speech sounds. In John Goldsmith (ed.), *The Handbook of Phonological Theory*, 245–306. Cambridge, MA: Basil Blackwell.
- Clements, G. Nick, Alexis Michaud & Cédric Patin. 2011. Do we need tone features? Ms, CNRS/University of Paris.

- Duanmu, San. 2000. *The phonology of Standard Chinese*. Oxford: Oxford University Press.
- Goldsmith, John A. 1979. *Autosegmental phonology*. New York: Garland.
- Hallé, Pierre. A. 1994. Evidence for tone-specific activity of the sternohyoid muscle in Modern Standard Chinese. *Language and Speech* 37(2). 103–123.
- Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie. 2005. *The world atlas of language structures*. Oxford: Oxford University Press.
- Hyman, Larry M. 2011. The representation of tone. In Marc van Oostendorp, Colin J. Ewen, Elisabeth V. Hume & Keren D. Rice (eds.), *The Blackwell Companion to Phonology*, vol. 2. London: Wiley-Blackwell.
- Jeng, Heng-hsiung. 1979. The acquisition of Chinese phonology in relation to Jakobson's laws of irreversible solidarity. *Proceedings of the 9th International Congress of Phonetic Sciences*. Copenhagen: University of Copenhagen.
- Lee, Wai-Sum & Eric Zee. 2003. Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33(1). 109–112. DOI: 10.1017/S0025100303001208
- Levelt, Clara C., Niels O. Schiller & Willem J. Levelt. 1999. The acquisition of syllable types. *Language Acquisition* 8(3). 237–264. DOI: 10.1207/S15327817LA0803\_2
- Li, C. N. & Sandra A. Thompson. 1976. The acquisition of tone in Mandarin-speaking children. *Journal of Child Language* 4(2). 185–199.
- Li, C. N. & Sandra A. Thompson. 1989. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- Liang, Jie & Vincent J. van Heuven. 2004. Evidence for separate tonal and segmental tiers in the lexical specification of words: A case study of a brain-damaged Chinese speaker. *Brain and Language* 91(3). 282–293. DOI: 10.1016/j.bandl.2004.03.006
- Ohala, John J. & William G. Ewan. 1973. Speed of pitch change. *Journal of the Acoustical Society of America* 53(1). 345. DOI: 10.1121/1.1982441
- Saffran, Jenny R. 2001. Words in a sea of sounds: The output of statistical learning. *Cognition* 81. 149–169. DOI: 10.1016/S0010-0277(01)00132-9
- Sagart, Laurent, Pierre Hallé, Bénédicte de Boysson-Bardies & Catherine Arabia-Guidet. 1986. Tone production in Modern Standard Chinese: An electromyographic investigation. *Cahiers de Linguistique — Asie Orientale* 15(2). 205–221. DOI: 10.3406/clao.1986.1204
- Sagey, Elisabeth C. 1986. *The representation of features and relations in non-linear phonology*. Cambridge, MA: MIT dissertation.
- Sloos, Marjoleine. 2013. *Phonological grammar and frequency: An integrated approach. Evidence from German, Indonesian, and Japanese*. PhD dissertation, University of Groningen.
- Sloos, Marjoleine & Jeroen M. van de Weijer. Submitted. What adult speech can tell us about child language acquisition.
- Tardif, Twila. 1996. Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology* 32(3). 492–504. DOI: 10.1037/0012-1649.32.3.492
- van de Weijer, Jeroen M. 2012. *Grammar as selection: Combining Optimality Theory and Exemplar Theory*. Nagoya: Kougaku Shuppan.
- van de Weijer, Jeroen M. & Marjoleine Sloos. 2013. Learning markedness constraints: The case of French. In Suzanne Aalberse & Anita Auer (eds.), *Linguistics in the Netherlands 2013*, 188–200. Amsterdam and Philadelphia: John Benjamins. DOI: 10.1075/avt.30.14van
- Wang, Jialing. 1997. The representation of the neutral tone in Chinese putonghua. In Jialing Wang & Norval Smith (eds.), *Studies in Chinese phonology*, 157–183. Berlin and New York: Mouton de Gruyter.

- Wong, Pui-san. 2012. Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics* 40(1). 141–151.  
DOI: 10.1016/j.wocn.2011.10.005
- Xu, Yi. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25(1). 61–84.  
DOI: 10.1006/jpho.1996.0034
- Yip, Moira. 1980. *The tonal phonology of Chinese*. Cambridge, MA: MIT dissertation.
- Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.  
DOI: 10.1017/CBO9781139164559
- Zhang, Jie. 2001. *The effects of duration and sonority on contour tone distribution; Typological survey and formal analysis*. Los Angeles, CA: UCLA dissertation.
- Zhu, Hua. 2002. *Phonological development in specific contexts: Studies of Chinese-speaking children*. Clevedon: Multilingual Matters.

### *Authors' addresses*

Jeroen van de Weijer  
School of English Studies  
Shanghai International Studies University  
Building 1, Office 320  
550 Dalian Xi Lu  
Shanghai 200083, China  
jeroen.van.de.weijer@gmail.com

Marjoleine Sloos  
Interacting Minds Centre  
Aarhus University  
Jens Chr. Skous Vej 4, Building 1483, 3rd floor  
8000 Aarhus C, Denmark  
marj.sloos@gmail.com