Chinese (Mandarin), Phonology of

(For *Encyclopedia of Language and Linguistics*, 2nd edition, Elsevier Publishing House)
San Duanmu, University of Michigan, MI USA
February 2005

Chinese is the first language of over 1,000 million speakers. There are several dialect families of Chinese (each in turn consisting of many dialects), which are often mutually unintelligible. However, there are systematic correspondences among the dialects and it is easy for speakers of one dialect to pick up another dialect rather quickly. The largest dialect family is the Northern family (also called the Mandarin family), which consists of over 70% of all Chinese speakers. Standard Chinese (also called Mandarin Chinese) is a member of the Northern family; it is based on the pronunciation of the Beijing dialect. There are, therefore, two meanings of Mandarin Chinese, one referring to the Northern dialect family and one referring to the standard dialect. To avoid the ambiguity, I shall use Standard Chinese (SC) for the latter meaning. SC is spoken by most of those whose first tongue is another dialect. In principle, over 1,000 million people speak SC, but in fact less than 1% of them do so without some accent. This is because even Beijing natives do not all speak SC.

SC has five vowels, shown in **Table 1** in IPA symbols (Chao 1968, Cheng 1973, Lin 1989, Duanmu 2002). [y] is a front rounded vowel. When the high vowels occur before another vowel, they behave as glides [j, ɥ, w]. [i] and [u] can also follow a non-high vowel to form a diphthong. The mid vowel can change frontness and rounding depending on the environment. The low vowel can change frontness but not rounding.

Table 1 Standard Chinese Vowels

High	i	y	u
Mid	ə		
Low	a		

Table 2 Standard Chinese Consonants

Labial	Dental	Palatal	Retroflex	Velar
p	t			k
p^h	t^h			k^h
	ts	(tɕ)	tʂ	
	ts^h	($tɕ^h$)	$tʂ^h$	
f	s	(ɕ)	ʂ	x
m	n			(ŋ)
	l		r	

The consonants of SC are listed in **Table 2**, where sounds with limited distribution are in parentheses. For most speakers [ŋ] cannot occur in syllable initial position. In syllable coda

position only [n] and [ŋ] can occur. The palatals do not contrast with [ts, ts^h, s], [ʈʂ, $ʈʂ^h$, ʂ], or [k, k^h, x]; the palatals only occur with front vowels or front glides, but the other three sets do not. Since some speakers pronounce the palatals as [ts^j, ts^{jh}, s^j], it is possible to analyze a palatal as a combination of a dental and a front vowel. The retroflex liquid (sometimes written as a fricative [ʐ̩]) is not a trill but an approximant (and has no lip rounding, unlike that in English); since SC does not have a trill, I transcribe it as [ɹ] instead of [ɻ]. The retroflex series [ʈʂ, $ʈʂ^h$, ʂ ɹ] is a major characteristic of SC speakers from Beijing. SC speakers from other areas often replace [ʈʂ, $ʈʂ^h$, ʂ, r] with the dentals [ts, ts^h, s, z]. The unaspirated stops and affricates [p, t, k, ʈʂ, ts, tɕ] can become voiced [b, d, g, ɖʐ, dz, dʑ] when they occur in an unstressed syllable.

SC also has two syllabic consonants [z] and [r] (or [ʐ̩]). They were previously thought to be special vowels, probably because it was believed that every syllable must have a vowel. [z] is used when a syllable starts with [ts], [ts^h], or [s] and when there is no vowel in the rime; it can be seen as the extension of the [s] element into the rime, where it becomes voiced. [r] is used when a syllable starts with [ʈʂ], [$ʈʂ^h$], [ʂ], or [r] and when there is no vowel in the rime; it can be seen as the extension of the retroflex element into the rime. SC also has a couple of syllabic nasals, which are usually interjections.

The SC syllable can made of up to four sounds—CGVX, where C is a consonant, G a glide, V a vowel, X a nasal or an offglide of a diphthong, and VX the rime. When both C and G are present, they are realized as one sound C^G, where G is the secondary articulation. Thus, the SC word [s^wei] ‘age’ is phonetically quite different from the English word [swei] *sway*. SC also has a suffix [r], which will change the rime of the syllable it is attached to from VX to Vr and add a retroflex quality to the nuclear vowel. In other words, the [r] suffix can lead to loss of contrast in the original coda.

SC has two kinds of syllables, which can be called full (or regular) syllables and weak syllables. Full syllables (mostly monosyllabic content words) have tone and are long. Weak syllables (mostly grammatical words) are short and do not have their own tones. A full syllable can sometimes change to a weak syllable, in which case it will lose its underlying tone and undergo rime reduction and shortening. In syllable theory, full syllables are heavy and have two moras each, whereas weak syllables are light and have one mora each. In other words, in weak syllables the vowel is short. In full syllables the vowel is short when the rime is VC or VG and long when the rime is V.

In the electronic dictionary CMUDICT, English has around 10,000 monosyllables (excluding homophones). In contrast, SC has a very small syllable inventory, just around 400 syllables excluding tones (or around 1,300 syllables including tones). It is a puzzle why SC uses so few syllables, especially when many times more seem to be available. For example, given about 20 Cs, three Gs, five Vs, and five Xs, there are about 2000 possible CGVX combinations (excluding tones), yet just 400 are used. It turns out that two thirds of the unused forms are ruled out by two requirements. The first is that C and G cannot have the same place of articulation, which follows from the analysis that C^G is a single sound, because in a single sound each place feature can be used just once. The second requirement is that V and X cannot have opposite values for [round] or [back].

There are four distinctive tones on full SC syllables. Weak syllables may get tone from certain intonational environment; otherwise they remain toneless, which is phonetically a low pitch. The four distinctive tones are high, rise, low, and fall. The pitch range of the tones can vary according to stress, whereby a syllable with greater stress has a wider pitch range (Shen 1985). Using standard tonal features, according to which contour tones are made of two (or more) level tones, the four tones are represented in (1), where the vowel is long because it is in a V rime.

(1)

Tone name	First	Second	Third	Fourth
Tone feature	H	LH	L	HL
	/\	| |	/\	| |
Syllable	maa	maa	maa	maa
Gloss	'mother'	'hemp'	'horse'	'scold'

In the alphabet system Pinyin, the words in (1) can be transcribed as *ma1*, *ma2*, *ma3*, and *ma4*, where vowel length is not represented and tones are represented by the digits 1-4. The first, second, and fourth tones have normal voice quality, but the third tone has a murmured voice quality. The third tone may also begin with a slight dip in pitch, which does not seem to be phonologically relevant. In final position the third tone can optionally end with a rise, in which case it is phonetically extra long. Phonologically we may represent it with three moras, exemplified in (2). Such a syllable often has an amplitude break or a glottal stop in the middle, and to some people it like two syllables.

(2)

```
  L H
  /\ |
maaa
```

'horse'

Of the 1,300 or so SC syllables (including tones), most are full syllables, where the four tones are fairly evenly distributed, as shown in **Table 3** and **Table 4**. In **Table 3** we see that there are slightly fewer second tones than other tones, but not by a lot. In **Table 4** we see that most syllables have four or three tones each, and only a small number of syllables have two or one tone each.

Table 3 Frequency of Tones

Tone	*First*	*Second*	*Third*	*Fourth*	*All*
Number	337	255	316	347	1255

Table 4 Frequency of Tones

Tones per syllable	*4*	*3*	*2*	*1*	*All*
Number of such syllables	178	130	59	35	402

According to a text corpus of over 45 million Chinese character tokens (Da, 2000), there are over 6,000 different Chinese characters, most of which are monosyllabic words. This means

that each SC syllable represents about 15 words excluding tones, or 5 words including tones. The homophone load is not distributed evenly, as **Figure 1** shows.

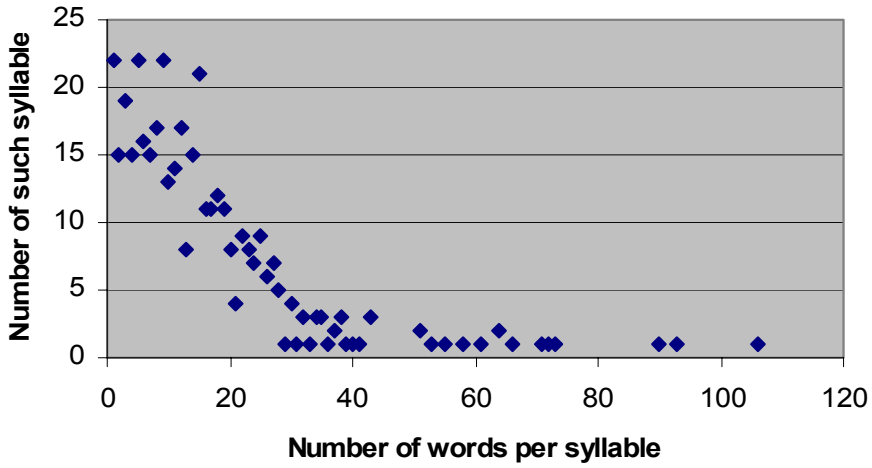

Figure 1: Homophone density in Standard Chinese (ignoring tones), based on the analysis of 6000 characters listed in Da (2000). Most syllables represent fewer than 20 words each, but the syllable [ji] represents over 100 words.

The top fifteen SC syllables are shown in (3), where the number of words a syllable represents (ignoring tones) is shown in parentheses. One might think that the most frequent syllables are the most 'natural' or 'unmarked', namely, those that children learn first or those that are most common in the world's languages, such as [ba], [ma], or [ta]. However, many of those in (3) do not seem to be unmarked syllables.

(3) pi (51), tɕy (51), wei (53), wu (55), ʂɻ̩ (58), tɕan (61), ɕi (64), ɥan (64), tɕʰi (66), li (71), ʈʂɻ̩ (72), fu (73), ɥy (90), tɕi (93), ji (106)

Most English monosyllables represent just one word each. Since Chinese has so many homophones, a natural question is, how does Chinese avoid ambiguity in speech? The answer seems to be that most ambiguities are clarified by context. For example, although *sun* and *son* are homophones in English, there is hardly any context in which they would cause ambiguity.

Despite the large number of homophones, the syllable inventory of Chinese continues to decrease. SC no longer allows [p t k] or [m] in syllable final position, although some other dialects do. Shanghai has lost all diphthongs, and its tonal inventory has reduced to just two. In all likelihood, SC is moving in the direction of further reduction. For example, SC does not make use of such contrasts as [wi] vs. [wei], or [ji] vs. [i], which English does (consider *we* vs. *way*, and *yeast* vs. *east*). In addition, about 200 of the 1300 syllables are now rarely used. From a

functional point of view, it is a mystery why high homophone density has not prevented syllable loss in SC or at least slowed it down. A possible answer, paradoxically, is that high homophone density may in fact speed up syllable loss. Studies on frequency effects show that frequent words are more likely to undergo reduction than infrequent words (Bybee 2001). Because Chinese has fewer syllables than English, Chinese syllables are used more frequently, and so they are more likely to undergo reduction and loss of contrasts.

In disyllabic English words and phrases, three kinds of stress differences can be distinguished. In words like *Peter*, *Anna*, and *panda*, one is clearly stressed and the other not. In such cases, the stressed syllable is longer, has an unreduced vowel, and has a pitch accent. In contrast, the unstressed syllable is short, has a reduced vowel, and has no pitch accent. In words like *blackboard* and *pancake*, the stress difference is also clear. Although both syllables are heavy and have an unreduced vowel, one syllable has a pitch accent and the other does not. In expressions like *Red Cross*, *real deal*, and *red-hot* (adjective), the stress difference is no longer obvious; in each case, both syllables are heavy, have an unreduced vowel, and have a pitch accent. As a result, they are sometimes thought to have equally stress.

When a full SC syllable occurs next to a weak one, their stress difference is like the first English case (as in *Peter*, *Anna*, and *panda*). When two (or more) full SC syllables occur together, they all have tones, and so their stress difference is not obvious; this is similar to the third English case (as in *Red Cross* and *real deal*). Now in English the first two cases are quite common, and so stressed syllables often stand out. In Chinese, on the other hand, full syllables often occur together, and so stressed syllables often do not stand out. This may have contributed to a common view that there is no stress in Chinese.

The difference between full and weak syllables in SC can be explained in terms of moraic trochee: a full syllable has two moras, so it forms a foot and has stress. However, Chinese also uses disyllabic feet. For example, the disyllabic foot is used in poetic templates, and it is also a domain for certain kinds of tone sandhi. In addition, a minimal expression should be disyllabic. If a noun is monosyllabic, a semantically redundant syllable is often added. Thus, one usually cannot say *Fa* 'France' or *Wang* 'Wang' but must say *Fa Guo* 'France Country', *Lao Wang* 'old Wang', or *Xiao Wang* 'Little Wang'. In contrast, *Sudan* 'Sudan' and *Yindu* 'India' can be said by themselves (and adding 'country' to them would be odd). The disyllabic requirement has created a large dual vocabulary, whereby many words have two forms, a monosyllabic form and a disyllabic form. The disyllabic form is a compound in structure but a single noun in meaning and it can be called a pseudo-compound. Some examples are shown in **Table 5**.

Table 5 Pseudo-Compounds

Pseudo-compound	Literal meaning	Gloss
mei-tan	coal-charcoal	'coal'
shi-jian	time-interval	'time'
shang-dian	business-store	'store'
jin-qian	gold-money	'money'
lao-hu	old-tiger	'tiger'

It is commonly thought that the creation of disyllabic words is triggered by homophone density. However, the common view cannot explain why monosyllabic names need another

syllable, even though there is no ambiguity, such as when you address someone in person. A more likely reason for the creation of pseudo-compounds is to fill a disyllabic foot (Duanmu 1999b).

A disyllabic noun (or compound) can be heavy-heavy (two full syllables), or heavy-light (a full syllable and a weak syllable), but not light-heavy. This suggests that the disyllabic foot is trochaic. However, when a heavy-heavy noun is spoken in isolation, the second syllable is often longer and appears to have slightly more stress. This has lead to the view that Chinese has final stress (Chao 1968, Hoa 1983). But when a heavy-heavy noun is in nonfinal position, its second syllable no longer has extra duration (Feng 1985; Wang and Wang 1993). It seems therefore that the extra duration of a final full syllable is due to pre-pause lengthening, and the trochaic analysis fares better overall.

The discussion so far suggests that a disyllabic word contains both moraic trochee and syllabic trochee (Duanmu 1999a). The structure is shown in (4), which can be called a dual-trochee.

(4)

	heavy-light	heavy-heavy
	x	x
Syllabic trochee:	(σ σ)	(σ σ)
	/\ |	/\ /\
Moraic trochee:	(mm) . m	(mm) . (mm)
	x	x x

The dual-trochee distinguishes three degrees of stress: (a) a heavy syllable that heads a syllabic foot, (b) a heavy syllable that does not head a syllabic foot, and (c) a light syllable. The cases are easy to distinguish in English: (a) has an unreduced vowel and a pitch accent (first syllable in *Peter*, *pancake*, or even *city*, if the last word is syllabified as *cit.y*), (b) has an unreduced vowel but no pitch accent (second syllable in *pancake*), and (c) has a reduced vowel and no pitch accent (second syllable in *Peter* or *Anna*). In Chinese, (a) and (b) are hard to distinguish because they both have tones and unreduced rimes; however, (c) is easy to distinguish from (a) and (b) because it has a reduced rime and no tone.

Since Chinese uses pitch contour (tones) to contrast word meanings, intonation is often expressed not as pitch variation on lexical words themselves, but as boundary tones that are added after lexical tones. Two examples are shown in (10) and (11).

(5)

Tone		Intonation		
LH	+	L	→	LHL
nan				nan
'difficult'		'affirmation'		'Surely difficult!'

(6)

Tone		Intonation		
HL	+	H	→	HLH
mai				mai
'sell'		'question'		'Sell?'

The boundary tones can also occur on what might be called intonation syllables. For example, the boundary tone in (10) can occur on [a] or [ou], and that in (11) can occur on [ma].

Many Chinese dialects have tone sandhi, whereby syllable tones change in context. The most well-known tone sandhi in SC is the Third Tone Sandhi, by which a third tone changes to a second tone when another third tone follows, or T3 T3 → T2 T3. In an expression made of many third tones, the resulting change can be quite complicated. To understand the change one must first understand the formation of syllabic feet, which in turn depends on syntax (Shih 1986, Shen 1994, Chen 2000, Duanmu 2002). Thus, Third Tone Sandhi offers an excellent case for the study of the interaction between phonology and syntax.

In summary, Chinese differs from languages like English in a number of ways (such as a lack of polysyllabic words, a small inventory of syllables, high homophone density, a dual vocabulary, and the use of distinctive tones). However, the difference is only apparent. Under careful analysis, Chinese also observes similar linguistic principles as other languages do (such as foot structure, the behavior of heavy vs. light syllables, and the effect of frequency on syllable reduction).

Bibliography

Bybee, J. (2001). *Cambridge Studies in Linguistics 94: Phonology and language use*. Cambridge, UK: Cambridge University Press.

Chao, Y. R. (1930). 'A system of tone letters' *Le maître phonétique* 45, 24-27.

Chao, Y. R. (1933). 'Tone and intonation in Chinese' *Bulletin of the Institute of History and Philology*, Academia Sinica 4.2, 121-134.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Chen, M. (2000). *Cambridge Studies in Linguistics 92: Tone sandhi: patterns across Chinese dialects*. Cambridge, UK: Cambridge University Press.

Cheng, C. C. (1973). *A synchronic phonology of Mandarin Chinese*. (Monographs on linguistic analysis no. 4). The Hague: Mouton.

Da, J. (2000). 'Chinese text computing' http://lingua.mtsu.edu/chinese-computing/. Murfreesboro, TN: Department of Foreign Languages and Literatures, Middle Tennessee State University.

Duanmu, S. (1999a). 'Metrical structure and tone: evidence from Mandarin and Shanghai' *Journal of East Asian Linguistics* 8.1, 1-38.

Duanmu, S. (1999b). 'Stress and the development of disyllabic vocabulary in Chinese' *Diachronica* XVI.1, 1-35.

Duanmu, S. (2002). *The phonology of Standard Chinese* (Paperback edition.) Oxford: Oxford University Press.

Duanmu, S. (2004). 'A corpus study of Chinese regulated verse: phrasal stress and the analysis of variability' *Phonology* 21: 43-89.

Feng, L. (1985). 'Beijinghua yuliu zhong shengyundiao de shichang' [Duration of initials, finals, and tones in Beijing dialect]. In Lin, T. & Wang, L. J. (eds.), 131-195.

Hoa, M. 1983. *L'accentuation en pékinois*. Paris: Editions Langages Croisés. Distributed by Centre de Recherches Linguistiques sur l'Asie Orientale, Paris.

Lin, M. C. & Yan, J. Z. (1988). 'The characteristic features of the final reduction in the neutral-tone syllable of Beijing Mandarin' *Phonetic Laboratory Annual Report of Phonetic Research*, 37-51. Beijing: Phonetic Laboratory, Institute of Linguistics, Chinese Academy of Social Sciences.

Lin, T. & Shen, J. (1995). 'Beijinghua er hua yun de yuyin fenqi' [Variations in the [er]-suffixed rimes in the Beijing dialect] *Zhongguo Yuwen* 1995.3, 170-179.

Lin, T. & Wang, L. J. (eds.) (1985). *Beijing Yuyin Shiyanlu* [Working Papers in Experimental Phonetics]. Beijing: Beijing University Press.

Lin, Y. H. (1989). *Autosegmental treatment of segmental processes in Chinese phonology*. Doctoral dissertation, University of Texas, Austin.

Shen, J. (1985). 'Beijinghua shengdiao de yinyu he yudiao' [Pitch range of tone and intonation in Beijing dialect]. In Lin, T. & Wang, L. J. (eds.), 73-130.

Shen, J. (1994). 'Beijinghua shangsheng liandu de diaoxing zuhe he jiezou xingshi' [Tonal patterns and rhythmic structure in successive third tones in the Beijing dialect] *Zhongguo Yuwen* 1994.4, 274-281.

Shih, C. L. (1986). *The prosodic domain of tone sandhi in Chinese*. Doctoral dissertation, University of California, San Diego.

Wang, J. & Wang, L. J. (1993). 'Putonghua duo yinjie ci yinjie shi chang fenbu moshi' [The types of relative lengths of syllables in polysyllabic words in Putonghua] *Zhongguo Yuwen* 1993.2 (233), 112-116.

Wang, L. J. & He, N. J. (1985). 'Beijinghua er-huayun de tingbian shiyan he shengxue fenxi' [Auditory discrimination experiments and acoustic analysis of Mandarin retroflex endings]. In Lin, T. & Wang, L. J. (eds.), 27-72.

Yan, J. Z. & Lin, M. C. (1988). 'Beijinghua sanzizu zhongyin de shengxue biaoxian' [Acoustic characteristics of the stress in Beijing trisyllables] *Fangyan* 1988.3, 227-237.

Yang, S. A. (1992). 'Beijinghua duoyinjie zuhe yunlu tezheng de shiyan yanjiu' [An experiment on the prosody of polysyllables in the Beijing dialect] *Fangyan* 1992.2, 128-137.

Yip, M. (1980). *Tonal phonology of Chinese*. Doctoral dissertation, MIT, Cambridge, Mass.

Yip, M. 2002. *Tone*. Cambridge, UK: Cambridge University Press.