

Modeling the Sound Pattern of Standard Chinese

Theory Group

Computer Science, Earlham College

Stress Languages – Rhythm

In English, as in other stressed languages, the sound pattern of words can be thought of as a kind of rhythm: words are broken into syllables, some of which get more emphasis than the others. Consider the words “phonology” and “phonological”.

- PRI - SEC - PRI - -
pho NO lo gy PHO no LO gi cal
pho 'no lo gy 'pho no 'lo gi cal

In “phonology”, the primary stress (PRI) falls on the secondary syllable, but in “phonological”, the primary stress goes to the third syllable and a secondary stress (SEC) is placed on the first syllable. Placing the stresses anywhere else, on the other hand, would make it sound funny. Previously, our group has developed computational models for a library of stress patterns that includes most of the world’s stressed languages.

Tonal Languages – Melody

Standard Chinese (SC), in contrast, is a tonal language: the sound pattern can be thought of as a melody with syllables pronounced with varying pitch. A prime example of tones in SC is “hàn yǔ”. The first syllable “han”has the fourth tone while the second “yu”has the third tone, and with these two specific tones the word means ‘Chinese’. If we were to change the fourth tone on “han” to the second, resulting in “hán”, the meaning of the word would change to ‘Korean’.

4 3 2 3
hàn yǔ hán yǔ
‘Chinese’ ‘Korean’

The Differences

A main difference between English and Chinese is that in English a light syllable can occur in any position of a word. For example, in banana, a light syllable occurs in the first and third positions, and in America, a light syllable occurs in the first, third, and fourth positions. In contrast, a light syllable in Chinese does not occur word initially. As a result, there are fewer word-stress patterns in Chinese than in English. The reason Chinese does not have unstressed initial syllables seems to be that in non-final positions all disyllabic words have initial stress, even though some such words have final stress in isolation. For example, in SC xīn-zàng ‘heart’ has final stress in isolation but initial stress in xīn-zàng bìng ‘heart disease’.

1 4 1 4 4
xīn zàng xīn zàng bìng
‘heart’ ‘heart disease’

The Similarities

In *Phonology of Standard Chinese*, San Duanmu cites previous works that argued that the realization of stress could vary in different languages. As an example, the linguists (e.g. Beckman 1986; Hayes 1995) suggested that English could be realized phonetically while Chinese abstractly, which makes it hard to perceive. Consequently, to a native SC speaker, it is not immediately obvious that there is even a clear distinction between syllables and words, let alone distinctions in stress between syllables. However, San concludes that SC does, in fact, exhibit syllabic stress and that the stress pattern provides a clear distinction between words and syllables and explains, in part, the distribution of tones. Cases exist where a heavy syllable is followed by a light syllable, and when this occurs it is easy to tell that the heavy syllable has stress and that the light syllable does not. In fact, some minimally contrastive pairs are available in SC. In the example to the right, the stress pattern is heavy-light for ‘careless’ and heavy-heavy for ‘main idea’.

4 0 4 4
dà yì dà yì
‘careless’ ‘main idea’

The Principles

In his book, San Duanmu’s Tone-Stress Principle states that:

- 1 A stressed syllable can be assigned a lexical tone or pitch accent.
- 2 An unstressed syllable is not assigned a lexical tone or pitch accent.

San further illustrates how to form feet and uses the concept to develop his allowed list of stress patterns. First, a foot is usually formed from the left; secondly, the final syllable usually forms another foot (with the empty beat); thirdly, binary feet are formed for other syllables, with some flexibility, as long as there is no stress clash (two adjacent stresses) or two free syllables in succession. Ignoring the distinction between 2 and 3, which is rather subtle and often optional, the foot structures of the patterns are shown in the table below. The symbols are explained on the right.

Length	Foot Structure	
2 syllables	H ⁺ H ⁺	
3 syllables	H ⁺ H ⁺ H ⁺	Foot iamb (hH)
4 syllables	H ⁺ H ⁺ H ⁺ H ⁺	trochee (Hh)
5 syllables	H ⁺ H ⁺ H ⁺ H ⁺	H unstressed heavy syllable
6 syllables	H ⁺ H ⁺ H ⁺ H ⁺ H ⁺	H ⁺ secondary stress heavy
	H ⁺ H ⁺ H ⁺ H ⁺ H ⁺	H ⁺ primary stress heavy
7 syllables	H ⁺ H ⁺ H ⁺ H ⁺ H ⁺ H ⁺	

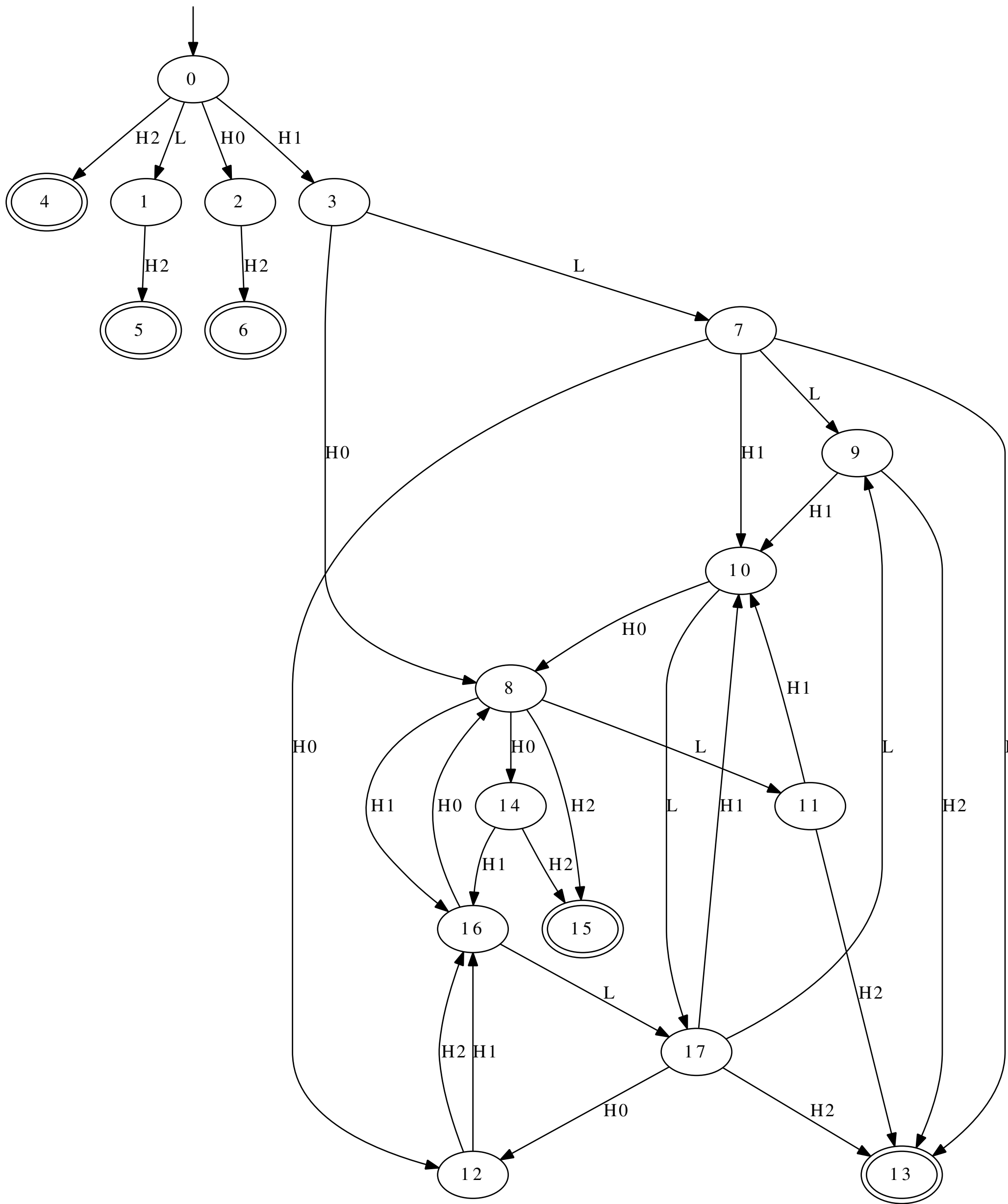
Based on the table of stress patterns, we came up with a list of forbidden patterns for SC, i.e.: L⁺, ×H×, ×H×, $\acute{\sigma}^* \acute{\sigma}^*, \acute{\sigma}^+ \acute{\sigma}^+, \sigma\sigma\sigma, \times\sigma\acute{\sigma}^*$. We used the same form of notation in our previous work and also in the workbench Dakotah Lambert (Earlham ’15) developed. This table explains the symbols:

Symbol	Meaning
×	start of word
×	end of word
*	any syllable
$\acute{\sigma}^+$	syllable with any stress
σ	unstressed syllable
$\acute{\sigma}$	syllable with primary stress
$\grave{\sigma}$	syllable with secondary stress

Here we can trasnlate the list of forbidden patterns for SC into English:

- 1 L⁺ - no stressed light syllable
- 2 ×H× - no unstressed heavy monosyllable
- 3 ×H× - no heavy monosyllable with secondary stress
- 4 $\acute{\sigma}^* \acute{\sigma}^*$ - no nonfinal primary stress
- 5 $\acute{\sigma}^+ \acute{\sigma}^+$ - no two consecutive stressed syllables
- 6 $\sigma\sigma\sigma$ - no three consecutive unstressed syllables
- 7 $\times\sigma\acute{\sigma}^*$ - no starting unstressed syllable followed by two or more syllables

The DFA



We drew the DFA based on the list of constraints. The result is a DFA with 17 states with 5 accept states that yield acceptable strings of varied lengths in SC. Below are some examples of strings we ran through the DFA.

Word	Path	Accepted?	Violated Constraint
H	0 \xrightarrow{H} 2	No	×H×
H ⁺ H ⁺	0 \xrightarrow{H} 2 \xrightarrow{H} 6	Yes	
H ⁺ H	0 \xrightarrow{H} 4 \xrightarrow{H} ?	No	$\acute{\sigma}^* \acute{\sigma}^*$
H ⁺ H ⁺	0 \xrightarrow{H} 3 \xrightarrow{H} ?	No	$\acute{\sigma}^+ \acute{\sigma}^+$
H ⁺ H ⁺ H ⁺ H ⁺	0 \xrightarrow{H} 3 \xrightarrow{H} 8 \xrightarrow{H} 14 \xrightarrow{H} ?	No	$\sigma\sigma\sigma$
H ⁺ H ⁺ H ⁺ H ⁺	0 \xrightarrow{H} 3 \xrightarrow{H} 8 \xrightarrow{H} 14 \xrightarrow{H} 15	Yes	
H ⁺ H ⁺ H ⁺ H ⁺ H ⁺	0 \xrightarrow{H} 2 \xrightarrow{H} ?	No	$\times\sigma \acute{\sigma}^*$
H ⁺ H ⁺ H ⁺ H ⁺ H ⁺	0 \xrightarrow{H} 3 \xrightarrow{H} 8 \xrightarrow{H} 16 \xrightarrow{H} 8 \xrightarrow{H} 14 \xrightarrow{H} 15	Yes	

References

[1] San Duanmu.
The Phonology of Standard Chinese, Second Edition.
Oxford University Press, 2nd edition, 2007.