

THE SIZE OF MORPHEMES IN MANDARIN: PERSPECTIVES FROM TONAL UR LEARNING

Boer Fu

Massachusetts Institute of Technology



Research Question: Syllable = Morpheme?

- Conventional view of Mandarin Chinese:
 - Every syllable is a meaning-bearing unit, or a morpheme.
 - A disyllabic word is composed of two morphemes.
- Loanwords are obvious exceptions (Packard 2000).
 - ʃweɿ4 tʃɛn3* 'Sweden': single disyllabic morpheme.
- I argue that disyllabic morphemes can be found in the native lexicon as well.
- Using evidence from tonal underlying representation (UR) learning.

Why Tonal UR Learning?

- Learning UR-SR mapping requires morphophonological alternations.
- Only when a Mandarin syllable is also a morpheme, can it be used to establish morphophonological alternations.
- By investigating the tonal UR Mandarin speakers have acquired during phonological learning,
- We can identify whether syllables are learned as morphemes or not.

Roadmap

- Semantic opacity in Mandarin compounds
- Tone 3 sandhi neutralization
- Novel AABB reduplication diagnostic
- Speaker judgement survey results
- Implication on morpheme size
- Implication on Chinese orthography

Mandarin Compounds: Musical Instrument names



ǎjaw3 xaw4
小 号
little horn
'trumpet'



tʃʰɑŋ2 xaw4
长 号
long horn
'trombone'



ta4 xaw4
大 号
big horn
'tuba'



ʏɛn2 xaw4
圆 号
round horn
'French horn'

semantically transparent

More Musical Instruments



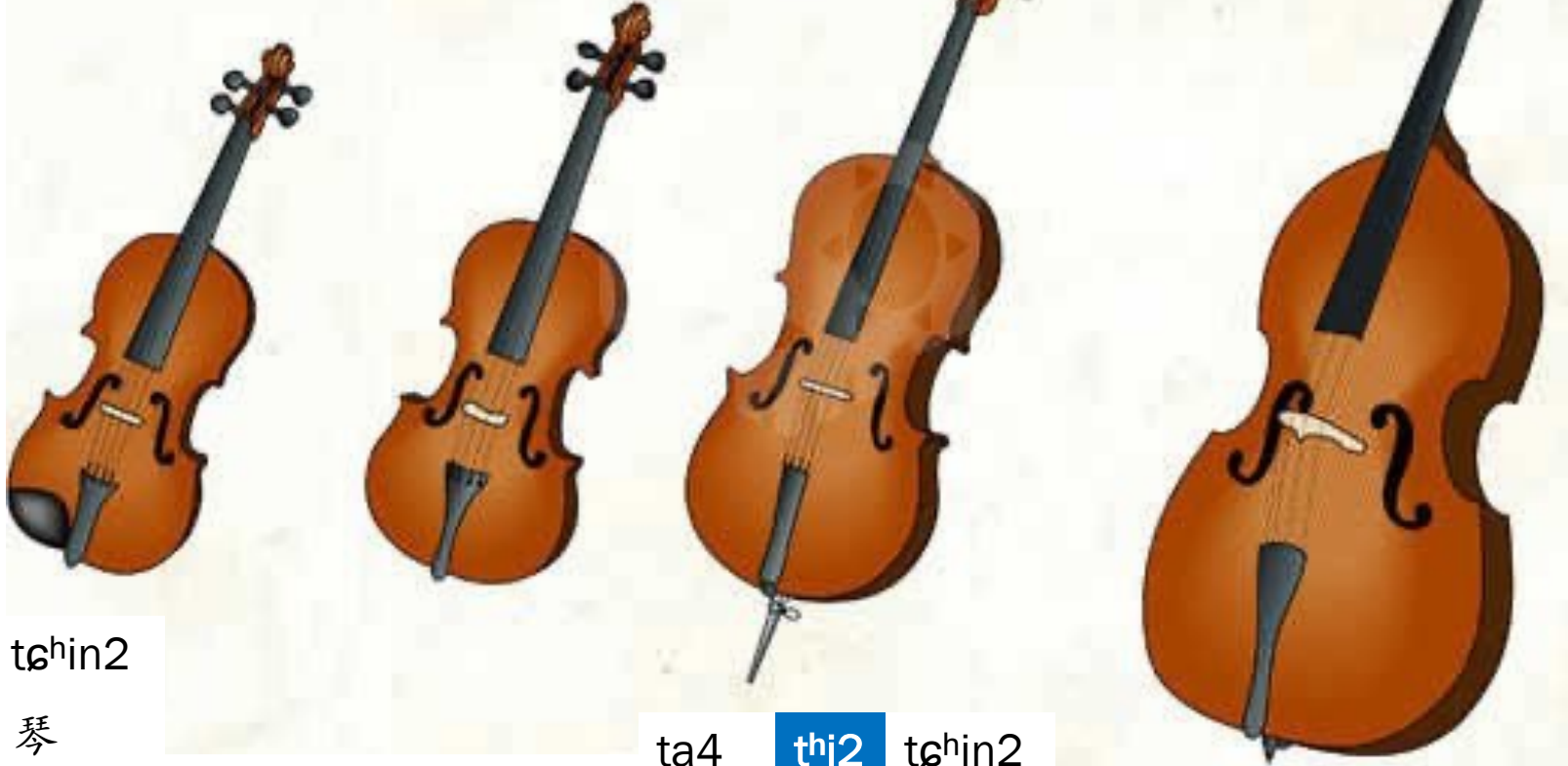
ɕjaw3 thi2 tɕhin2
 小 提 琴
 small lift qin
 'violin'

tʂuŋ1 thi2 tɕhin2
 中 提 琴
 medium lift qin
 'viola'

ta4 thi2 tɕhin2
 大 提 琴
 big lift qin
 'cello'

ti1 jin1 thi2 tɕhin2
 低 音 提 琴
 low voice lift qin
 'double bass'

Semantic Opacity!



ɕjaw3 thi2 tɕhin2
 小 提 琴
 small lift qin
 'violin'

tɕsun1 thi2 tɕhin2
 中 提 琴
 medium lift qin
 'viola'

ta4 thi2 tɕhin2
 大 提 琴
 big lift qin
 'cello'

ti1 jin1 thi2 tɕhin2
 低 音 提 琴
 low voice lift qin
 'double bass'

Is *thi2* A Morpheme?

- On its own, *thi2* is a morpheme, meaning ‘to lift’.
 - It can combine with aspect markers like -zhe and -le.
- But in ‘cello’ and ‘double bass’, *thi2* is not a morpheme.

(1)

a. ta4 [thi2 tɕhin2]

lift

big string instrument

‘cello’

b. ti1 jin1 [thi2 tɕhin2]

~~lift~~

low voice string instrument

‘double bass’

Semantic opacity in Compounds

- *thi*² is semantically opaque in *thi*² *tʂhin*² ‘string instrument’.
- Lexicalization: individual word components have lost their meaning.
- ❖ Question: how can we tell that lexicalization has taken place?
- Lexicalization can be observed in tonal UR learning of tone 3 sandhi words.

Mandarin Tone 3 Sandhi

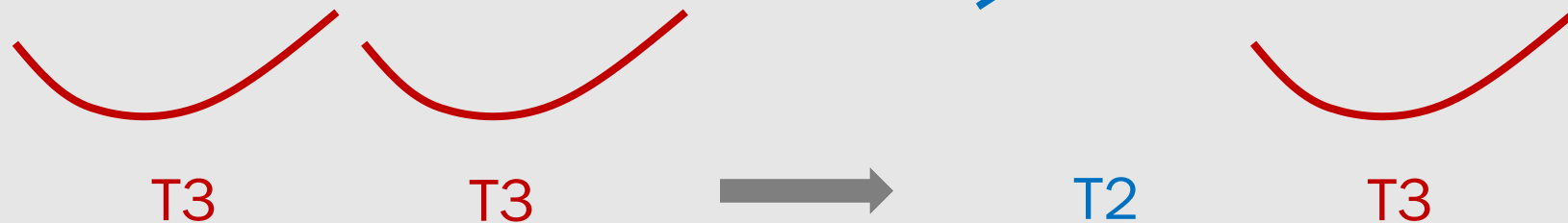
➤ The ingredients:

- *Tone 2: rising tone*
- *Tone 3: low, dipping tone, often accompanied with creaky voice.*

➤ The rule:

- $T3 \rightarrow T2 / _ T3$

(2)



Tonal Neutralization

- Tone 3 sandhi leads to neutralization in disyllabic words:

- /T3 T3/ → [T2 T3]

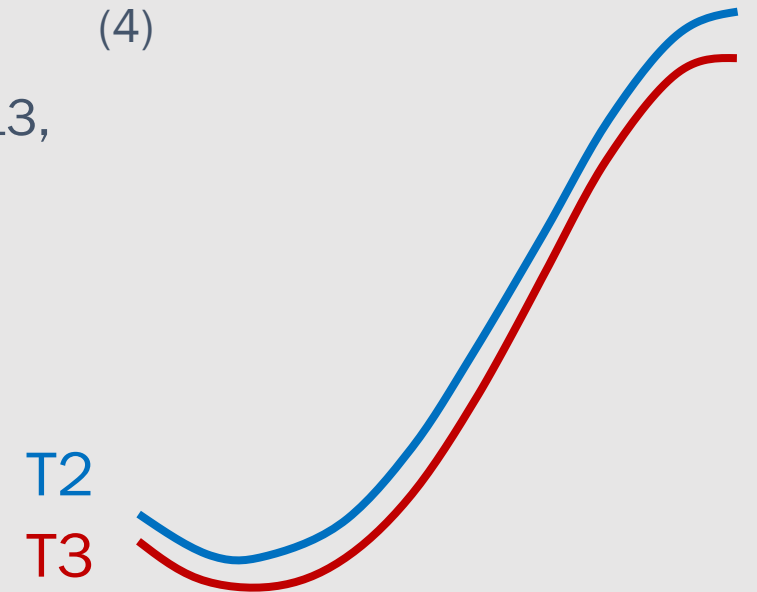
- /T2 T3/ → [T2 T3]

(3) Chinese	UR	SR	English
a. 百马	/paj3 ma3/	[paj2 ma3]	'a hundred horses'
b. 白马	/paj2 ma3/	[paj2 ma3]	'white horse'

- Phonological learning question: how did speakers learn the tonal UR?

Incomplete Neutralization

- Any acoustic cues to help speakers learn the tonal UR? (4)
- Incomplete neutralization reported in M. Lin 1980, Liu 2013, Wang & Li 1967, Yuan & Chen 2011.
- Some speakers produce a small acoustic difference:
 - /T3/ has a lower rising range than /T2/.
- But no speaker can perceive the difference.
- Speakers receive no help from acoustic information.



Pitch track roughly copied from Liu 2013

Compositional Transparency Helps

(5) Target word

Related word

a. 百 马
[pai2 ma3]
'hundred' 'horse'

b. 八 百
[pa1 pai3]
'eight' 'hundred'

c. 白 马
[pai2 ma3]
'white' 'horse'

d. 白 色
[pai2 se4]
'white' 'color'

- In morphologically related words, tone 3 sandhi is not applied. Tonal UR surfaces.

Prediction & Problem

- Speakers rely on morphophonological alternations to learn unfaithful UR-SR mappings.
- ❖ Prediction: if a disyllabic word with [T2 T3] is not compositionally transparent, then it would be learned as /T2 T3/ by the learner, no matter what the dictionary says.
- ❖ Problem: We don't know what UR speakers have learned, since all /T3 T3/ words only surface as [T2 T3].
- Solution: a novel AABB reduplication diagnostic to reveal tonal UR.

What Is AABB Reduplication?

- AABB reduplication: a semi-productive process for adjectives.

AB Base form	English	AABB Reduplicated form	English
(6) a. kan1 tɕiŋ4	‘clean’	b. kan1 kan1 tɕiŋ4 tɕiŋ4	‘very clean’ (‘ <i>clean everywhere</i> ’)
(7) a. ɰen4 tʂen1	‘careful’	b. ɰen4 ɰen4 tʂen1 tʂen1	‘very careful’ (‘ <i>careful at every moment</i> ’)

Reduplication to the Rescue

- /T3 T3/ base and /T2 T3/ base have different AABB reduplicated forms.

AB Base UR	English	AABB reduplicated SR
(8) a. / two3 ʃan3 /	'evasive'	b. two2 two2 ʃan2 ʃan3
'hide' 'dodge'		c. two2 two3 ʃan2 ʃan3
(9) a. / xuŋ2 xwo3 /	'flourishing (business)'	b. xuŋ2 xuŋ2 xwo2 xwo3
'red' 'fire'		c. * xuŋ2 xuŋ3 xwo2 xwo3

Derivation of Reduplication

(10) Reduplicate via SR

Base	[two ² şan ³] ‘evasive’	[xʊŋ ² xwo ³] ‘flourishing’
AABB reduplication	two ² two ² şan ³ şan ³	xʊŋ ² xʊŋ ² xwo ³ xwo ³
Tone 3 sandhi	two ² two ² şan ² şan ³	xʊŋ ² xʊŋ ² xwo ² xwo ³
Derived	[two ² two ² şan ² şan ³]	[xʊŋ ² xʊŋ ² xwo ² xwo ³]

(11) Reduplicate via UR

Base	/two ³ şan ³ / ‘evasive’	/xʊŋ ² xwo ³ / ‘flourishing’
AABB reduplication	two ³ two ³ şan ³ şan ³	xʊŋ ² xʊŋ ² xwo ³ xwo ³
Tone 3 sandhi	two ² two ³ şan ² şan ³	xʊŋ ² xʊŋ ² xwo ² xwo ³
Derived	[two ² two ³ şan ² şan ³]	[xʊŋ ² xʊŋ ² xwo ² xwo ³]

- A disyllabic word with a /T3 T3/ UR has two AABB reduplicated variant forms.
- A /T2 T3/ word only has one AABB reduplicated form.

AABB Reduplication Diagnostic

(12)

	[T2 T2 T2 T3]	[T2 T3 T2 T3]
/T3 T3/	Accept	Accept
/T2 T3/	Accept	REJECT!

❖ The diagnostic:

- If the speaker rejects [T2 T3 T2 T3], then their UR is /T2 T3/.

Speaker Judgement Survey

- I adapt the AABB reduplication diagnostic to nouns.
- I surveyed 6 native Mandarin speakers for their judgement on AABB nouns.
- Instruction to speakers: an AABB noun means ‘every AB’, with examples.
- 40-word list:
 - Both “T3 T3” words and “T2 T3” words, as listed in the dictionary.
 - Loanwords, animal and plant names, place names, compounds.
 - Some compounds are very transparent, some less so.

Survey Format

4. “蚂蚁”一词，组成AABB的叠词后，读音是：

má má yí yǐ 蚂蚁蚂蚁 Option A: [T2 T2 T2 T3]

má mǎ yí yǐ 蚂蚁蚂蚁 Option B: [T2 T3 T2 T3]

两种形式都可以 Option C: Both forms are fine

The word ‘ant’, after AABB reduplication, is pronounced as:

⇒ Speaker’s UR for ‘ant’ is /T2 T3/

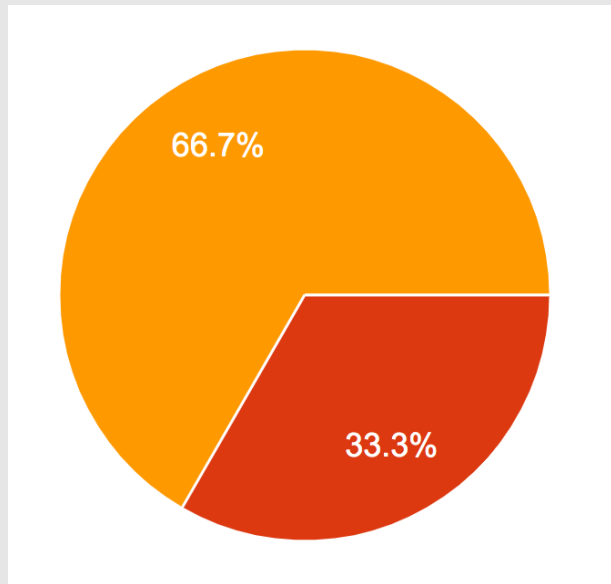
⇒ Speaker’s UR for ‘ant’ is /T3 T3/

⇒ Speaker’s UR for ‘ant’ is /T3 T3/

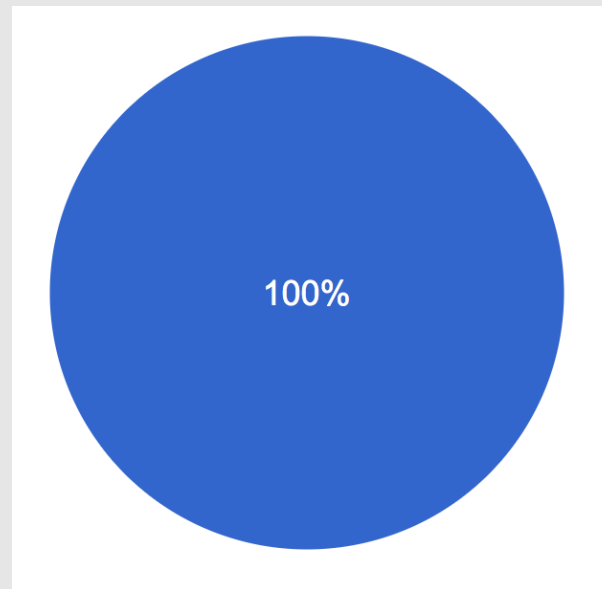
Speaker judgement

- Speakers show clear judgement on tones of AABB nouns.
- Semantically unnatural, yet phonologically natural.

(13) a. /ɥy3 san3/ 'rain umbrella'



b. /jan2 san3/ 'sun umbrella'



Blue: Option A
[T2 T2 T2 T3]

Red: Option B
[T2 T3 T2 T3]

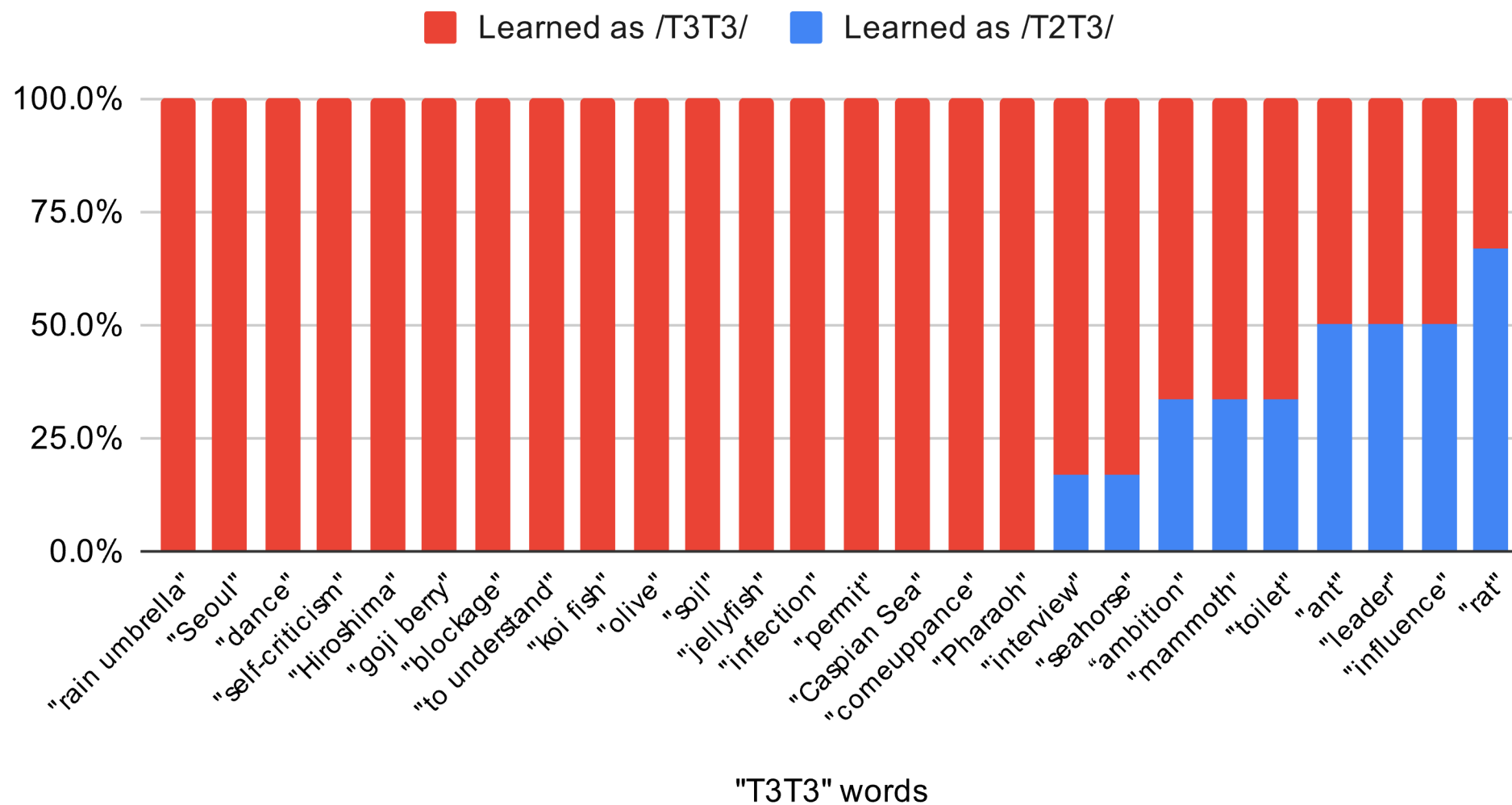
Orange: Option C
Both forms are okay

UR Learning and Compositionality

- *Recall that*
- ❖ Prediction: if a disyllabic word with [T2 T3] is not compositionally transparent, then it would be learned as /T2 T3/ by the learner, no matter what the dictionary says.
- ❖ Focus on “T3 T3” words: are they learned as /T3 T3/ or /T2 T3/?

UR learned by speakers for "T3T3" words

(14)



Compositionally Opaque Words

- Initial syllable does not contribute to meaning
→ prone to being learned as /T2 T3/

(15)

Dictionary tone	English	First syllable	Second syllable	/T2 T3/ learning rate
a. mɤŋ3 ma3 猛 犸	'mammoth'	mɤŋ3 '?'	ma3 '?'	33% speakers
b. ma3 ji3 蚂 蚁	'ant'	ma3 '?'	ji3 'termite'	50% speakers
c. law3 ʂu3 老 鼠	'rat'	law3 'old'	ʂu3 'rodent'	67% speakers
d. ma3 tʰŋ3 马 桶	'toilet'	ma3 'horse'	tʰŋ3 'bucket'	33% speakers

Compositionally Transparent Words

- Never learned as /T2 T3/.

(16)

Dictionary tone	English	First syllable	Second syllable	/T2 T3/ learning rate
a. ɥy3 san3 雨 伞	'rain umbrella'	ɥy3 'rain'	san3 'umbrella'	0% speakers
b. t ^h u3 ɭaŋ3 土 壤	'soil'	t ^h u3 'dirt'	ɭaŋ3 'soil'	0% speakers
c. ʂwej3 mu3 水 母	'jellyfish'	ʂwej3 'water'	mu3 'mother'	0% speakers
d. li3 xaj3 里 海	'Caspian Sea'	li3 'inside'	xaj3 'sea'	0% speakers

Somewhere in between

(17)

Dictionary tone	English	First syllable	Second syllable	/T2 T3/ learning rate
a. li3 ɕjaŋ3 理想	'ambition'	li3 'theory, ideal'	ɕjaŋ3 'think'	33% speakers
b. jiŋ3 ɕjaŋ3 影响	'influence'	jiŋ3 'shadow'	ɕjaŋ3 'sound'	50% speakers
c. liŋ3 taw3 领导	'leader'	liŋ3 'lead'	taw3 'direct'	50% speakers
d. ts ^h aj3 faŋ3 采访	'interview'	ts ^h aj3 'to pick'	faŋ3 'to visit'	17% speakers

- All highly lexicalized, but initial syllable still contributes to the overall meaning of the word.
- Relatively transparent to an educated adult speaker, but opaque to a child learner.

The case of 理 *li3* ‘theory, ideal’

(18)

Dictionary tone	English	First syllable	Second syllable	/T2 T3/ learning rate
a. <i>li3 tɕje3</i> 理解	‘to understand’	<i>li3</i> ‘theory, ideal’	<i>tɕje3</i> ‘to solve’	0% speakers
b. <i>li3 ɕjaŋ3</i> 理想	‘ambition’	<i>li3</i> ‘theory, ideal’	<i>ɕjaŋ3</i> ‘think’	33% speakers

- ‘To understand’: compositionally transparent.
- ‘ambition’ much less so: *thinking about one’s ideal* --> *ambition*
- At the age of acquisition, some children can only figure out the meaning contribution of *li3* in ‘to understand’, but not in ‘ambition’.
- *li3* is a morpheme in ‘to understand’, but not a morpheme in ‘ambition’.

Implication on Morpheme Size

- ❖ Conclusion: Not every syllable in Mandarin is a morpheme.

- Echoed in Packard (2000):
 - *Phonetic loanwords and strongly lexicalized disyllabic words are not composed of two morphemes.*
 - *The meaning of one or both syllables are opaque to the speaker.*
 - *The same word might be transparent to one speaker, but opaque to another.*

Implication on Morpheme Size

- Data from tonal UR learning provide concrete evidence for Packard's observations.
 - If the initial syllable in a disyllabic tone 3 sandhi word is not considered as a morpheme by the individual speaker, then no morphophonological alternation can be established.
 - Thus the speaker learns /T2 T3/ as the UR, identical to SR.
- Speaker variation in UR learning
 - Compositional transparency is determined by the individual speaker
 - Compositional transparency is determined by the child learner, not the grammarian or the adult speaker.

Implication on Chinese Orthography

❖ Common misconception: Chinese characters denote meanings, not sounds.

■ But now that we know:

- A Mandarin syllable does not necessarily map onto a morpheme,
- Every Mandarin syllable is transcribed by a Chinese character.

■ ⇒ There exists syllables that are transcribed by meaningless characters.

■ ‘Horse bucket’:

- *ma2* is transcribed as 马.
- But it does not have the pronunciation or the meaning of 马.

(19) *ma2* tʰɯŋ3
 马 桶
 ‘horse’ ‘bucket’?
 ‘toilet’

(20) *ma3*
 马
 ‘horse’

■ Chinese characters are merely tools to transcribe (or approximate) sound.



Thank you!

- **SELECTED REFERENCES:**

- Lin, Maocan, Lianhe Lin, Guangrong Xia, & Yusheng Cao (1980). Putonghua erzici biandiao de shiyan yanjiu. [Experimental study of Mandarin disyllabic word tone sandhi.] *Zhongguo Yuwen* [Studies of the Chinese Language] 1980:3.
- Liu, Xiaohui (2013). Shangsheng biandiao de shengxue yu ganzhi shiyan yanjiu. [Acoustic and perceptual experimental study of third tone sandhi.] *Wenjiao Ziliao* [Data of Culture and Education] 625. 139-142.
- Packard, Jerome (2000). *The Morphology of Chinese: A Linguistic and Cognitive Approach*. Cambridge University Press.
- Wang, William Shi-Yuan, & Kung-Pu Li (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research*. 10:3. 629-636.
- Yuan, Jiahong, & Yiya Chen (2011). 3rd tone sandhi in Standard Chinese: a corpus approach.

Many thanks to Adam Albright, Fulang Chen, Edward Flemming, Filipe Kobayashi, Michael Kenstowicz, Anton Kukhto, Patrick Niedzielski, Catarina Soares, and Donca Steriade for valuable discussions. The project also benefited from audience feedback at CLS58, the MIT reading groups Ling-Lunch, Morphun ,and Phonology Circle. All remaining mistakes are my own.

Exceptions!

- These 2 words are compositionally opaque, but no speaker learned them as /T2 T3/. Why?

(21)

Dictionary tone	English	First syllable	Second syllable	/T2 T3/ learning rate
a. kan3 lan3 橄 榄	'olive'	kan3 '?'	lan3 '?'	0% speakers
b. kow3 tɕ ^h i3 枸 杞	'goji berry'	kow3 '?'	tɕ ^h i3 '?'	0% speakers

Tone 2 Alternative

(22) Unavailable tone 2 alternative

0% /T2 T3/	Tone3	Tone2
a. kan3 lan3 'olive'	kan3 'dare'	*kan2
b. kow3 tɕ ^h i3 'goji'	kow3 'dog'	*kow2

(23) Available tone 2 alternative

> 0% /T2 T3/	Tone3	Tone2
a. ma3 ji3 'ant'	ma3 'horse'	ma2 'hemp'
b. law3 ʂu3 'rat'	law3 'old'	law2 'labor'

- Neither *kan2 nor *kow2 corresponds to an attested lexical item.
- If there is no tone 2 alternative, then speaker posits tone 3 as UR.

What about Orthography?

(24)

"T3 T3" word	Related character
a. kan3 lan3 橄 榄 'olive'	c. kan3 敢 'dare'
b. kow3 tɕhi3 枸 杞 'goji berry'	d. kow3 狗 'dog'

(25)

"T3 T3" word	Related characters		
a. ma3 ji3 蚂 蚁 'ant'	b. ma3 马 'horse'	c. ma3 码 'code'	d. ma3 玛 loanword character
a. ma3 tʰuŋ3 马 桶 'toilet'	b. ma1 妈 'mum'	c. ma4 骂 'scold'	d. ma0 吗 question particle

- But the tones of related characters might not be conclusive.
- Even when the character of the first syllable itself is a common character like 马 'horse', speakers still ignore it.

Observable Variation

- The tonal free variation in some words can be accounted for by variation in UR learning in related words that might have been acquired earlier.

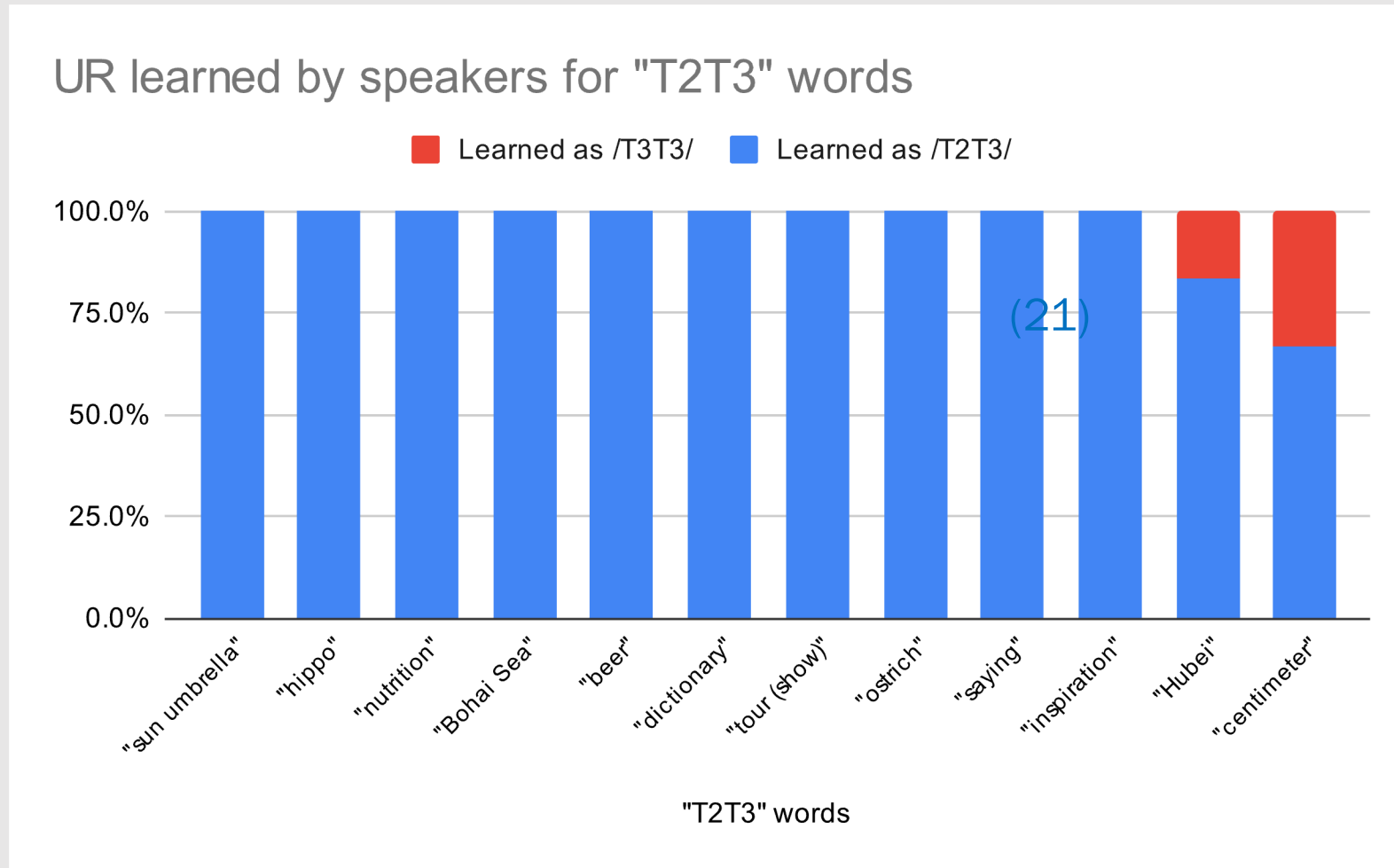
(26)

Words with variant tone	Related word	UR: /23/	UR: /33/
a. tɕ ^h ɛn ^{2/3} fu ² 潜 伏 'to go undercover'	a. tɕ ^h ɛn ² ɕweɰ ³ 潜 水 'to scuba dive'	33%	67%
b. tɕ ^h i ^{2/3} fu ² 祈 福 'to pray for blessings'	b. tɕ ^h i ² taw ³ 祈 祷 'to pray'	67%	33%

Acquired first

What about “T2 T3” Words?

(27)



Remaining Puzzle: Verbs vs. Nouns

(17)

Dictionary tone	English	First syllable	Second syllable	/T2 T3/ learning rate
a. li3 ɕjaŋ3 理想	'ambition'	li3 'theory, ideal'	ɕjaŋ3 'to think'	33% speakers
b. jiŋ3 ɕjaŋ3 影响	'influence'	jiŋ3 'shadow'	ɕjaŋ3 'sound'	50% speakers
c. liŋ3 taw3 领导	'leader'	liŋ3 'to lead'	taw3 'to direct'	50% speakers
d. ts ^h aj3 faŋ3 采访	'interview'	ts ^h aj3 'to pick'	faŋ3 'to visit'	17% speakers

- The above items also all have the dual identity of being verbs and nouns.
- When *liŋ3 taw3* is a verb, [T2 T3 T2 T3] is more acceptable than when it is a noun (F. Chen p.c.) --> Noun: /T2 T3/; Verb: /T3 T3/

Remaining Puzzle: Verbs vs. Nouns

- A-not-A construction is a way to ask yes/no questions.
- A disyllabic verb of the shape AB is usually made into “A not AB” in this construction.
- The lexical item of “not” does not trigger sandhi on A. This is a place one might be able to observe the tonal UR of A (Kobayashi p.c.)
- Judgement from the 6 speakers: general preference for [T2 not T2 T3]
- Possibility of identity requirement between the the syllable before negation and after negation (Kukhto p.c.)

(28)

Dictionary tone	English	AABB diagnostic	A not AB tone judgement
a. li ³ ɕjaŋ ³ 理想	‘to dream’	33% speakers: /T2 T3/ 67% speakers: /T3 T3/	50% choose [T2 not T2 T3] 50% choose [T3 not T2 T3]
b. jɿŋ ³ ɕjaŋ ³ 影响	‘to influence’	50% speakers: /T2 T3/ 50% speakers: /T3 T3/	100% choose [T2 not T2 T3] 0% choose [T3 not T2 T3]
c. liŋ ³ taw ³ 领导	‘to lead’	50% speakers: /T2 T3/ 50% speakers: /T3 T3/	100% choose [T2 not T2 T3] 0% choose [T3 not T2 T3]
d. tshaj ³ faŋ ³ 采访	‘to interview’	17% speakers: /T2 T3/ 83% speakers: /T3 T3/	67% choose [T2 not T2 T3] 33% choose [T3 not T2 T3]