UR UNDERLEARNING OF MANDARIN CHINESE TONE 3 SANDHI WORDS

BOER FU

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

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

GOAL OF THE PROJECT

- To investigate whether speakers can learn underlying representations that are different from surface representations without morphological alternation
- With a case study of Mandarin tone 3 sandhi

- *Why is this an interesting question to ask?*
- *Why can Mandarin tone 3 sandhi help us answer this question?*

UR LEARNING

- How do speakers learn Underlying Representations (UR) that are different from Surface Representations (SR)?
- Morphological alternations help.

(1) German final devoicing

a.	[taːk]	‘day.$_{SG}$’	[taːgə]	‘day.$_{PL}$’	c.	[fʁɔʏnt]	‘friend.$_{SG}$’	[fʁɔʏndə]	‘friend.$_{PL}$’
c.	[baŋk]	‘bench.$_{SG}$’	[bɛŋkə]	‘bench.$_{PL}$’	d.	[huːt]	‘hat.$_{SG}$’	[hyːtə]	‘hat.$_{PL}$’

UR LEARNING WITHOUT ALTERNATION

- But what if morphological alternation is not available to the learner?
- Would speakers still be able to recover non-identical UR-SR mapping?

❖ Case study: Mandarin tone 3 sandhi

- Mandarin has a shortage of morphological alternations, making it the perfect test case for UR learning without alternation.

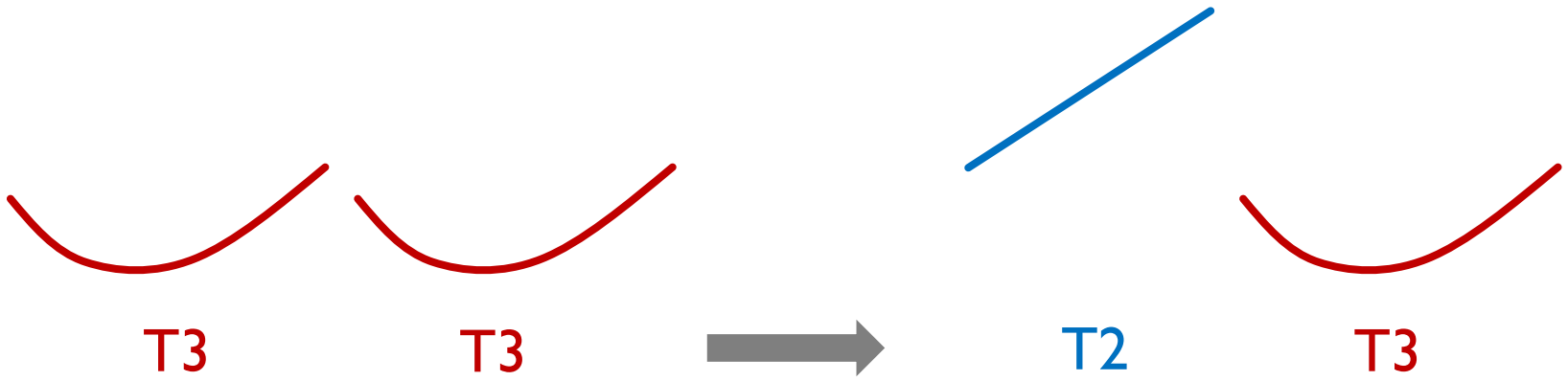
ROADMAP

- Mandarin tone 3 sandhi
- Incomplete neutralization
- Problem: Speaker tonal UR not known
- Solution: novel AABB reduplication diagnostic
- Survey of speaker judgement
- Implication on UR learning

MANDARIN TONE 3 SANDHI

- The ingredients:
  - Tone 2: rising tone
  - Tone 3: low, dipping tone, often accompanied with creaky voice.
- The rule:
  - T3 → 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.	五笔	/wu3 pi3/	[wu2 pi3]	'five-stroke input system'
b.	无比	/wu2 pi3/	[wu2 pi3]	'unparalleled'

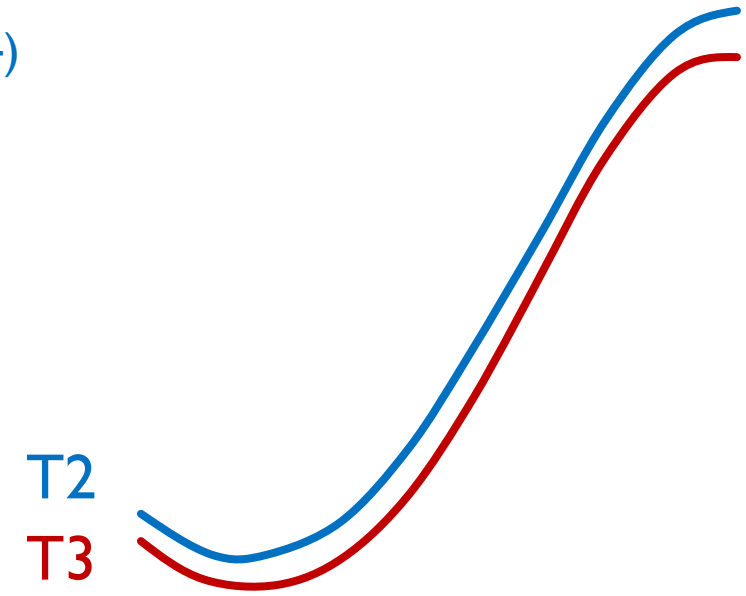
- No affixes to reveal the underlying tone.

➢ How do speakers know which is which?

INCOMPLETE NEUTRALIZATION

- Any acoustic cues to help speakers learn the tonal UR?
- 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.

(4)



Pitch track roughly copied from Liu 2013

PROBLEM AND SOLUTION

- Incomplete neutralization not only poses a challenge for Mandarin learners, but also presents a problem for phonologists.
- ❖ Problem: we don't know what tonal UR the speakers have learned.
  - Previous research mostly took dictionary entries at face value.
- ➢ Solution: novel AABB reduplication diagnostic
  - A new method that can reveal the speaker's tonal UR for a surface [T2T3] word in a phonologically natural, yet semantically contrived environment.

WHAT IS AABB REDUPLICATION?

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

	AB Base form	English	AABB Reduplicated form	English
(5) a.	kan1 tɕiŋ4	'clean'	b. kan1 kan1 tɕiŋ4 tɕiŋ4	'very clean' ('*clean everywhere*')
(6) a.	ɻen4 tʂen1	'careful'	b. ɻen4 ɻen4 tʂen1 tʂen1	'very careful' ('*careful at every moment*')

REDUPLICATION TO THE RESCUE

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

	AB Base UR	English	AABB reduplicated SR
(7) a.	/two3 ʂan3/	'evasive'	b. two2 two2 ʂan2 ʂan3
	'hide' 'dodge'		c. two2 two3 ʂan2 ʂan3
(8) a.	/xʊŋ2 xwo3/	'flourishing (business)'	b. xʊŋ2 xʊŋ2 xwo2 xwo3
	'red' 'fire'		c. **xʊŋ2 xʊŋ3 xwo2 xwo3*

DERIVATION OF REDUPLICATION (VIA SR OR UR)

(9) *Reduplicate via SR*

Base	[two2 ʂan3] 'evasive'	[xʊŋ2 xwo3] 'flourishing'
AABB reduplication	two2 two2 ʂan3 ʂan3	xʊŋ2 xʊŋ2 xwo3 xwo3
Tone 3 sandhi	two2 two2 ʂan2 ʂan3	xʊŋ2 xʊŋ2 xwo2 xwo3
Derived	[two2 two2 ʂan2 ʂan3]	[xʊŋ2 xʊŋ2 xwo2 xwo3]

(10) *Reduplicate via UR*

Base	/two3 ʂan3/ 'evasive'	/xʊŋ2 xwo3/ 'flourishing
AABB reduplication	two3 two3 ʂan3 ʂan3	xʊŋ2 xʊŋ2 xwo3 xwo3
Tone 3 sandhi	two2 two3 ʂan2 ʂan3	xʊŋ2 xʊŋ2 xwo2 xwo3
Derived	[two2 two3 ʂan2 ʂan3]	[xʊŋ2 xʊŋ2 xwo2 xwo3]

- 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

(11)

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

❖ **The diagnostic:**

- For a disyllabic word with surface [T2 T3], if the speaker rejects [T2 T3 T2 T3] as an AABB reduplicated form, then the speaker has posited a /T2 T3/ UR for the lexical item. Otherwise the UR is /T3 T3/.

SPEAKER JUDGEMENT SURVEY

- The AABB reduplication diagnostic can be adapted to reveal tonal UR for nouns.
- I surveyed 6 native Mandarin speakers for their judgement on AABB nouns.
- They were given the instruction that an AABB noun means 'every AB'.
- The survey has 40 words, including both "T3 T3" words and "T2 T3" words, as listed in the dictionary.
- Focus on words listed as "T3 T3".
- How well have the speakers learned /T3 T3/ as the UR?

SURVEY FORMAT

(12)

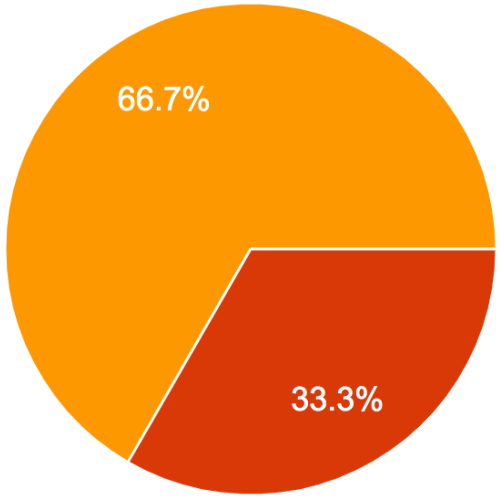
4. “蚂蚁”一词，组成AABB的叠词后，读音是： The word ‘ant’, after AABB reduplication, is pronounced as:

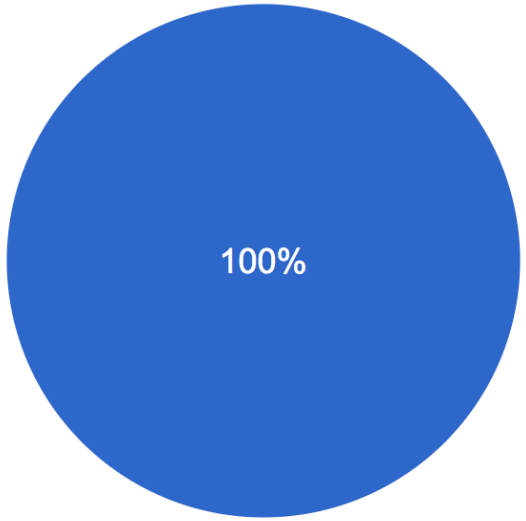
○ má má yí yǐ 蚂蚂蚁蚁	Option A: [T2 T2 T2 T3]	⇒ Speaker’s UR for ‘ant’ is /T2 T3/
○ má mǎ yí yǐ 蚂蚂蚁蚁	Option B: [T2 T3 T2 T3]	⇒ Speaker’s UR for ‘ant’ is /T3 T3/
○ 两种形式都可以	Option C: Both forms are fine	⇒ Speaker’s UR for ‘ant’ is /T3 T3/

SPEAKER JUDGEMENT

- The speakers are shown to have clear judgement on tones in AABB reduplicated forms, even though these items are semantically unnatural.
- Speakers' choice for the first two words in the survey: both compositionally transparent.

(13) a. /ɥy3 san3/ 'rain umbrella' b. /jaŋ2 san3/ 'sun umbrella'



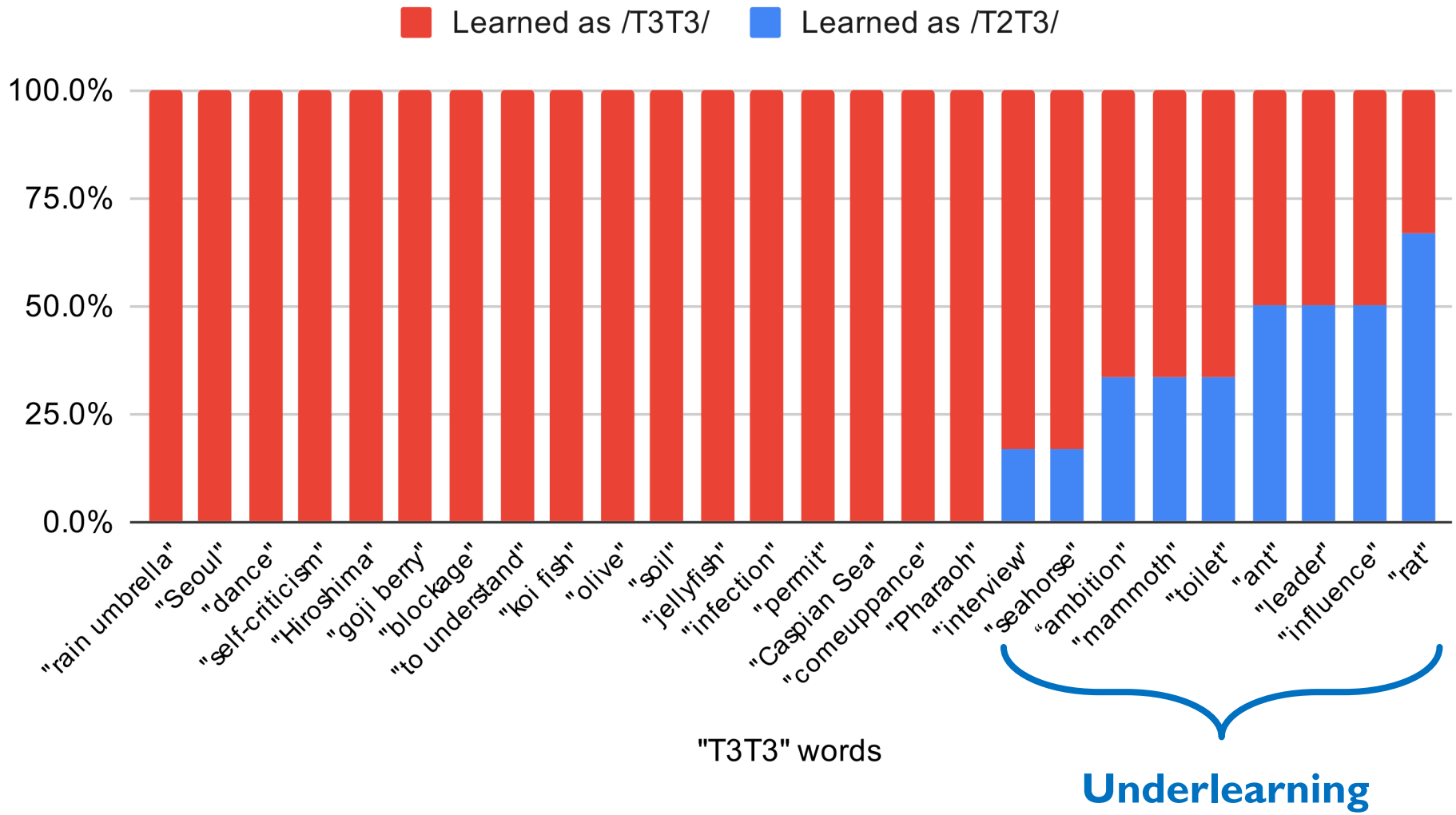


Blue: Option A: [T2 T2 T2 T3]
Red: Option B: [T2 T3 T2 T3]
Orange: Option C: both forms are okay

SURVEY RESULTS

(14)

UR learned by speakers for "T3T3" words



"UNDERLEARNING"

- A dictionary "T3 T3" learned as /T2 T3/ by the speaker.
- The word is purported to have undergone the phonological process of tone 3 sandhi, yet the speaker has acquired a grammar in which the sandhi rule is underapplied, hence the name "underlearning"
- Underlearn ≠ underperform
- When a case of "underlearning" is identified, the dictionary is in the wrong, not the speaker.

UR UNDERLEARNING

- "Underlearning" takes place in spite of speaker familiarity with dictionary entry.
- "T3T3" words that are prone to "underlearning" can be characterized as:
  (i) Compositionally opaque
  (ii) The first syllable has a tone 2 alternative

WORDS PRONE TO UNDERLEARNING: COMPOSITIONALLY OPAQUE

- The 4 words below all include an initial syllable that does not contribute to its meaning.

(15)

Dictionary tone	English	First syllable	Second syllable	Underlearning rate
a. ma3 ji3 蚂 蚁	'ant'	ma3 '?'	ji3 'termite'	50% speakers underlearn
b. law3 ʂu3 老 鼠	'rat'	law3 'old'	ʂu3 'rodent'	67% speakers underlearn
c. mɤŋ3 ma3 猛 犸	'mammoth'	mɤŋ3 '?'	ma3 '?'	33% speakers underlearn
d. ma3 tʊŋ3 马 桶	'toilet'	ma3 'horse'	t^{h}ʊŋ3 'bucket'	33% speakers underlearn

WORDS NEVER UNDERLEARNED: COMPOSITIONALLY TRANSPARENT

- The following compositionally transparent words are never underlearned.

(16)

Dictionary tone	English	First syllable	Second syllable	Underlearning rate
a. ɥy3 san3 雨 伞	'rain umbrella'	ɥy3 'rain'	san3 'umbrella'	0% speakers underlearn
b. tʰu3 ɻaŋ3 土 壤	'soil'	tʰu3 'dirt'	ɻaŋ3 'soil'	0% speakers underlearn
c. ʂwej3 mu3 水 母	'jellyfish'	ʂwej3 'water'	mu3 'mother'	0% speakers underlearn
d. li3 xaj3 里 海	'Caspian Sea'	li3 'inside'	xaj3 'sea'	0% speakers underlearn

EXCEPTIONS!

- These 2 words are compositionally opaque, but no speaker underlearn them. Why?

(17)

Dictionary tone	English	First syllable	Second syllable	Underlearning rate
a. kan3 lan3 橄　欖	'olive'	kan3 '?'	lan3 '?'	0% speakers underlearn
b. kow3 tɕʰi3 枸　杞	'goji berry'	kow3 '?'	tɕʰi3 '?'	0% speakers underlearn

TONE 2 ALTERNATIVE

(18) Unavailable tone 2 alternative

0% underlearn	Tone3	Tone2
a. kan3 lan3 'olive'	kan3 'dare'	*kan2
b. kow3 tɕʰi3 'goji'	kow3 'dog'	*kow2

(19) Available tone 2 alternative

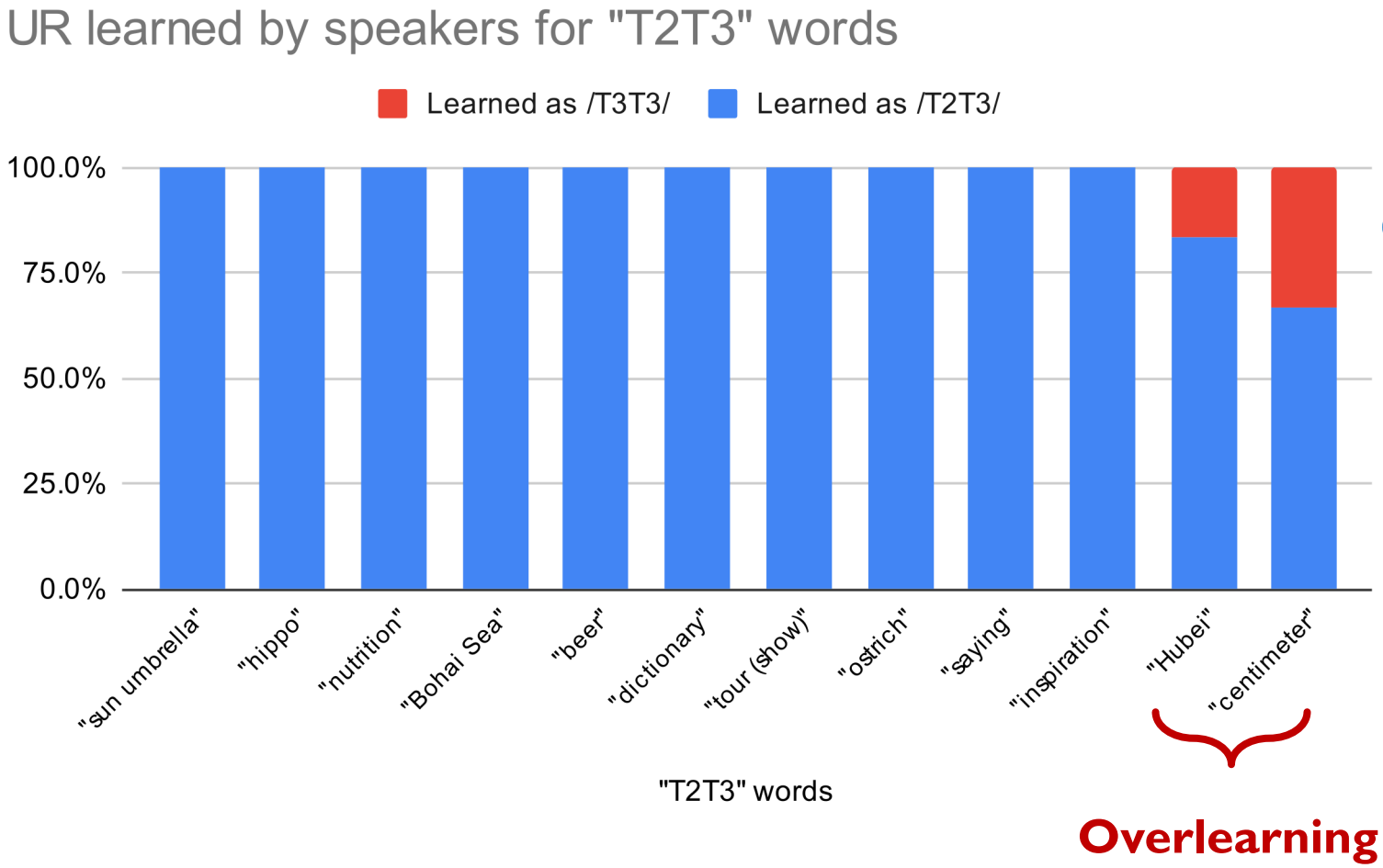
> 0% underlearn	Tone3	Tone2
a. ma3 ji3 'ant'	ma3 'horse'	ma2 'hemp'
b. law3 ʂu3 'rat'	law3 'old'	law2 'labor'

- 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 OVERLEARNING?

- "T2 T3" learned as /T3 T3/: the speaker has learned a UR that involves overapplication of sandhi:

(20)



(21)

First	Second	English
li2 'ancient unit'	mi3 'meter'	'centimeter'
xu2 'lake'	pej3 'north'	'Hubei'

IMPLICATION ON PHONOLOGICAL LEARNING: BIAS FOR FAITHFUL MAPPING

- Speakers tend to underlearn "T3 T3" as /T2 T3/, but they rarely overlearn "T2 T3" as /T3 T3/.
- Learning bias for establishing faithful mappings between SR and UR: Lexicon optimization.
- Unless given evidence to suggest otherwise, speakers prefer to posit URs that are identical to SRs.
- Evidence that suggests unfaithful mapping between SR and UR include:
  - A morphologically related lexical item that shows the first syllable in surface [T3]

    → Compositionally opaque words are prone to underlearning.
  - For the first syllable, there is no surface [T2] anywhere else in the lexicon

    → Syllables with tone 2 alternatives are prone to underlearning.

IMPLICATION ON PHONOLOGICAL LEARNING: LEARNERS NEED ALTERNATION

Research question:

Can speakers learn underlying representation without morphological alternation?

Answer: Not really.

- Mandarin alternations are found in morphologically-related, compositionally transparent compounds, as opposed to systematic inflections.
- Learners of Mandarin still rely on alternation to learn unfaithful UR-SR mapping.

IMPLICATION ON PHONOLOGICAL LEARNING: LEARNERS DECIDE ON WHAT COUNTS AS ALTERNATION

- Alternation can only serve as evidence for phonological learning if they can be identified as such by the learner.
- The same word is judged to be compositionally transparent by some learners, and opaque by others.
- Only those learners who judge the word to be compositionally transparent can identify alternation with morphologically-related words, and posit unfaithful UR-SR mapping.
- It is the child learner who decides on what counts as alternation, not the grammarian or the adult educated speaker. ('toilet' ≠ horse + bucket)
- German vs. Mandarin:
  - German: same alternation for every learner.
  - Mandarin: alternation varies between learners.

THANK YOU!

- **SELECTED REFERENCES:**

- Horslund, Camilla Søballe, Rasmus Puggaard & Henrik Jørgensen (2021). An alternative, phonetically based phoneme analysis of the Danish consonant system. 12th International Conference of Nordic and General Linguistics. Universitetet i Oslo (virtual).
- Inkelas, Sharon (1995). The Consequences of Optimization for Underspecification. In E. Buckley and S. Iatridou (eds.), *Proceedings of the Twenty-Fifth Northeastern Linguistics Society.* Amherst: GLSA. 287-302.
- 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.
- 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.

BONUS CONTENT
WHAT ABOUT ORTHOGRAPHY?

- An alternative account of the exception in underlearning of opaque is that speakers had help from orthography.

(21)

"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'

(22)

"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 tʰʊŋ3 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 underlearn.

BONUS CONTENT
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.

(23)

Words with variant tone	Related word	UR: /23/	UR: /33/
a. tɕʰjɛn2/3 fu2 潜 伏 'to go undercover'	a. tɕʰjɛn2 ʂwej3 潜 水 'to scuba dive'	33%	67%
b. tɕʰi2/3 fu2 祈 福 'to pray for blessings'	b. tɕʰi2 taw3 祈 祷 'to pray'	67%	33%

Acquired first

BONUS CONTENT
VERBS IN A-NOT-A QUESTION CONSTRUCTION

- 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:

(24)

Dictionary tone	English	Underlearning rate	A not AB tone
a. li3 ɕjɑŋ3 理 想	'to dream'	33% speakers underlearn	50% choose [T2 not T2 T3] 50% choose [T3 not T2 T3]
b. jiŋ3 ɕjɑŋ3 影　响	'to influence'	50% speakers underlearn	100% choose [T2 not T2 T3] 0% choose [T3 not T2 T3]
c. liŋ3 taw3 领　导	'to lead'	50% speakers underlearn	100% choose [T2 not T2 T3] 0% choose [T3 not T2 T3]
d. tsʰaj3 fɑŋ3 采　　访	'to interview'	16.7% speakers underlearn	66.7% choose [T2 not T2 T3] 33.3% choose [T3 not T2 T3]