Phonological representation of Chinese disyllabic words

Shu-Chuan Tseng
Institute of Linguistics, Academia Sinica, Taiwan

Introduction

Spoken words in spontaneous speech are often reduced, but not necessarily produced in any arbitrary form. This study explores whether there may exist a kind of invariability in surface forms, which is possibly associated with word-level processing in spoken word production and retrieval.

Coalescence of Chinese disyllables

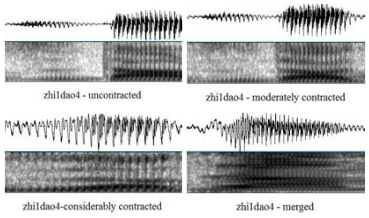
In ancient Chinese, a new/unknown character is transcribed by two already known characters, whose pronunciation has the onset from the first syllable and the rhyme from the second syllable. A similar coalescence mechanism can also be observed in some Chinese dialects, e.g., ki lai => kiai (to rise up) in Southern Min (Lien 1997).

Reduction of Chinese disyllabic words

- **Phonological rules** (coalescence, syllable merger)
  - Writing system (1) zhīhū 之乎 => zhū 諸
    之*pronoun*,乎*particle* => 諸"*all*", "*every*"
    (2) zhèyàngzi 這樣子"*this way*" =>jiàngzi 醬子)
  - Edge-in theory (Merger = S1-onset + S2-rhyme)
  - Sonority principle (Nuclei merging by sonority)
- **Phonetic reduction**
  - Gradient degrees in terms of acoustic features
- **Automatic speech recognition** (ASR)
  - Pronunciation variation modeling, dictionary
- **Mental representation of spoken words**
  - Categorical exemplar clouds

Syllable contraction in conversational speech: Four cases of reduced speech

zhīdào (to know) /tʂɯ.tau/ (Tseng *et al*. 2013)



Data

Taiwan Mandarin Conversational Corpus, 573K words

- Academic licenses are available via the Association for Computational Linguistics and Chinese Language Processing (ACLCLP)

- **Sinica MCDC8**: Eight hours of speech data annotated with interpausal unit boundary information (**IPU**), 123K syllables (Tseng 2019)

- Phone-aligned Chinese Conversational Speech Database: 3.5 hours of speech data with human-verified, signal-aligned phone boundary information; used as training data for the ILAS phone aligner

Taiwan Mandarin Conversational Corpus

	Mandarin Map Task Corpus (MMTC)	Mandarin Topic-oriented Conversation Corpus (MTTC)	Mandarin Conversational Dialogue Corpus (Sinica MCDC8 + MCDC22)
Speech type	Task-oriented	Topic-specific	Free conversation
Speaker	28F/24M	33F/25M	37F/23M
Recording	2002	2002	2001
Data	3 hrs. (29 dialogues)	10 hrs. (26 conversations)	30 hrs. (30 conversations)
Signal alignment	IPU/**word**/ syllable/phoneme	IPU/**word**/ syllable/phoneme	IPU/**word**/ syllable/phoneme

Acknowledgements:
This study was funded by the Ministry of Science and Technology, under grant MOST 106-2410-H-001-045-MY2.

Research motivation

Our hypothesis is that for Chinese disyllabic words, the coalescence form may also be a type of phonological representation, in addition to the canonical form. **Phonetic similarity scores** calculated on the basis of free phone recognition results and **word reduction type** (RT) determined by syllable contraction degrees were used to study gradient acoustic features and the derived phonological categories of disyllabic words in a Chinese conversational speech corpus.

Reduction types (RT)

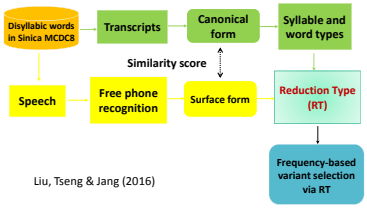
RTs are defined according to different degrees of syllable contraction observed in spontaneous speech: Canonical form type (**CAN**), marginal segment deletion type (**MSD**), nucleus merger type (**NUM**), and syllable merger type (**SYM**).

Constraints in RT derivation algorithm

More reduced ↑

RT	Word-internal syllable boundary	Segment omission	Nucleus length constraint	Cross-boundary segment preservation	*Xian zai* (now) /ejen.tsai/
SYM	-	Yes	<2	No	/ejai/
NUM	+/-	Yes	≥2	No	/ejenai/
MSD	+	Yes	≥2	Yes	/ejen.ai/
CAN	+	No	≥ 2	Yes	/ejen.tsai/

Deriving categorical RT



Liu, Tseng & Jang (2016)

"Typical" variants derived by means of RT

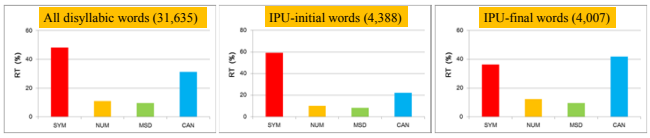
- 31,635 tokens of disyllabic words produced in the **Sinica MCDC8** were processed
- Typical variants of all disyllabic words were individually selected by signal-based scores and phonologically based algorithms, for which word-level variability and frequency effect are taken into consideration (Bybee 2001, Pierrehumbert 2001).
- Data-driven SYM forms may not always correspond to the phonologically predicted target forms (coalescence of two original syllables).

Majority RT and typical variants for disyllabic words that occur > 100 times in the Sinica MCDC8)

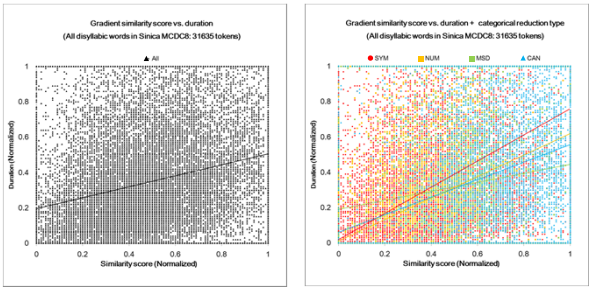
Word	Gloss	MCDC8 count	Canonical form	Majority RT	RT (%)	Typical variant
jiù shì	that is	1039	/tɕ j ou ʂ i/	SYM	50.82	/tɕ ɯ/
wǒ mén	we	878	/w o m ə n/	SYM	84.97	/ŋ/
rán hòu	then	733	/ʐ a n x ou/	SYM	79.81	/ʈ au/
juè dé	to feel	676	/tɕ ye t ə/	SYM	64.64	/tɕ ye/
yīn wèi	because	651	/i n w ei/	SYM	67.28	/ʐ ei/
méi yǒu	not/have not	571	/m ei j ou/	SYM	66.90	/m ə/
suǒ yǐ	so	479	/s w o i/	SYM	90.61	/ʦ ei/
kě shì	but	466	/kʰ ɔ ʂ i/	SYM	61.16	/kʰ ɯ/
qí shí	in fact	431	/tɕʰ i ʂ i/	SYM	54.06	/tɕʰ ɣ/
shí hòu	time	420	/ʂ i x ou/	SYM	83.10	/ʂ ou/
tā mén	they	415	/tʰ a m ə n/	SYM	83.37	/tʰ a ŋ/
xiàn zài	now	397	/ɕ j e n ʦ ai/	SYM	69.02	/ɕ j ai/
bǐ jiào	more	371	/p i tɕ j au/	SYM	63.34	/tɕ j au/
shén me	what	346	/ʂ ə n m o/	SYM	90.75	/ʂ ʂ/
zhè yàng	in this way	339	/ʈʂ ə j a ŋ/	SYM	87.02	/tɕ j a/
zhēn de	really	281	/ʈʂ ə n t ə/	SYM	64.77	/ʈʂ ə/
nà biān	there	269	/n a p j e n/	SYM	46.84	/n/
kě yǐ	can	260	/kʰ ɔ i/	SYM	85.00	/kʰ y/
hǎo xiàng	it seems	259	/x au ɕ j a ŋ/	SYM	57.53	/x a/
zhī dào	to know	234	/ʦ ɯ t au/	SYM	51.28	/ʦ au/
hěn duō	a lot	232	/x ə n t w o/	SYM	43.53	/ŋ t ou/
kě néng	possibly	223	/kʰ ɔ n ə ŋ/	SYM	84.75	/kʰ ə/
hái shi	still/or	218	/x ai ʂ ɯ/	CAN	47.71	/x ai ʂ i/
yì xiē	some	202	/i ɕ j e/	CAN	62.87	/i ɕ j e/
bù huì	will not	199	/p u x w ei/	SYM	64.32	/p w ei/
ér qiě	however	194	/ɚ tɕʰ j e/	CAN	59.28	/ɚ ɕ ye/
xué xiào	school	172	/ɕ ye ɕ j au/	SYM	55.81	/ɕ j au/
yīng gāi	should	168	/i ŋ k ai/	SYM	41.07	/n ə/
zì jǐ	-self	168	/ʦ i tɕ i/	SYM	62.50	/ʦ i/
dōng xi	thing	149	/t o ŋ ɕ i/	MSD	41.61	/w ɕ i/
yǐ qián	before	147	/i tɕʰ j e n/	CAN	78.91	/i ɕ ye n/
rú guǒ	if	139	/ʐ ɥ k w o/	SYM	81.29	/w o/
nǐ mén	you (plural)	138	/n i m ə n/	SYM	85.51	/n/
dàn shì	but	137	/t a n ʂ ɯ/	SYM	44.53	/t ə/
lǐ miàn	inside	126	/l i m j e n/	CAN	60.32	/i m e/
zěn mó	how	124	/ʦ ə n m o/	SYM	91.13	/ʦ ŋ/
gōng sī	company	121	/k o ŋ s i/	MSD	46.28	/k u ʂ i/
yǐ jīng	already	117	/i tɕ i ŋ/	SYM	49.57	/i/
yí yàng	the same	115	/i j a ŋ/	CAN	54.78	/i j a/
rén jiā	people	112	/ʐ ə n tɕ j a/	SYM	51.79	/n j ai/
dà gài	probably	108	/t a k ai/	SYM	62.96	/t ai/
dà jiā	everyone	100	/t a tɕ j a/	SYM	72.00	/t a/

(Modified from the results presented in Liu, Tseng & Jang 2016)

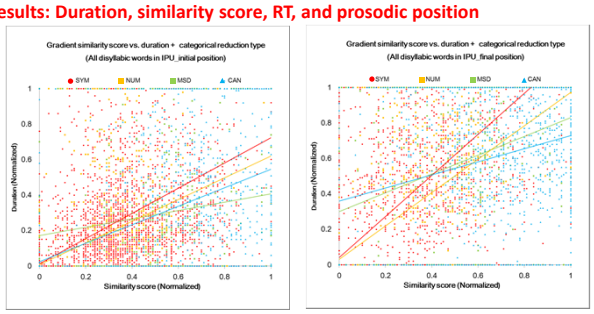
Results: RT distribution (Sinica MCDC8)



Results: Duration and similarity score (gradient) vs. RT (categorical)

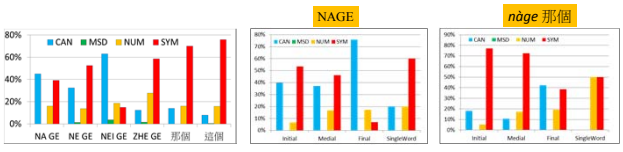


Results: Duration, similarity score, RT, and prosodic position



RT of disyllabic demonstratives (word category, discourse function, RT, prosodic position)

- 3% of tokens in the Sinica MCDC8 are demonstratives: Lexical and filler uses
  - **Lexical use** of demonstratives:
    Distal: *that* 那 *nà*/na/, *nè*/nə/, *nèi*/nei/
    Proximal: *this* 這 *zhè*/tʂə/, *zhèi*/tʂei/
    } w/o 個 *ge* (a classifier for general use)
  - **Demonstrative fillers**: NA (NE/NEI), NAGE (NEGE/NEIGE), ZHEGE (ZHEIGE)



Observations

- **CAN and SYM are both preferred**.
- In general, the larger the similarity score is, the longer the duration is.
- RT corresponds to the similarity score by reduction degree and duration.
- For words in the IPU-initial position, SYM is preferred and produced with a short duration, while those in the final position, both of CAN and SYM are preferred: Cues for **boundary effects** in conversational speech (initial shortening and final lengthening).
- Phonological representation of disyllabic words may be conditioned by **production frequency** as well as by the associated **prosodic positions**.
- **Uses of demonstratives differ in terms of RT distribution pattern (discourse function?)**

Implications

The more reduced a word is, the more difficult to recognize it? Not necessarily.

A word recognition experiment was conducted by Lee (2018), using CAN, MSD, and SYM disyllabic stimuli whose surface form was derived by the Edge-in theory (Chung 1997), the Sonority Principle (Hsu 2003), and the Discriminative Trace Principle (Tseng 2005) for high- and low-frequency words selected from the Sinica MCDC8. **CAN and SYM were statistically more quickly recognized than MSD irrespective of production frequency.**

Conclusion: The extreme case of phonetic reduction of Chinese disyllabic words (**SYM**) that resembles the coalescence form may possibly be a categorical type of spoken word variant that is phonologically represented in the mental lexicon.