Phonetic fusion in Chinese conversational speech

Shu-Chuan Tseng
Academia Sinica

This paper presents a corpus-based perspective on the phonetic fusion of disyllabic words in a Chinese conversational speech corpus. Four categorical types that reflect the phonological features of reduction degrees are automatically derived from gradient, acoustic properties. A transcription experiment is conducted with the most common disyllabic words. Both automatic derivation by acoustic signals and human transcription by perceptual judgment refer to the same sound inventory. We have shown that the complete form of fusion occurring in conversation need not be legitimate syllables and it appears consistently in the form of syllable merger that represents a group of phonetic variants.

Keywords: conversational speech, phonetic variation, syllable merger

1. Introduction

Lexicon constituents are part of an intrinsic knowledge system facilitating simultaneous operations of processing and understanding the substance of the delivered message in interactive speech. Integrative lexical information is connected with multiple linguistic levels, e.g., the phonological representation, the syntagmatic and paradigmatic constructional property, the meaning, and the sociopragmatic context. During the course of verbal communication activities, experiences and knowledge of these kinds may be envisaged in overt and covert manners. Presumably, language-intrinsic regularities should predict the phonetic forms of spoken words. However, speakers have to comply with language-external contextual factors with spontaneity and semantic predictability. Speech communication may act as a trivial social activity but actually involves a set of complex and complicated mechanisms. To project this activity onto a focused domain, this paper primarily concerns the lexical representation of spoken words produced in conversation. Spontaneously produced speech is often phonetically reduced. From an

https://doi.org/10.1075/cld.21004.tse | Published online: 14 September 2021
Chinese Language and Discourse ISSN 1877-7031 | E-ISSN 1877-8798

empirical standpoint, reduced speech is an ideal interface between the phonetic form and phonological representation of spoken words for observation. In particular, phonetic fusion in disyllables is explicitly associated with phonological regularity, word morphology and orthography in Chinese. In this paper, we attempt to analyze spoken data of disyllabic words from a conversational speech corpus and provide empirical evidence that reveals possible phonological representation forms as accessible constituents other than the citation form of Chinese words.

1.1 Mandarin Chinese

Taiwan is a multilingual society. Some of the major varieties of Chinese, including Mandarin Chinese (*Mandarin*), Southern Min (*Mǐn*), and Hakka (*Hakka*) as well as Formosan languages, are spoken in Taiwan (Li and Thompson, 2009). As the corpora of Mandarin Chinese we use for the analysis in this paper are collected in Taiwan, the conclusions we draw from the results apply exclusively to this particular variety of Chinese. Mandarin Chinese is an isolating language with rare inflectional morphology. It has four lexical tones and one neutral tone. Four lexical tones (Tone 1, Tone 2, Tone 3, and Tone 4) are represented with meaning-differentiating tonal contours: high-level, low-rising, low-dipping, and high-falling. Orthographically, tones are transcribed by adding diacritics to the main vowel of the respective syllable in *Hanyu Pinyin*, a transcription system of Mandarin Chinese used in Mainland China nationwide and in the international academic community, *i.e., shī* (teacher), *shí* (stone), *shǐ* (history), and *shì* (event), in the order of Tone 1, Tone 2, Tone 3, and Tone 4. The only tone sandhi rule in Mandarin Chinese predicts a surface tone, Tone 2, for Tone 3 if it is followed by another Tone 3. An eligible Mandarin Chinese syllable consists of a maximum number of four segments, an onset **C**, a glide **G**, a vowel **V**, and a nasal coda **N**. In compliance with pre-assumed phonological theories, the inventory may vary slightly (Duamu 2007, Lin 2007). In this paper, we use the same phoneme system for training acoustic models and transcribing spoken words. **C**, the onset position, can be occupied by one of six plosives /p, p^h, t, t^h, k, k^h/, six fricatives /f, s, ʂ, ɕ, x, ʐ/, six affricates /ts, ts^h, tʂ, $tʂ^h$, tɕ, $tɕ^h$/, two nasals /m, n/ and one lateral /l/, or it can be vacant. Only /n/ and /ŋ/ are allowed in the coda position **N. G** contains two glides /j, w/ and **V** contains 15 vowels including mono- and diphthongs /i, ɨ, ɯ, u, y, a, o, ə, e, ɚ, ai, ei, au, ou, ye/.

From the perspective of orthography, a tonal syllable in Mandarin Chinese and other varieties of Chinese normally corresponds to a character in the writing system, which is often also equivalent to a morpheme. In Yin's (1984) quantitative study of contemporary Mandarin Chinese, approximately 1,300 tonal syllables account for 5,000 morphemes that make use of only 400 distinctive syllable struc-

ture types (**CGVN**), leading to a large number of homophones. Compared to old Chinese, modern Mandarin Chinese has more disyllabic than monosyllabic words in both written and spoken language uses (Packard 2000, Institute of Language Teaching and Research 1986, Li 2007, Chen *et al.* 1996, Tseng 2019). The homonym avoidance principle and phonological and prosodic disyllabification constraints have been proposed in the literature to explain the increasing use of disyllabic words (Duanmu 1999, Arcodia 2007). Disyllabic words can be mono- or dimorphemic. For example, *pútáo* (the grapes) and *shòusī* (sushi) are monomorphemic words; *mǔ-jī* (female-chicken, hens) and *kǒng-lóng* (scary-dragon, dinosaurs) are dimorphemic, which are usually semantically transparent, as their meaning is composed of two component morphemes (Wang *et al.* 2019). Common word formation patterns in Mandarin Chinese include the subject-predicate compound *tóu-tòng* (head-pain, headache), resultative verb compound *dǎ-pò* (hit-broken), parallel verb compound *bāng-zhù* (help-help), modifier-head compound *pí-xíe* (leather-shoes), verb-object compound *pīn-tú* (spell-picture, jigsaw puzzle), *sǎo-dì* (sweep-floor), reduplication *tiān-tiān* (day-day, every day), prefixes *kě-ài* (-able-love, lovable) and suffixes *kuài-zi* (chopsticks-diminutive) (Yuan and Huang 1998, Duanmu 1999, Baxter and Sagart 1998, Packard 2000, Arcodia 2007, Li 2007, Li and Thompson, 2009, Huang *et al.* 2017, Yip 2000).

Moreover, the definition of "word" in Mandarin Chinese is not uniform. Given a text, word segmentation varies according to existing theories of word morphology. In this paper, we adopted the linguistic framework proposed by Huang *et al.* (2017) that served as the groundwork in the implementation of the CKIP (Chinese Knowledge and Information Processing) automatic word segmentation and part of speech tagging system at the Institute of Information Science, Academia Sinica. We applied the CKIP automatic system to process our texts with post-editing modifications added to the system outputs in order to accommodate conversation-specific constructions, e.g., directional complements. Previous works have proposed that lexical properties affect phonetic forms of spoken words at different linguistic levels, including phonological features, morphological structure, syntactic category, semantic meaning, and production frequency (Arcodia 2007, Amiot 2005, Jurafsky *et al.* 2001, Bybee and Hopper 2001, Levelt 1989). Some of these lexical properties will be examined in this paper with a focus on the phenomenon of phonetic fusion within disyllabic words. Quantitative evidence obtained by automatic text and speech processing as well as qualitative transcription of spoken words will be discussed to extend our understanding of potentially representative phonetic forms of spoken words.

1.2 Fusion and reduction

1.2.1 *Word fusion*

Coalescence normally refers to a phonological phenomenon, in which two words are merged into one, e.g., *don't* for "do not". Collocations associated with discourse contexts may affect the actually produced phonetic forms of coalescence words (Scheibman 2000). From the viewpoints of word formation and speech errors (Levelt 1989), portmanteaux or blends are formed by fusing two words or two morphemes that may occur at word-internal, compound word, or phrasal levels. Word blends can convey new concepts that are derived from the meaning of the original words, e.g., *brunch* for "breakfast-lunch" and *motel* for "motor" and "hotel", in which the initial part of the first word is merged with the final part of the second word. Alternatively, word blends also appear in speech errors, e.g., *lection* standing for *lecture-lesson* (Garret 1975, 138). Disyllabic word mergers in Chinese dialects employ a similar mechanism, in which phonological and orthographical properties participate in a different manner. Phonologically speaking, a disyllabic merger takes the onset from the first syllable and the rhyme from the second syllable, targeting a legitimate tonal syllable in the respective varieties of Chinese (He 2013, Sun 2014). The merger syllable theoretically should be accommodated with a tone complying with the rhyme-tone legitimacy. However, to our knowledge, no full-fledged rules in this regard have been proposed so far. Orthographically speaking, the newly resulting character can be a mixture of the original characters, representing a meaning derived from both morphemes, often semantically transparent to some extent. Unlike the nouns *motel* or *brunch*, lexicalized word blends of these kinds are frequently found at the syntactic level as well, e.g., in imperative uses, utterance-final particles, measure words, and demonstratives in the varieties of Chinese (Cui 1994, Xu 1999). Compound words that form disyllabic coalescence, but not lexicalized in terms of new characters, appear extensively in unrestricted word classes (Li 2013). Below are some examples of word blends from You (2018, 200), Cui (1994, 118), and the Dictionary of common words of Taiwanese Southern Min (Ministry of Education 2008).

勿fɤʔ⁴(do not)-要iæ⁵¹³(want)=>勨fiæ⁵¹³(don't, imperative) [Suzhou dialect]
不pʏ²¹³(negation)-要iɔ²¹³(want)=>耍pɔ²¹³ (don't, imperative) [Xining dialect]
不pu³⁵(negation)-用yoŋ⁵¹(use)=>甭pəŋ³⁵(not necessary) [Beijing dialect]
唔/不m²¹³(negation)-好ho⁵³(good)=>孬mo⁵³(not good enough) [Chaoyang dialect]
了le (particle)-啊a (discourse particle)=>啦la (discourse particle) [Beijing dialect]
昨昏 cha-hng=>chắng (yesterday-night=>last night) [Southern Min]

Similar to the mechanism of word coalescence and blends, traditional Chinese phonology utilizes the technique *fǎnqīe*, which takes the onset from the syllable of a known character and the rhyme from another to phonetically transcribe an unknown character. The spirit of *fǎnqīe* is close to that of the Edge-in theory, which explains the phonological process of syllable contraction by keeping the segments from the two edges (onset of the first syllable/rhyme of the second syllable) and then merging the inner segments (Yip 1988, Chung 1997). To predict contracted forms in Taiwanese Southern Min, Hsu (2003) further proposed a nuclei merging model that follows the sonority hierarchy a > ɔ > e > o > i > u. The cross-syllable boundary of a disyllable disappears, and the remaining nuclei from the two original syllables merge, similar to the Edge-in theory. The onset part of the first syllable and the rhyme part of the second syllable form the predicted contracted form, which ideally should be a phonologically legitimate syllable. *Fǎnqīe*, the Edge-in theory, and the nuclei merging sonority hierarchy share similar phonological mechanisms.

1.2.2 *Phonetic reduction*

Coalescence is considered a phonological phenomenon that predicts target forms from the scope of words (Chung 1997, Hsu 2003). It is possible that these target forms are found among a wide range of phonetically reduced pronunciation variants that are actually produced in continuous speech, accompanied by reduced duration and articulatory undershoot of vowel reduction or centralization. This interface between phonology and phonetics provides an interesting angle to look at language processing and production. Myers and Li (2009) examined the acoustic properties of Southern Min syllable contraction data and proposed that lexical frequency and degree of segmental reduction are correlated, but they found no evidence for a categorical alternation in fully uncontracted *versus* fully contracted forms. Similarly, Bürki *et al.* (2011) proposed that the schwa alternation in French does not always follow the phonologically predicted categorical outputs of schwa *versus* non-schwa variants, and this gradual, complex phonetic reduction did cause difficulties for listeners in judging the presence of schwa. Word-specific factors such as neighborhood density, morphological structure, word category and production frequency, context-specific factors such as previous mention and discourse function as well as speaker- and speaking style-specific factors have all been identified to play a role in the likelihood and actual production of phonetically reduced words in previous research (Bell *et al.* 2009, Gahl *et al.* 2012, Meunier and Espesser 2011, Clopper *et al.* 2017, Clopper *et al.* 2018, Levelt 1989).

For Mandarin Chinese, multisyllabic word reduction ranges from marginal segment omission and partial fusion to syllable merger (Tseng *et al.* 2013). Lee

(2018) reported that syllable merger forms are more quickly recognized than partial fusion forms and words with a high production frequency are more quickly recognized than those with a low frequency. Pluymaekers *et al.* (2005) reported that a high frequency of words in Dutch leads to a reduced duration of affixes. Suffixed forms of verbs and nouns in American English also show patterns of duration reduction in the stems (Cohen and Carlson 2016). But the relationship between predictability and phonetic prominence may not always hold. For instance, Kuperman *et al.* (2007) reported that the most predictable interfix in Dutch compounds is produced for a longer duration, a result counter to the common consensus that the more predictable a construction is, the more reduced it will be (Jaeger and Buz 2017). Apparently, multiple factors are involved in speech production, leading to varying spoken word forms. In this study, we use authentic data of disyllabic words from a conversational speech corpus to study the specific phenomenon of phonetic fusion concerning the following research questions. Do the completely fused words produced in realistic spoken data always correspond to phonologically predicted forms? How often are they used? Whether it is possible that there are more than one fusion forms for a word? If there do exist representative forms among the phonetic variants, it may be likely that these forms are phonologically represented in the mental lexicon. Methodologically speaking, three different approaches are conducted in our analysis of phonetic variants: (1) computational derivation of categorical reduction types, (2) analysis of duration and lexical frequency of reduced words, and (3) impressionistic phonemic transcription of common disyllabic words.

2. Background, data and methodology

2.1 Relevance of cross-syllable boundaries for phonetic fusion

The target form of coalescences or blends normally has the onset from the first syllable (σ1) and the rhyme from the second syllable (σ2), as predicted by the Edge-in theory (Yip 1988, Chung 1997). In the process of word fusion, the σ1 coda and the σ2 onset, if available, are subject to omission in the first place. Individual syllable structure plays a role, but it is too overwhelming a task to consider the finer details and all syllable combinations of disyllabic words. Therefore, we propose to focus on the presence of the word-internal syllable boundaries and non-vocalic segments at the boundary instead.

To obtain an overview of boundary types of disyllabic words in authentic use, two Taiwan Mandarin Chinese corpora: a text corpus containing 5 M words, the *Sinica Balanced Corpus* (hereafter *SBC*, Chen *et al.* 1996), and a conversational

corpus containing 500 K words, the *Taiwan Mandarin Conversational Corpus*[1] (hereafter *TMC*, Tseng 2019) are examined as shown in **Table 1**. Approximately 85% of syllable tokens, 92% by syllable type, have an onset consonant. This is a significantly large proportion. Due to the quasi-correspondence between syllables (phonology), characters (orthography), morphemes (morphology) and meaning (semantics) in Mandarin Chinese, the large number of non-zero onset syllables faithfully reflects the presence of onset consonants in favor of enhancing the meaning-differentiating function. In contrast, approximately 40% of syllable tokens, equivalent to 60% of syllable types, are closed syllables. As a whole, no conspicuous differences in phonological properties at the level of syllables are observed in written and spoken language uses.

Table 1. Syllable types in text and conversation

Syllables with a coda			Syllables with an onset		
	SBC	**TMC**		**SBC**	**TMC**
Tokens	43.5%	40.1%	Tokens	84.9%	84.1%
Types	59.7%	59.5%	Types	92.2%	92.6%

To further focus on segments at the cross-syllable boundary, four types of disyllabic words are categorized by the presence of σ1 coda (Coda) and σ2 onset (Onset), in which Zero denotes an empty σ1 coda or an empty σ2 onset, resulting in four subgroups: Coda#Onset, Coda#Zero, Zero#Onset, and Zero#Zero. For instance, Coda#Onset represents a disyllabic word in which both σ1 coda and σ2 onset are present, e.g., *kǒng-lóng* (dinosaur). A disyllabic word is grouped as Zero#Zero if both σ1 coda and σ2 onset are empty, e.g., *kě-ài* (lovable). **Figure 1** shows the distribution patterns of these four boundary types of disyllabic words in *SBC* and in an annotated subset of *TMC* that will be used for the analysis in the present study, the *Sinica MCDC8 Corpus*[2] (hereafter *MCDC*, 90 K words, Tseng 2019). Coda#Onset and Zero#Onset together make up the majority in both corpora. There are no significant differences in terms of boundary types in written and spoken language uses, either.

1. The *Sinica Chinese Spoken Wordlist* released by Academia Sinica (24T-1080124) reports lexical information about the contents of the *TMC*. Academic license is free of charge.

2. The *Sinica MCDC8* is released by Academia Sinica (24T-1031223). It is publicly distributed via the Association for Computational Linguistics and Chinese Language Processing.

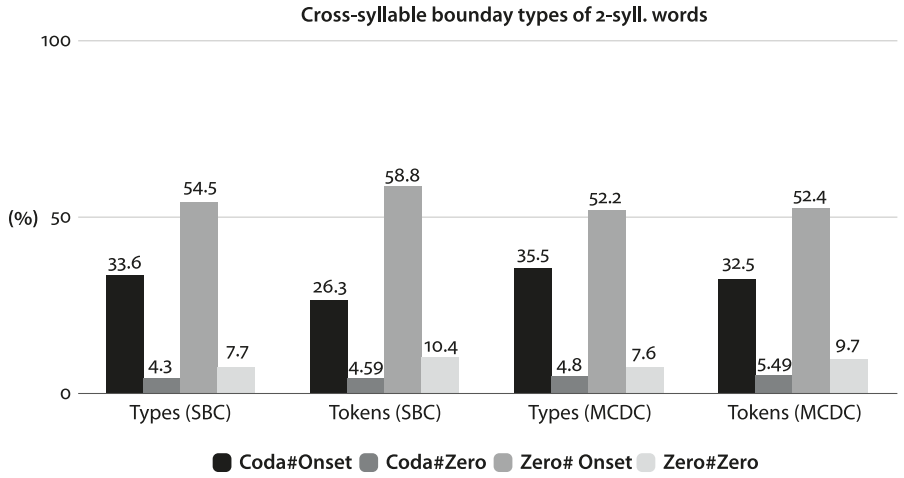

Figure 1. Types of disyllabic words in text and conversation

2.2 Categories of disyllabic contraction

Similar to phonetic reduction, fusion is a progressive process. We attempt to analyze reduced speech by categories that accordingly represent the gradient properties of the corresponding acoustic-phonetic characteristics. Instead of presenting the number of deleted segments, we adopted the concept of syllable contraction (Tseng 2005, List 2017) to represent the degree of phonetic reduction, with the extreme case of reduction to be the complete form of fusion. Four degrees of syllable contraction occurring in Chinese disyllables were proposed in terms of syllable number and nuclei merging (Tseng *et al.* 2013). The four-way contraction taxonomy was substantially transformed into disyllabic reduction types by Liu *et al.* (2016), including the citation form type (CAN), the marginal segment deletion type (MSD), the nucleus merger type (NUM) and the syllable merger type (SYM). CAN is produced with no omission of any existing cross-boundary nonvocalic segments, while MSD omits at least one of the existing cross-boundary nonvocalic segments. NUM loses all existing cross-boundary nonvocalic segments, and the nuclei start to merge, resulting in a blurred syllable boundary. In SYM, the complete fusion form, two syllables are merged into one. The overall system is shown in **Figure 2**. To take *zhīdào* (to know) [tʂɯ.tau] as an example, the most often produced phonetic form is an SYM [tʂau], a phonologically legitimate syllable in Mandarin Chinese. [tʂɯ.au] is an MSD that has only the plosive onset omitted. [tʂəau] is an NUM where the only one nonvocalic segment at the cross-syllable boundary is deleted and the two nuclei merge, but it is not yet a completely merged syllable.

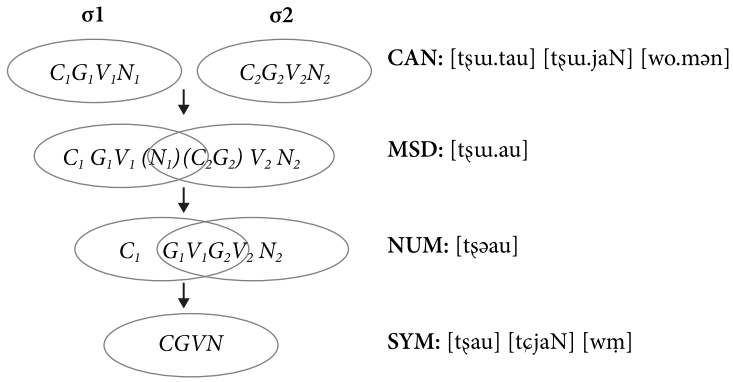

Figure 2. Reduction types towards fusion

The process of fusion may not always follow the same route, as the lexical properties of words are different. For instance, production frequency and word morphology are reported to affect the resulting spoken word forms (Pierrehumbert 2001, Bybee and Hopper 2001, Jurafsky *et al.* 2001, Tseng 2014, Lee 2018). When σ2 is a suffix or a complement, the process towards fusion may bypass MSD and NUM. The onset of the merger may remain $\mathbf{C}_1$ or be an alternative **C**. For instance, *zhè-yàng* (this way) [tʂɯ.jaN] is often phonetically produced as [tɕjaN] in *MCDC*. The legitimate syllable [tɕjaN] is used instead because the phonologically predicted merger [tʂjaN] violates the phonotactics of Mandarin Chinese. In social media, the word fusion [tɕjaN] is commonly written in a Chinese character that has the meaning of *sauce* (Tseng 2005). This lexicalization is stabilized, because readers would immediately recognize it as a merger representing *this way* rather than *sauce* from the context. Similarly, for *wǒ-men* (I-plural suffix, we) [wo.mən], the merger [wm̩] and the syllabic nasal [m̩] are not legitimate syllables either, but they are very frequently produced forms in *MCDC*. Both the bilabial coda in the merger and the syllabic nasal are phonetic cues that lead to immediate distinctiveness of *wǒ-men*, as they clearly violate the Mandarin Chinese phonotactics and therefore can be easily recognized. The merger form has retained the σ2 onset, the bilabial nasal, showing that the Edge-in principle does not always hold in spontaneous speech production.

2.3 From acoustic models, surface forms to categorical reduction types

The four reduction types mentioned above are automatically derived by comparing surface forms with citation forms. As shown in **Table 2**, the derivation algorithms are in principle developed by (1) whether the cross-syllable boundary is

present, (2) whether any of the cross-boundary nonvocalic segments is absent, and (3) whether the number of vowels is reduced to one (Liu *et al.* 2016).

Table 2. Criteria for deriving reduction types

Reduction type	Cross-syllable boundary	Nonvocalic segment omission	Number of nuclei
CAN	+	No	≥ 2
MSD	+	Yes	≥ 2
NUM	+/−	Yes	≥ 2
SYM	−	Yes	< 2

Surface forms are obtained automatically by applying the *ILAS phone aligner* (ILAS phone aligner 2020). It is an automatic system that assigns labels to the boundaries of phones aligned with the signal, given the text of speech content, by referring to pretrained acoustic models obtained from annotated spontaneous speech data (Tseng 2019). The phone set used for training acoustic models is identical to the aforementioned sound inventory of Mandarin Chinese (22 consonants, 2 glides, and 15 vowels). Applying the *ILAS phone aligner* but with no text information in the input, the output of the aligner, in which the acoustic properties of the phone sequences best match the input signal of a spoken word, is used as the surface form of the spoken word. Please note that surface forms are obtained solely from acoustic signals by referring to phonetic similarity scores that are computed by a system of features and salience settings of Mandarin Chinese proposed by Liu *et al.* (2016).

From *MCDC*, 30,264 tokens (3,806 types) of disyllabic words were individually segmented out of the original speech stretches and then processed by the *ILAS phone aligner*. Finally, freely recognized surface forms were compared to citation forms to derive reduction types. Please note that human-verified signal-aligned word boundary annotations in *MCDC* were utilized for extracting the speech signal of disyllabic word tokens. We thus ensure that the speech input used for acoustic processing and reduction type derivation contains adequate acoustic information of the words at issue. The remaining processing tasks are accomplished fully automatically with no human intervention. To take *suǒyǐ* (so) as an example, there are 476 tokens of *suǒyǐ* in *MCDC*. In total, 173 different sequences of phones are derived from acoustic processing; thus, 173 different surface forms are obtained. Reduction type derivation results in assignments of 434 SYM, 2 NUM, and 40 CAN. If we were to phonetically transcribe speech data of this amount, it would be challenging to achieve satisfactory levels of consistency within and among labelers.

2.4 Phonemically transcribing common disyllabic words

We also conducted a transcription experiment with the ten most commonly used disyllabic words in *MCDC* as listed in **Table 3**[3] to examine whether there are discrepancies between the physical properties of acoustics and perceptual judgments of spoken word forms. Two experienced phoneticians transcribed 6,320 word tokens independently. All four cross-syllable boundary types are represented in the wordlist. Among them, five are monomorphemic connectives. The other five dimorphemic words cover four word formation patterns. It is uncontroversial that *yīnwèi, ránhòu, suǒyǐ* and *qíshí* are monomorphemic and *méi-yǒu, wǒ-men* and *tā-men* are dimorphemic. For the others, the eligibility of negated forms by adding *bù* is tested. *Kě-bù-shì* means "you don't say" instead of "but not", while *jiù-bù-shì* is an eligible negation meaning "that is not". Thus, *kěshì* is regarded as an inseparable monomorpheme and *jiù-shì* as a dimorphemic word in this paper. For *júe-dé*, more than one reading is possible depending on whether it is followed by an adjective predicate or a sentence. As the negated form *júe-bù-júe-dé* is eligible, *júe-dé* is regarded as dimorphemic as well. A computer-aided panel of phoneme symbols of the Mandarin sound inventory was provided to the transcribers to phonemically transcribe the word tokens that were presented without contextual information. No other phonemes, suprasegmental properties or broad phonetic annotations were allowed in the transcription experiment.

Table 3. Words used in the phonemic transcription experiment

Type	Word	Citation form	Meaning	Word formation	Morph.	Count
Coda#Onset	*ránhòu*	ʐan.xou	then	Connective	1-mor.	731
Coda#Zero	*yīnwèi*	in.wei	because	Connective	1-mor.	649
Zero#Onset	*tā-men*	tʰa.mən	they	Pronoun-suffix	2-mor.	414
Zero#Onset	*wǒ-men*	wo.mən	we	Pronoun-suffix	2-mor.	877
Zero#Onset	*kěshì*	kʰə.ʂɨ	but	Connective	1-mor.	465
Zero#Onset	*jiù-shì*	tɕjou.ʂɨ	that is	Adverb-copula	2-mor.	1,039
Zero#Onset	*qíshí*	tɕʰi.ʂɨ	in fact	Connective	1-mor.	428
Zero#Onset	*júe-dé*	tɕye.tə	to think	Verb-resultative	2-mor.	675
Zero#Zero	*méi-yǒu*	mei.jou	don't have	Negation-verb	2-mor.	566
Zero#Zero	*suǒyǐ*	swo.i	so	Connective	1-mor.	476

3. *Kě* alone can also be used for *kěshì* (but) in Mainland China. However, it is not common in Taiwan.

3. Results and analysis

3.1 SYM is most often used, but CAN is most representative

We have defined four categories of reduction types (RT): the canonical form-like type CAN, the marginal segment deletion type MSD, the nucleus merger type NUM, and the syllable merger type SYM. From the perspective of phonetic reduction, the process of fusion theoretically ends with a complete merger of two syllables because no more segments can be omitted. **Figure 3a** shows that SYM represents the majority irrespective of boundary types. Though not necessarily appearing in the phonologically predicted form that is composed of σ1 onset and σ2 rhyme, all nonvocalic segments at the cross-syllable boundary are absent, and two nuclei merge into one. Regardless of the number of cross-boundary nonvocalic segments, SYM types are preferred. Over 85% of Coda#Zero words omit existing nonvocalic segments at the boundary, suggesting that it is highly likely for a σ1 coda to be omitted in the process of fusion. But it is not the case for σ2 onsets in Zero#Onset words. They tend to retain, possibly due to the meaning-differentiating function of onset consonants.

Syllable mergers are the most frequently produced form in conversational speech, accounting for approximately 44~69% of the overall words in each boundary type. However, this could be an effect of high-frequency words. To examine how individual word types are represented with reduction types, we defined a term called majority RT that signifies the particular reduction type that makes up the majority of reduction types assigned to a given word type. The results in **Figure 3b**, including disyllabic discourse markers, provide substantive pieces of empirical evidence. CAN and MSD types that are the closest to citation forms are still the majority of phonetic representations of spoken words in conversation, although spontaneous speech is expected to be greatly reduced. Please note that reduction types are derived automatically by acoustically obtained surface forms and algorithms of comparison to citation forms. Perceptual bias is scarcely possible.

Similar results were revealed by Johnson (2004, 39), who compared citation forms to phonetically transcribed words in American English conversational speech. Seventy-two percent of disyllabic function words maintain two syllables with no overt syllable deletion, and only 27% of them are reduced to one syllable. Our results suggest that citation form-like CAN and MSD together make up nearly 70% of word types. Twenty to forty percent of word types have the complete fusion form SYM as the majority RT. The only exception is the Coda#Zero type, which forms a minority of only 5% of the overall data. Citation form is normally acknowledged as the standard phonological representation of lexical

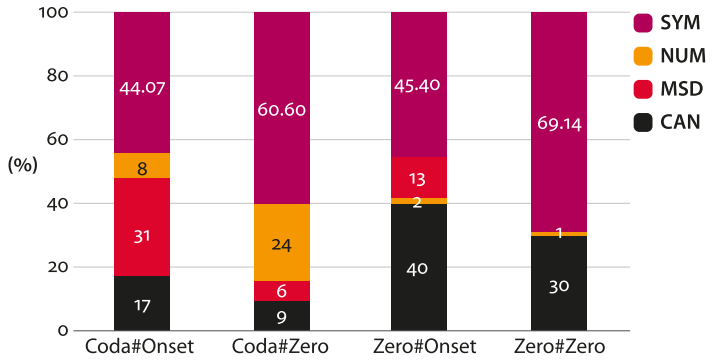

Figure 3a. RT: 30,264 disyllabic word tokens

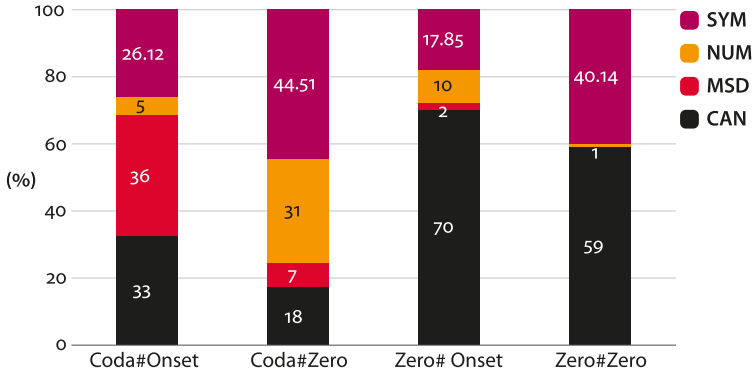

Figure 3b. Majority RT: 3,806 disyllabic word types

units, which is directly associated with canonical word meaning. As shown in our results, syllable mergers may also be considered as alternative representative forms due to their high occurrence frequency in overall RT and in majority RT assignments.

3.2 Duration and production frequency

The more frequently a word is spoken, the more likely the word is to be reduced (Cohen and Carlson 2016, Pluymaekers *et al.* 2005). Segmental reduction usually leads to a decrease in duration. This consensus seems to be uncontroversial in the literature (Dilts 2013). **Figure 4** presents the results of normalized duration and normalized similarity scores defined for obtaining surface forms of the 6,320 word tokens in **Table 3** with a statistical summary in **Table 4**. Phonologically predicted

disyllabic contraction forms are mergers of σ1 onset and σ2 rhyme (Chung 1997, Hsu 2003, Tseng 2005). Nonetheless, as we observed from the results for reduction types, phonetic fusion can be marginal by deleting the σ1 coda, σ2 onset or both, e.g., MSD. It can also be a complete fusion, e.g., SYM. The trend of the duration pattern across reduction types clearly follows the reduction categories we defined. The more reduced a word is, the shorter its duration will be.

We also observed that categorical reduction types do not always conform to phonetic similarity scores in **Figure 4**. The results of reduction types were derived from comparisons of phonological features relevant to syllable contraction, as listed in **Table 2**. Thus, it is not taken for granted that the degree of reduction always corresponds to the number of omitted segments. **Figure 4** shows that more reduced words in the sense of reduction types can also be produced long in some cases. These results show that contextual factors in interactive, spontaneous speech may affect the actual phonetic forms, e.g., prosodic position, syntactic construction or discourse function. Words that have an asymmetrical distribution of reduction type, phonetic similarity, and duration are candidates whose lexical representation and discourse function in conversation should be further investigated.

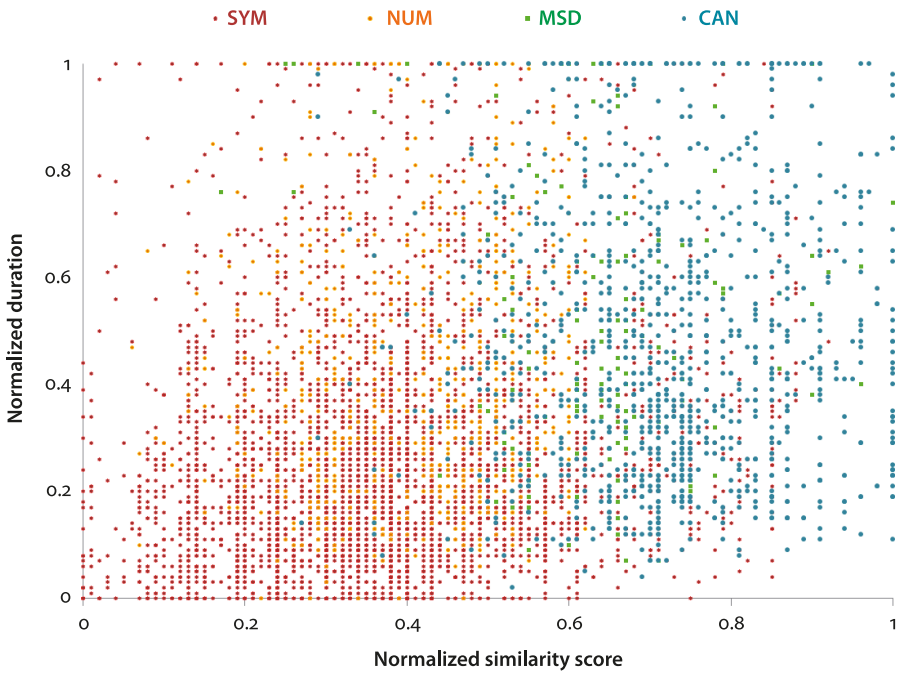

Figure 4. Correlation between the duration and phonetic similarity of classified RTs

Table 4. Statistics for duration (sec.) in four reduction types

RT	Count	Mean	SD
CAN	1,073	0.472	0.251
MSD	135	0.504	0.249
NUM	698	0.415	0.240
SYM	4,414	0.269	0.212

The results of majority RTs are shown by referring to the frequency information of 3,391 disyllabic word types in *SBC* and 3,640 in *TMC*. **Table 5** summarizes statistical information about corpus shares and reduction types. **Figures 5a** and **5b** show the mean values of corpus share percentages in both corpora by boundary type. Across boundary types, the preference for high-frequency words to be produced in the form of SYM remains the same regardless of the corpus type we refer to. The corpus share of word types produced with an SYM form is similar in both corpora, approximately 0.02%. However, shares of the other three reduction types differ. We mentioned earlier that the distribution patterns of the syllable structure and boundary types are similar in *SBC* and *TMC*, but when the issue involves phonetic phenomena such as reduction, the type of language resources used for obtaining reference information, e.g., lexical frequency, does matter.

Table 5. Corpus shares (%) of majority RT by word type in *SBC* and *TMC*

	SBC				TMC		
	Count	Mean	SD		Count	Mean	SD
CAN	1,829	0.0127	0.0253	CAN	1,988	0.0044	0.0128
MSD	468	0.0129	0.0232	MSD	492	0.0049	0.0122
NUM	295	0.0138	0.0217	NUM	308	0.0037	0.0071
SYM	799	0.0261	0.0566	SYM	852	0.0236	0.0953

3.3 Fusion inspected via acoustics and perception

In our transcription experiment, we compared the results of phoneme transcription with acoustically derived majority RT forms and phonologically predicted merger forms. The predicted forms of disyllabic words are primarily derived by applying the Edge-in theory (Yip 1988, Chung 1997), the nuclei merging sonority model (Hsu 2003) and trace principles (Tseng 2005) as used by Lee (2018). The Edge-in theory and the sonority model mainly focus on the phonological mechanism of merging the two original syllables from the outermost edges. Trace

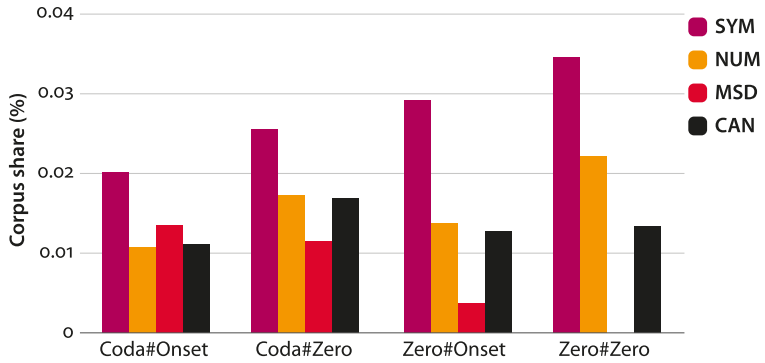

Figure 5a. Majority RT: Mean values of corpus share in *SBC*

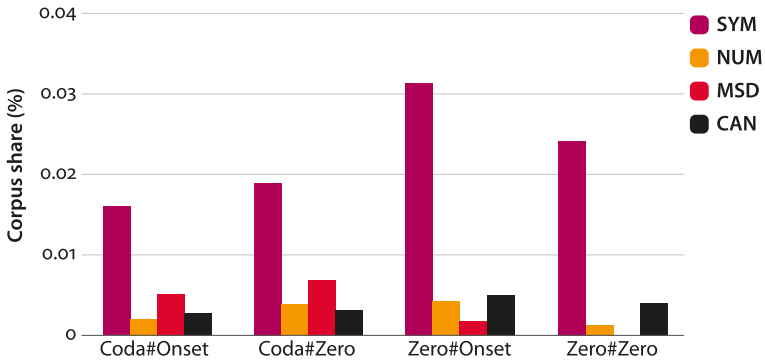

Figure 5b. Majority RT: Mean values of corpus share in *TMC*

principles that are listener-oriented and intelligibility-based suggest that phonetic properties representing distinctive phonological properties from each of the original syllables should be preserved in the produced forms to facilitate effective lexical retrieval in the merging process, e.g., prolongation of segments and retention of vocalic contrast (Tseng 2005).

In addition to phonologically predicted forms, **Table 6** also lists the majority RT of each word, including the percentages and the most representative surface form of the majority RT, derived by acoustic properties. The results of phonemic transcription, on the other hand, are based on perceptual judgments of the two transcribers. They had to choose phonemes from a restricted sound inventory as those provided by the user panel. The transcription results are rather diverse. The decision of selecting phonemes instead of phones to account for the phonetic details the transcribers heard varies individually, as the thresholds determined by the transcribers, which map a sound onto a phoneme or a word, may be different.

In the present study, we discuss only the most common phonemic transcription results that were agreed upon by both transcribers.

Table 6. Phonetic fusion forms obtained from the three methods

Pinyin	Citation form	Majority RT	Majority RT %	Phonetic form (RT)	Phonological prediction	Phonemic transcription
yīnwèi	in.wei	SYM	67%	ɀei	iwei	i
ránhòu	ɀan.xou	SYM	80%	tau	ɀau	naxou
suǒyǐ	swo.i	SYM	91%	tsei	swei	swei/sə
kěshì	kʰə.ʂɨ	SYM	61%	kʰɯ	kʰə:	kʰəsɨ
qíshí	tɕʰi.ʂɨ	SYM	54%	tɕʰy	tɕʰi:	tɕʰisɨ
jiù-shì	tɕjou.ʂɨ	SYM	50%	tɕɯ	tɕjə:	tɕjousɨ
méi-yǒu	mei.jou	SYM	67%	mə	mjə:	meijou
júe-dé	tɕye.tə	SYM	65%	tɕye	tɕyə	tɕye
wǒ-men	wo.mən	SYM	85%	ŋ	wən	om
tā-men	tʰa.mən	SYM	83%	tʰaŋ	tʰan	tʰam

Four words have their majority RT percentage higher than 80%: *ránhòu, suǒyǐ, wǒ-men*, and *tā-men* in **Figure 6**. These words are clearly subject to extreme reduction in spontaneous speech. *Ránhòu* and *suǒyǐ* are predicted to take two legitimate syllables [ɀau] and [swei] as the merger forms, ranking the 308th and the 193rd by production frequency, respectively, out of the 401 produced syllable types in *TMC*. The acoustically derived representative phonetic forms of majority RT are [tau] and [tsei], which are considerably close to the predicted phonological forms, taking into consideration that spontaneous speech is often reduced. Concerning perceptual judgments, the transcribers most often agreed on [naxou] and [swei/sə]. To be careful in interpreting this result, we only draw the conclusion that *suǒyǐ* seems more likely to be perceived as one syllable than *ránhòu*. This discrepancy may be due to the higher production frequency of [swei] compared to [ɀau] in spoken use, as those in *TMC*. Whether a fusion form (variation) will lead to a type of lexicalization in terms of language change, multiple factors need to be taken into consideration, e.g., the target syllable's production frequency, its distinctiveness in the spoken context, and its impact on the writing system etc.

On the other hand, phonological features do matter in the complete fusion forms, e.g., the nasal feature. Once the nasal feature is present in the production of *wǒ-men* and *tā-men* (first and third person), the clue for decoding a plural

suffix seems to be delivered explicitly enough. It is noteworthy that the acoustically derived forms in **Table 6** have velar features, but human transcription tends to perceive a bilabial nasal as it appears in the suffix morpheme. Both velar and bilabial variants differ from the phonologically predicted dental nasal coda. *Suǒyǐ, ránhòu, wǒ-men* and *tā-men* form different boundary types, so the number and type of cross-syllable segments may not be the most crucial factors for the final merger form. It is possible that the frequent use of the merger syllable such as [swei] for *suǒyǐ* as well as the easily accessible connection between word morphology and meaning such as the bilabial nasal in [tʰam] for *tā-men* are in fact decisive.

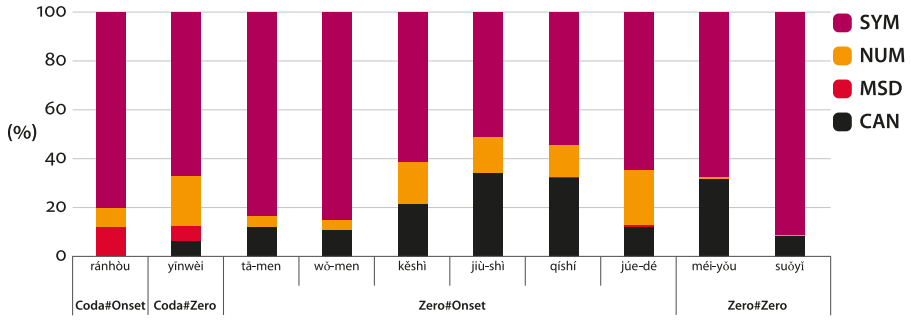

Figure 6. RT distribution of 10 common disyllabic words

Kěshì, qíshí and *jiù-shì* are Zero#Onset words with the same σ2, an unintended coincidence. For all three of them, only the citation forms were mostly agreed upon by the two transcribers. Although the acoustically derived representative phonetic forms are syllable mergers, compared with the other words, these three high-frequency words have a relatively low SYM assignment rate, between 50~60%. Their reduced forms vary to a great extent. The second syllable *shì*, the copula, tended to remain audible for the transcribers, although their acoustic salience may be different due to the RT proportions in **Figure 6**. The SYM assignment is the lowest for the dimorphemic word *jiù-shì*, suggesting that acoustic properties signaling the presence of the copula *shì* in the production of *jiù-shì* may be more conspicuous than those in the monomorphemic *kěshì* and *qíshí*. For the remaining *yīnwèi, méi-yǒu* and *júe-dé*, their SYM majority RT rates lie at approximately 65%. Based on our data, it is possible that there are more than one phonological representation forms for words, as the phonetic variants of the words need to comply with the associated discourse context. For instance, when *yīnwèi* refers to a cause-effect rhetorical relation, it is not as greatly reduced as the acoustically derived phonetic form [ɻei]. However, as a discourse juncture marker with no implication of a cause-effect relation, *yīnwèi* can also appear in an extremely reduced form, e.g., the transcription result [i]. For *júe-dé*, as a dimorphemic word,

the acoustically derived phonetic form is the same as the transcribed form, [tɕye], seemingly a syllable merger. However, we found that *júe-dé* is often produced with a prolonged vocalic segment [tɕye:] in our data, in which the prolongation stands for the resultative component *dé*. Our transcribers did not have the option to annotate suprasegmental properties such as prolongation or nasalization, and the automatic derivation process did not take into account any acoustic-prosodic features, either. We may further consider to implement prosodic properties of these kinds and those of the tonal contours in the revised version of the current approach.

4. Discussion

Signal-based acoustic features are gradient properties. In this paper, we propose that phonetic fusion can be represented by categorical linguistic properties. We have adopted an approach to assigning linguistic categories based on gradient acoustic properties derived from spoken words and have demonstrated that categorical reduction types defined by phonological properties of syllable contraction at the word level reflect differences in cross-syllable boundary type, duration, production frequency, and morphology. As stated in the literature, exposure to language use is correlated with the degree of reduction and forms of pronunciation variants (Myers and Li 2009, Jurafsky *et al.* 2001) and it has also proven to enhance access to lexical retrieval (Perruchet and Poulin-Charronnat 2012, Lorenz and Tizón-Couto 2019). In our study, the most essential finding is the properties of the complete form of fusion, syllable merger. The tendency of a preference for syllable mergers is the same in mono- and dimorphemic words. Our analysis by boundary types, composed of varying morphological structures and different informative extents of semantic transparency, has revealed that the complete phonetic fusion of disyllabic words is not exclusively reserved for any of the aforementioned linguistic properties. For example, not only high-frequency words, rarely used words also appear in a fused form, e.g., *huàzhí* (resolution) [xaus], *zǎo-rì* (in the near future) [tʂauN], and *chū-wài* (to go out) [tʂ^{h}wa]. Under certain circumstances, segmental features alone can serve as a discriminating cue as well, e.g., the nasal characteristics marking a suffixed plural pronoun, e.g., [wm̩] for *wǒ-men*. Unlike the examples of discourse particles and negation forms in Chinese dialects mentioned earlier in this paper, our results have also shown that word category does not seem to restrict occurrences of syllable mergers of disyllabic words in conversational speech. We have only considered disyllabic words in this study. But phonetic fusion is also observed across words, in particular in combination with function words. Syllable merger forms occurring above

the word level may have different representations. For instance, *yí-ge* in the nominal phrase *yí-ge-rén* (one-classifier-person) are pronounced in syllable merger forms /i:, iə, iʔ/ in *MCDC*. Grammatically speaking, when the classifier *ge* is omitted completely, *yí* is changed to an alternate tone *yì in yì-rén*. However, it is also possible in Beijing Mandarin that a phonetically reduced form retains the high-rising tone of *yí* in *yí-ge-rén*, resulting in the form *yí rér*. The phenomenon of phonetic fusion is closely related to phonology and grammatical construction in Mandarin Chinese. Further studies on these issues are needed to elaborate on the relationship between language production, variation and change.

Our human transcription results have pointed out discrepancies between objective physical signal information and subjective linguistic interpretation. The discrepancies may result from distinctive thresholds determined by the pretrained acoustic models and by the abstract phonological system of human listeners, given the same sound inventory. In addition, the effect of orthography in terms of Chinese characters and the related phoneme restoration effect may also play a role in the perception of fused words. We did not consider these effects in our transcription experiment. Nonetheless, it is uncontroversial that syllable mergers are a form often used by native speakers. A lexical decision experiment conducted by Lee (2018) showed that in addition to the canonical form CAN, the syllable merger form SYM is more easily accessible than phonetic forms that are only marginally reduced, such as MSD. Lexical retrieval and activation in conversational use may be triggered and enhanced by more experienced phonetic forms that are most likely encountered in everyday speech communication. With our results, we would like to propose that these familiar forms, the complete fusion form, may be considered as a candidate form that is accessible in speakers' mental lexicon (Levelt 1989). Zhou and Marslen-Wilson (1997) tested three hypotheses regarding how tone alternation concerning the Tone 3-Tone 3 sandhi tone rule is phonologically represented: the surface, canonical, and abstract representation views. The abstract representation view grants both surface and citation forms as phonological representation forms. Their results of the priming experiments did not consistently favor any one of the three hypotheses. Our study on a large-scale set of production data suggests that syllable mergers are a highly preferred surface form for disyllabic words, which can further be considered for testing the hypotheses. Different from the phonological mechanism that predicts a unique coalescence form of two original syllables (Chung 1997, Hsu 2003), according to our results, it is likely that the surface representation may not physically refer to specific phonetic forms, but rather as a category that comprises of variants of syllable mergers. A similar notion of the Edge-in theory, but more related to the memory of words, was proposed by Aitchison (2012) stating that we remember better the begin and the end of a word and the begin is more important than the end, also known as

"the bathtub effect". Based on the production data presented in this study, the phonetically derived syllable merger forms preserving mainly the outermost segments of disyllabic words may possibly also be stored in the mental lexicon as categorical phonological forms.

We would also like to emphasize the notion that syllable merger is a category, not a specific phonetic form. A similar notion has been proposed before. Phonetic forms of words may be predictable in the sense of categorical clusters. Johnson's model of perception by exemplars (1997) accounts for speaker, dialect, and contextual variability. It is proposed that a production-perception link is possible with the model. Our data analysis offers more supportive evidence for phonetic details, as those reflected in acoustic signals may be processed with a certain degree of flexibility that is allowed by centering on representative forms of spoken words. Reduction types are derived based on gradient acoustic information and categorical phonological comparison. Among a cluster of phonetic variants of a given word type, its majority RT can be regarded as the representative form, for instance the SYM type for most of the high-frequency words. Reduction in word duration and production frequency seems to accordingly change with the reduction types. The concept of majority RT is to some extent similar to that of the exemplar (Pierrehumbert 2001), but more from the level of words instead of phonemes. Syllable merger, as shown in the current study, may be considered a genuine category of phonetic variants of Chinese disyllabic word forms that are representative next to the category of citation forms. Please note that the orthography may affect the production and the perception of Chinese spoken words as well, as characters represent tonal syllables and may connect the meaning with the phonological form in certain ways. The effect of orthography on phonetic fusion may be further clarified by conducting psycholinguistic experiments that verify the relationship between character components and phonetic representation.

5. Conclusions

We have examined the qualitative properties of the phenomenon of phonetic fusion and have identified a number of factors that should be included in a complex model of phonetic fusion. Categorical reduction types offer an interface for studying the gradual process of phonetic fusion that is specific to conversational discourse. We would like to propose a possible link that connects the most often produced surface forms with the lexical representation of spoken words. Our analysis has clearly shown that the phonetic fusion of Chinese disyllabic words tends to end with a syllable merger form. The target syllable may be legitimate, such as [swei], or illegitimate, such as [t^{h}am]. Preschool children in reading

tasks have shown that their morphological awareness favors vocabulary knowledge (Chung and Hu 2007). Does vocabulary knowledge in speaking tasks also include prototypical fusion forms of disyllabic words? During the course of language acquisition, is phonetic fusion learned as a phonological rule, or are individual word forms learned independently? Longitudinal child speech data may provide empirical evidence for how reduced disyllabic words are produced and what role the acquisition of word fusion plays in the phonological development process of children. Perception experiments of spoken word recognition in terms of the three forms proposed in this study (phonologically predicted, acoustically derived and manually transcribed forms) may further explain the relationship between production data and the mental representation of phonetic fusion.

Acknowledgements

We gratefully acknowledge the constructive comments of the reviewers and the financial support from the Ministry of Science and Technology, under grant MOST 106-2410-H-001-045-MY2.

References

Aitchison, Jean. 2012. Words in the mind: An introduction to the mental lexicon. 4th Ed. Oxford, UK. Basil Blackwell Publishers.

Amiot, Dany. 2005. Between compounding and derivation: Elements of word-formation corresponding to prepositions. *Morphology and its demarcations: Selected papers from the 11th morphology meeting, Vienna, February 2004*, vol. 264, 183–213. John Benjamins Publishing. https://doi.org/10.1075/cilt.264.12ami

Arcodia, Giorgio Francesco. 2007. Chinese: A language of compound words. *Selected proceedings of the 5th Décembrettes: Morphology in Toulouse*, 79–90.

Baxter, William H., and Laurent Sagart. 1998. Word formation in old Chinese. *New approaches to Chinese word formation: Morphology, phonology and the lexicon in modern and ancient Chinese*, ed. by Jerome L. Packard, 35–76. Mouton de Gruyter. https://doi.org/10.1515/9783110809084.35

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand, and Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60.1.92–111. https://doi.org/10.1016/j.jml.2008.06.003

Bürki, Audrey, Cécile Fougeron, Cedric Gendrot, and Ulrich H. Frauenfelder. 2011. Phonetic reduction versus phonological deletion of French schwa: Some methodological issues. *Journal of Phonetics* 39.3.279–288. https://doi.org/10.1016/j.wocn.2010.07.003

Bybee, Joan L., and Paul J. Hopper (eds.). 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.45

Chen, Keh-Jiann, Chu-Ren Huang, Li-Ping Chang, and Hui-Li Hsu. 1996. Sinica Corpus: Design methodology for balanced corpora. *Proceedings of the Eleventh Pacific Asia Conference on Language, Information and Computation* 167–176.

Chung, Wei-Lun, and Chieh-Fang Hu. 2007. Morphological awareness and learning to read Chinese. *Reading and Writing* 20.5.441–461. https://doi.org/10.1007/s11145-006-9037-7

Clopper, Cynthia G., and Rory Turnbull. 2018. Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In *Rethinking reduction*, ed. by Francesco Cangemi, Meghan Clayards, Oliver Niebuhr, Barbara Schuppler, and Margaret Zellers, 25–72. Mouton de Gruyter. https://doi.org/10.1515/9783110524178-002

Clopper, Cynthia G., Jane F. Mitsch, and Terrin N. Tamati. 2017. Effects of phonetic reduction and regional dialect on vowel production. *Journal of Phonetics* 60.38–59. https://doi.org/10.1016/j.wocn.2016.11.002

Cohen, Clara, and Matt Carlson. 2016. Phonetic reduction can lead to lengthening, and enhancement can lead to shortening. *INTERSPEECH 2016*, 1094–1098. https://doi.org/10.21437/Interspeech.2016-1146

Chung, Raung-Fu. 1997. Syllable Contraction in Chinese. *Chinese languages and linguistics III, morphology and lexicon*, edited by Feng-Fu Tsao and Samuel H. Wang, 199–235. Symposium series of the Institute of History and Philology, Academia Sinica, Taipei

Cui, Li. 1994. On portmanteau words in Chinese. Zhengzhou daxue xuebao [Newsletter of Zhengzhou University] (Philosophy and Social Sciences Edition) 30.118–120. (In Chinese)

Dilts, Philip. 2013. Modelling phonetic reduction in a corpus of spoken English using random forests and mixed-effects regression. Edmonton: University of Alberta Unpublished Ph.D. dissertation.

Duanmu, San. 1999. Stress and the development of disyllabic words in Chinese. *Diachronica* 16.1.1–35. https://doi.org/10.1075/dia.16.1.03dua

Duanmu, San. 2007. *The phonology of standard Chinese.* 2nd Edition. Oxford University Press.

Gahl, Susanne, Yao Yao, and Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of memory and language* 66.4.789–806. https://doi.org/10.1016/j.jml.2011.11.006

Garrett, Merrill F. 1975. The analysis of sentence production. *Psychology of learning and motivation*, ed. by Gordon H. Bower, 133–177. Academic Press.

He, Fuling. 2013. On the pronunciations of "twenty", "thirty" and "forty" in Chinese Dialects. *Newsletter of Chinese Language* 91.95–106. (In Chinese)

Hsu, Hui-chuan. 2003. A Sonority Model of Syllable Contraction in Taiwanese Southern Min. *Journal of East Asian Linguistics* 12.349–377. https://doi.org/10.1023/A:1026108613211

Huang, Chu-Ren, Shu-Kai Hsieh, and Keh-Jiann Chen. 2017. *Mandarin Chinese words and parts of speech: A corpus-based study*. Taylor & Francis.

ILAS phone aligner. 2020. Institute of Linguistics, Academia Sinica. Accessed 20 December 2020 <http://aligner.ling.sinica.edu.tw/>.

Institute of Language Teaching and Research. 1986. *A frequency dictionary of Modern Chinese.* Beijing: Beijing Language Institute Press. (In Chinese)

Jaeger, T. Florian, and Esteban Buz. 2017. Signal reduction and linguistic encoding. *Handbook of psycholinguistics*, ed. by Eva M. Fernandez, and Helen Smith Cairns, 38–81. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781118829516.ch3

Johnson, Keith. 2004. Massive reduction in conversational American English. *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium*, 29–54. National Institute for Japanese Language and Linguistics, Tokyo.
Johnson, Keith. 1997. Speech perception without speaker normalization. *Talker variability in speech processing*, ed. by Keith Johnson and John W. Mullennix, 145–166. San Diego, Academic Press.
Jurafsky, Dan, Alan Bell, Michelle Gregory, and William D. Raymond. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. *Frequency and the emergence of linguistic structure*, ed. by Joan Bybee and Paul Hopper, 229–54. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.45.13jur
Kuperman, Victor, Mark Pluymaekers, Mirjam Ernestus, and Harald Baayen. 2007. Morphological predictability and acoustic duration of interfixes in Dutch compounds. *Journal of the Acoustical Society of America* 121.4.2261–2271. https://doi.org/10.1121/1.2537393
Lee, Chien-Wen. 2018. The effects of word frequency and syllable reduction degrees on recognizing Mandarin spoken disyllabic words for normal hearing adults. Department of Speech Language Pathology and Audiology, National Taipei University of Nursing and Health Science. Master thesis.
Levelt, Willem J.M. 1989. *Speaking: From intention to articulation*. MIT press.
Li, Charles N., and Sandra A. Thompson. 2009. Chinese. *The world's major languages*, ed. by Bernard Comrie, 703–723. Routledge.
Li, Chunxiao. 2013. On the syneresis of Southern Min dialect in Fujian and Taiwan. Zhongguo fangyan xuebao [Newsletter of Chinese dialects] 1.145–156. (In Chinese)
Li, Hui. 2007. A study of the characteristics of lexicalization of disyllabic phrases in contemporary Chinese. Yuyan jiaoxue yu yanjiu [Language teaching and research] 2.50–55. (In Chinese)
Lin, Yen-Hwei. 2007. *The sounds of Chinese*, with audio CD, vol. 1. Cambridge University Press.
Liu, Yi-Fen, Shu-Chuan Tseng, and Roger Jang. 2016. Deriving disyllabic word variants from a Chinese conversational speech corpus. *Journal of the Acoustical Society of America* 140.1.308–321. https://doi.org/10.1121/1.4954745
List, Johann-Mattis. 2017. Contraction. *Encyclopedia of Chinese language and linguistics*, ed. by Rint Sybesma, 672–675. Brill.
Lorenz, David, and David Tizón-Couto. 2019. Chunking or predicting-frequency information and reduction in the perception of multi-word sequences. *Cognitive Linguistics* 30.4.751–784. https://doi.org/10.1515/cog-2017-0138
Meunier, Christine, and Robert Espesser. 2011. Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics* 39.3.271–278. https://doi.org/10.1016/j.wocn.2010.11.008
Ministry of Education. 2008. *Dictionary of common words of Taiwanese Southern Min*. https://twblg.dict.edu.tw/
Myers, James, and Yingshing Li. 2009. Lexical frequency effects in Taiwan Southern Min syllable contraction. *Journal of Phonetics* 37.2.212–230. https://doi.org/10.1016/j.wocn.2009.02.002
Packard, Jerome L. 2000. *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge University Press. https://doi.org/10.1017/CBO9780511486821

Perruchet, Pierre, and Bénédicte Poulin-Charronnat. 2012. Beyond transitional probability computations: Extracting word-like units when only statistical information is available. *Journal of Memory and Language* 66.4.807–818. https://doi.org/10.1016/j.jml.2012.02.010

Pierrehumbert, Janet. B. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. *Frequency effects and emergent grammar*, ed. by Joan Bybee and Paul Hopper, 137–157. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.45.08pie

Pluymaekers, Mark, Mirjam Ernestus, and Harald Baayen. 2005. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118.4.2561–2569. https://doi.org/10.1121/1.2011150

Scheibman, Joanne. 2000. I *dunno*: A usage-based account of the phonological reduction of don't in American English conversation. *Journal of pragmatics* 32.1.105–124. https://doi.org/10.1016/S0378-2166(99)00032-6

Sun, Hongju. 2014. A study of coalescence in Chinese. *Journal of Southwest University* (Social Sciences Edition) 40.1.115–124. (In Chinese)

Tseng, Shu-Chuan. 2005. Monosyllabic word merger in Mandarin. *Language Variation and Change* 17.3.231–256. https://doi.org/10.1017/S0954394505050143

Tseng, Shu-Chuan. 2014. Chinese disyllabic words in conversation. *Chinese Language and Discourse* 5.2.231–251. https://doi.org/10.1075/cld.5.2.05tse

Tseng, Shu-Chuan. 2019. ILAS Chinese spoken language resources. *Proceedings of LPSS 2019-the third International Symposium on Linguistic Patterns in Spontaneous Speech*, 13–20.

Tseng, Shu-Chuan, Alexander Soemer, and Tzu-Lun Lee. 2013. Tones of reduced T1-T4 Mandarin disyllables. *International Journal of Computational Linguistics and Chinese Language Processing* 18.3.81–105.

Wang, Shichang, Chu-Ren Huang, Yao Yao, and Angel Chan. 2019. The effect of morphological structure on semantic transparency ratings. *Language and Linguistics* 20.2.225–255.

Xu, Bo. 1999. An analysis of disyllabic words in Ningbo dialect. *Journal of Zhejing Ocean University* 16.4.46–50. (In Chinese).

Yin, Bingyong. 1984. Quantitative analysis of Chinese morphemes. Zhongguo yuwen [Chinese language and literature] 5.1.338–347. (In Chinese)

Yip, Moira. 1988. Template morphology and the direction of association. *Natural Language & Linguistic Theory* 6.4.551–577. https://doi.org/10.1007/BF00134493

Yip, Po-Ching. 2000. *The Chinese lexicon: A comprehensive survey*. Psychology Press.

You, Rujie. 2018. *An introduction to Chinese dialectology*. Shanghai: Shanghai Educational Publishing House. (In Chinese)

Yuan, Chunfa, and Changning Huangg. 1998. A study on Chinese morphemes and morphology based on a morpheme database. Yuan wenzi yingyong [Applied Linguistics] 3.86–91. (In Chinese)

Zhou, Xiaolin, and William Marslen-Wilson. 1997. The abstractness of phonological representation in the Chinese mental lexicon. *Cognitive Processing of Chinese and other Asian languages* 3–26.

Address for correspondence

Shu-Chuan Tseng
Institute of Linguistics
Academia Sinica
Nankang
115 Taipei
Taiwan
tsengsc@gate.sinica.edu.tw

Publication history

Date received: 6 January 2021
Date accepted: 16 July 2021
Published online: 14 September 2021