Wordhood and Disyllabicity in Chinese

James Myers
National Chung Cheng University

1 Introduction

Words pose a theoretical challenge in Chinese (i.e. Sinitic languages like Mandarin), but words pose a challenge in any language. Even though Chinese is written with monosyllabic, monomorphemic characters and no overt word boundaries, there is as much evidence here as there is in English or any other language for a level between the morpheme and the phrase, interfacing between the lexicon and the grammar. Yet their interface role makes words dynamic things, subject to distinct and often conflicting constraints from processing, semantics, phonology, morphology, and syntax. To emphasize the universality of this situation, I start in section 2 with a quick look at the dynamic nature of English words, then turn in section 3 to Chinese words, which a wide variety of data reveal as surprisingly English-like, including a strong preference for disyllabicity. In section 4 I sketch out a formalism that may help capture the universal yet dynamic nature of wordhood, trying it out on some of the Chinese facts. Section 5 gives some brief conclusions.

2 Wordhood in English

Chao (1968: 136) famously wrote that "[n]ot every language has a kind of unit which behaves in most (not to speak of all) respects as does the unit called 'word' when we talk or write in English about the subunits of English." While he is right that the universality of words cannot simply be assumed, he also seems to give the impression that the status of English words is obvious. As this section reviews, it is not.

Linguists have long recognized that words are tricky things. Bloomfield (1926) proposed that a word is a "minimum free form", but it is easy to think of words that cannot form utterances on their own (*the*, *to*, *know*, *cat*). The morphology textbook of Matthews (1991) manages to postpone the question "What are words?" until p. 208, and Haspelmath (2011) even questions whether words can be defined at all.

The orthography-based definition of English folk linguistics seems clear enough: a word is a string of letters surrounded by spaces or punctuation (not counting apostrophes or hyphens). However, word spacing intuitions vary; there are many compounds about which even copyeditors (or copy editors?) cannot agree. More importantly, orthographic spacing does not consistently correspond with other wordhood tests. To cite a textbook example, *white house* and *White House* are both written with internal spaces, but the first is argued to be a phrase because it is semantically compositional and has phrase-final stress, while the latter is argued to be a word because it has noncompositional semantics and compound-initial stress.

However, the arguments for the wordhood of *White House* are themselves less than fully conclusive. After all, phrases also seem capable of having noncompositional semantics, most obviously in idioms like *kick the bucket* (die) and collocations like the so-called "phrasal" verb *blow up* (cause to explode; cf. *blow something up*). Moreover, while *White House* is stressed as a phonological word, a prosodic unit containing one main stress, phonological words may contain syntactically free clitics (*Jack'll = Jack will* is pronounced like *jackal*). English words

do tend to be disyllabic like *White House*, and most of these are trochees (left-headed feet): the median number of syllables in the 133,852 words in the CMU Pronouncing Dictionary (Lenzo 2014) is two, with 74% of its 61,493 disyllabic words stressed only on the first syllable. Yet clearly there are many English words either shorter or longer than this.

English wordhood also receives ambiguous support from language processing. While word frequency effects (Monsell 1991) suggest that words are treated as wholes at some stage of processing, they are also observed for larger and smaller constituents: common phrases like *don't have to worry* are responded to more quickly than rarer ones like *don't have to wait* (Arnon and Snider 2010), and responses to words containing common morphemes are generally faster as well (Taft 1979).

For syntacticians, a constituent is a word if and only if syntax cannot manipulate or otherwise refer to its internal components, a principle known as the Lexicalist Hypothesis or the Lexical Integrity Hypothesis (Chomsky 1970, Bruening 2018, Huang 1984). However, the Lexicalist Hypothesis is not cut-and-dried either. An *atomic scientist*, for example, is someone who works on *atomic science*, not a scientist who is atomic; the derivational suffix *-ist* is somehow attached to an adjective + noun phrase (Spencer 1988: 663). Naive speakers are also confused over how to apply derivational affixes to phrasal verbs, leading to nonstandard forms like *blower-upper* (Bauer 1983: 71). Compounds can incorporate phrases too: if one *thinks outside the box*, one is an *outside-the-box thinker* (see Bruening 2018 for similar examples). Not only can *White House* and *Blue House* (the official residence of the South Korean president) be coordinated within a compound, as in *the White and Blue Houses*,[1] but so can affixes, as in *pro- and anti-democracy* (Duanmu 1998:139; see also Bruening 2018). Deverbal nouns may even preserve the syntactic behavior of their roots: *destroy* is causative, permitting a thematic object in *John's destruction of the city*, but *grow* is not, forbidding **John's growth of the tomatoes* (Marantz 1997: 216).

At the very least, harmonizing such evidence with the Lexicalist Hypothesis must acknowledge that the morphology/syntax distinction is not as simple as it first appears (as is done in Li 2005, Müller 2018, Newmeyer 2009). Since we have to make caveats anyway, I believe that we should also respect the semantic, phonological, morphological and processing complexities as providing important insights into the nature of words as well.

I took the time here to deconstruct English words because they "obviously" exist, so in the next section I will work to argue that Chinese words, which some believe do not exist, actually do. My rhetorical goal is to have the two languages meet somewhere in the middle, revealing less a typological contrast than the universal dynamism of wordhood itself.

3 Wordhood in Chinese

As in English, the Chinese folk-linguistic "word" (the "sociological word" of Chao 1968: 136) is orthographic, but a Chinese character is more like the linguist's morpheme. This has led some linguists to argue that Chinese has no English-like words at all (see Huang et al., this volume). In actual fact, however, words in Chinese have essentially the same nature as in English: dynamic, yes, but with universal word-like features. In arguing this I build on previous reviews of the Chinese wordhood question, including Chao (1968), Duanmu (1998, 2017), Packard (2000), Li (2005), Xu (2018), and the many works they cite (see also Li, this volume).

3.1 Some basic facts

[1] Source: https://www.theguardian.com/commentisfree/2016/mar/23/north-korea-rhetoric-pyongyang-us-china-nuclear-destruction. Accessed 8 May 2020.

Given that syllables organize articulatory gestures and leave clear acoustic traces, the monosyllabicity of virtually all Chinese morphemes makes them highly salient. This may be responsible for other typical Chinese features, like its preference for compounding over affixation (root morphemes tend to retain some stress and thus lexical syllable form) and its morpheme-based writing system. Note that monosyllabicity drives orthography and not the other way around: Vietnamese and Thai also have monosyllabic morphemes and a paucity of affixation, but their orthographies are phoneme-based and not morpheme-based.

Nevertheless, for over a century Chinese linguists have recognized words, christening them with the repurposed term 词 *ci* (Zádrapa 2017). Disyllabic words in fact appeared very early on (Duanmu 1999, Feng 1998), including a smattering of disyllabic morphemes (DeFrancis 1984: 183). While traditional Chinese dictionaries have separate entries for each character, word-based dictionaries are common, as are word-segmented corpora. Using such sources, it is often noted that the type frequency for multi-character words, that is, the number of distinct lexical items, is far higher than for one-character words (Yip 2000). Token frequency, or the number of times a lexical item appears in a corpus, is still highest for one-character words, but this is merely the consequence of Zipf's law of abbreviation (Zipf 1935); in particular, function words universally tend to be both frequent and short. In my own calculations using the Academia Sinica Balanced Corpus of Modern Chinese (Sinica Corpus, version 4.0, with around ten million word tokens, around half a million transcribed from speech; Huang et al. 1997), the overall mean word length is 1.6 characters.

A further sign that Chinese words are truly comparable to those in other languages comes from analyzing a parallel corpus compiled by Ziemski et al. (2016) of 8,000 translation-equivalent sentences randomly selected from United Nations documents in the six official UN languages: Arabic, Chinese, English, French, Russian, and Spanish. All but Chinese are written with word boundaries, making it easy to compare the number of characters per Chinese sentence with the number of orthographic words per matching sentence in the other five languages. When the number of English words is predicted from the number of words in each of the other languages with orthographic word marking, the slopes (B) of the best-fit lines are all close to one, implying roughly one-to-one correspondences (Arabic: $B = 1.06$; French: $B = 0.82$; Russian: $B = 1.02$; Spanish: $B = 0.80$). However, when the number of Chinese characters is predicted from the number of words in each of the other five languages, the slopes hover around 1.5 (Arabic: $B = 1.84$; English: $B = 1.63$; French: $B = 1.38$; Russian: $B = 1.75$; Spanish: $B = 1.34$), which happens to be the mean Chinese word length noted in the previous paragraph. This suggests that Chinese words reflect concepts similar in size to those expressed by orthographic words in a variety of other languages.

3.2 Psychological tests

Before considering Chinese words as grammatical objects, let us first see what we can learn about them from native-speaker intuitions (3.2.1) and language processing experiments (3.2.2).

3.2.1 Intuitions

Despite their morpheme-based writing system, Chinese speakers have surprisingly clear wordhood intuitions. Striking evidence for this comes from the bilingual Chinese/English title on an old tourist map of the Lion's Head Hill Scenic Area in Hsinchu, Taiwan.[2]. In this bit of anonymous ephemera (perhaps from the 1980s), the Chinese title is written horizontally as in (1a) (without the numbered brackets, of course). However, beneath it on the map is the bizarre English "translation" in (1b).

[2] Source: https://0.share.photo.xuite.net/dbfish66/10c1b52/15612621/836163205_m.jpg. Accessed 9 May 2020.

(1) a. [台省]$_1$ [名勝]$_2$ [獅頭山]$_3$ [遊覽圖]$_4$
Taisheng__mingsheng__shitoushan__youlantu
Taiwan-province__famous-site__Lion-Head-Hill__sightseeing-map
Sightseeing map for Lion's Head Hill, a scenic spot of Taiwan Province
b. [WANDER PERUSAL FIGURE]$_4$ [LION HEAD HILL]$_3$ [TITLE WIN]$_2$ [STAGE PROVINCE]$_1$

Apparently the translator started with the traditional assumption that each Chinese character corresponds to one word, translating 台省 *Taisheng* 'Taiwan Province' as 'stage province', and so on. The translator also had the idiosyncratic belief that English order is the reverse of Chinese. Yet the translator obeyed a crucial third principle as well: characters combine to form words. This is indicated by the numbered brackets in (1), showing that the translator segmented the title into right-headed nominal compounds, just as I did in my English gloss in (1a).

Wordhood intuitions are also reflected in the word segmentations in the Sinica Corpus. Despite the corpus creators' implementation of strict conventions (Huang et al. 2017), splitting the corpus by punctuation reveals a small number of character strings that are segmented differently in different places. Typical examples are shown in (2) (‖ marks word segmentations).

(2) a. 伴我成長 ~ 伴 ‖ 我 ‖ 成長
ban__wo__chengzhang
with__me__grow-up
grow up with me
b. 不要 ~ 不 ‖ 要
bu__yao
not__want
do not want

These variations are not random, however. Example (2a) shows a proper name (a song title) treated sometimes as a whole and sometimes as a syntactic phrase, while (2b) shows the optional cliticization of a function morpheme (other variably segmented examples like this in the corpus involve 就 *jiu* 'thus', the modifier marker 的 *de*, and the particle 了 *le*).

Word segmentation intuitions have also been explored experimentally. Hoosain (1992) and Lin et al. (2011) both claimed to find much inconsistency across readers, but their results actually suggest no more than the systematic variation seen in the Sinica Corpus. For example, Hoosain observed readers treating 就是 *jiushi* 'thus is' as a whole, even though Chao (1968) considers it syntactically separable (as in 就一定是 *jiu yiding shi* 'thus definitely is'). His readers' tendency to cliticize *jiu* to *shi* is nevertheless no more "wrong" than it would be for an English student to write *it is* as *it's*.

More recently, Wang et al. (2017) had over 1,000 readers segment out 152 target words, each in a separate sentence; despite the very large data set, they found virtually no variation at all. Admittedly this was after setting aside all participants who, in at least one sentence, separated every character from each other, or segmented out at least one string over seven characters long, and the target words were also all two-character nominal compounds, eliminating the difficulties raised by cliticization. Nevertheless, Chinese readers really do seem to split text into words quite consistently.

3.2.2 Language processing

Words have also been observed in Chinese psycholinguistic experiments. Word-driven models of listening comprehension are intrinsically more plausible than morpheme-driven ones for Chinese, simply because whole words have far fewer homophones than do individual morphemes (Packard 1999). Morpheme homophony is likely also why higher syllable frequency slows down the recognition of isolated bimorphemic spoken words (Zhou and Marslen-Wilson 1994). Evidence that Chinese listeners segment whole words from fluent speech comes from Ding et al. (2016), who found that brain waves track disyllabic/bimorphemic constituents when listening to simple noun-verb Chinese sentences, though they did not specifically test if these constituents were words and not prosodic or syntactic phrases. Zou et al. (2019) observed distinct brain activation patterns when listening to bimorphemic word pairs that shared whole-word meanings versus those that shared meaning in just one morpheme.

Readers also process words as units: as in English, the most robust finding in Chinese psycholinguistics is the facilitative effect of whole-word frequency (Myers 2017). Readers also recognize characters more readily if they are embedded in real bimorphemic words than in fake ones (Mok 2009), and text reading times are slowed if characters are separated at places other than word boundaries (Bai et al. 2008). Polymorphemic words can exert an indirect influence too; Li et al. (2017) is just one of many studies reporting slower recognition times for two-character words that have many lexical neighbors differing from the target in just one character.

As in English, Chinese also shows frequency effects in constituents both larger and smaller than words. Liu (2015) found that the frequency of idioms affected their acceptability and learning, and Myers et al. (2006) found that readers were faster to respond to a two-character verb followed by the durative aspect morpheme 著 *zhe* the more frequent the whole three-character string. Word reading is also undeniably affected by character-level processes. As reviewed in Myers (2017), while common characters usually speed word recognition, they slow responses when character and word meanings conflict, and rarer characters can also raise cross-character transition probability, speeding word recognition via enhanced internal cohesiveness.

In a particularly data-rich study, Li et al. (2014) found that eye movements in Chinese reading are influenced by word length, frequency, and contextual predictability, emphasizing that this is just as in English reading. Even the character-level effects that they observed, like character frequency and visual complexity, influenced eye movements only indirectly, by affecting the detection of words in upcoming text.

Establishing what Packard (2000: 13-14) calls “psycholinguistic words” is not enough, however, to show that such entities have more than a fleeting existence in the course of carrying out specific processing tasks. They also need not correspond precisely to words as reflected in semantics, phonology, morphology, or syntax. We must therefore consider evidence from these domains as well.

3.3 Semantic tests

Just as in English, idiosyncratic meanings can be associated not just with morphemes but also with morpheme strings, as illustrated by the oft-cited examples in (3). Safflowers are indeed red flowers, but they may also be yellow, and anyway all other types of red flowers have their own names; *chifan*, literally ‘eat rice’, includes the eating of noodles.

(3) a. 红花
hong__hua
red__flower
safflower
b. 吃饭
chi__fan

eat__rice
dine

As noted by Chao (1968), Duanmu (1998), Packard (2000) and others, idioms also have idiosyncractic semantics in Chinese, as in any language. To take an arbitrary example, 对牛弹琴 *duiniutanqin* literally means 'play a qin (a stringed instrument) to a cow', but figuratively means 'speak to somebody who does not understand.' Speakers thus have to memorize that it does not have some other figurative meaning, such as 'soothe an angry person with kind words.' Unsurprisingly, then, idioms are often listed in dictionaries and segmented in corpora as if they were indeed a species of word.

3.4 Phonological tests

As in English, in Chinese the prototypical word is disyllabic. This is hardly a typologically rare phenomenon: Garrett (1999) lists a variety of unrelated languages that require words to be at least disyllabic and Gordon (2002) lists many more with trochaic feet. Table 1 confirms that the most common word size is disyllabic throughout the Sinica Corpus, but given that this is a phonological property, the preference is stronger in the spoken portion.

	Number of syllables			
	1	2	3	4
Speech	.09	.64	.20	.07
Writing	.02	.47	.39	.11

Table 1. Proportional type frequencies of word sizes up to four syllables in the Sinica Corpus

Disyllabicity is also the most productive Chinese word size, but only in speech. Table 2 demonstrates this using a measure proposed by Baayen and Renouf (1996) that estimates potential coinage in a word class as the proportion of words with a token frequency of one (so-called hapax legomena). As shown in the table, disyllables are the potentially most readily coined in spoken Mandarin; Myers and Tsay (2015) report a similar disyllabic productivity bias in the CCU Taiwanese Spoken Corpus (Ruan et al. 2012). In written Mandarin, however, three-character words are actually more productive. The same surprising fact is seen in the estimated growth curves in Figure 1 that project the number of new words (types) per class expected in ever larger random samples of tokens (via Generalized Inverse Gauss-Poisson modeling, following Evert and Baroni 2007): only in speech is the slope steepest for disyllables. Apparently, in writing, where the influence of prosody is minimized, longer words are preferentially coined in order to express more nuanced lexical distinctions and to cope with polysyllabic foreign proper names.

	Number of syllables			
	1	2	3	4
Speech	.04	.58	.27	.11
Writing	.01	.38	.47	.14

Table 2. Proportions of hapax legomena for different word sizes in the Sinica Corpus

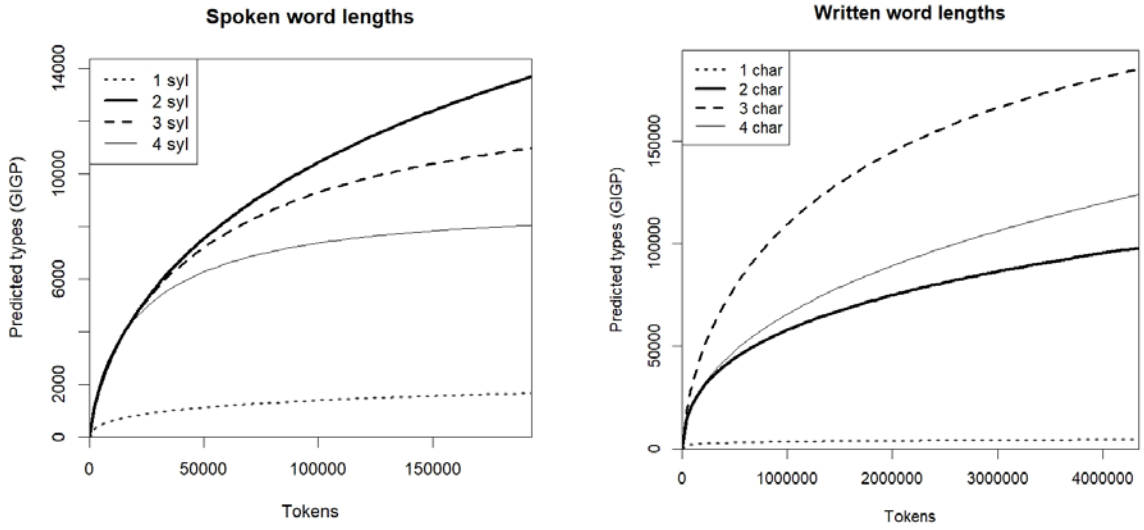

Figure 1. Projected growth curves for different word sizes in the Sinica Corpus

In Mandarin, at least, disyllabicity shares another property with English: trochaic stress (Duanmu 2012, Duanmu, this volume, though cf. Dell 2004). This is particularly audible in northern varieties, where words like that in (4a) end in a stressless syllable (i.e. neutral tone), but even in other varieties the stress is clearly stronger on the first syllable in disyllabic reduplicated forms like (4b).

(4) a. 先生
xiansheng
mister
b. 谢谢
xiexie
thanks

Disyllabicity helps shape longer Chinese words as well. Reduplication itself is generally disyllabic (e.g. 高兴 *gaoxing* 'happy' → 高高兴兴 *gaogaoxingxing* 'very happy') and the most common idiom form is quadrasyllabic, compounded of two disyllabic chunks (Liu 2015), as in the aforementioned *duiniutanqin* (for-cow play-qin). When forming nominal compounds from so-called elastic (Duanmu 2017) or telescopic (Huang et al. 2017) words, which come in disyllabic and monosyllabic variants, a monosyllable is disfavored at the left edge, as illustrated in (5). Qin and Duanmu (2017) confirmed this prosodic pattern in experimentally elicited native speaker judgments on novel compounds, taking into account factors like component frequency and compound interpretability.

(5) a. 技术工
ji-shu__gong
skill-technique__work
skilled worker
b. *技工人
ji__gong-ren
skill__work-person
skilled worker

Curiously, Qin and Duanmu (2017, 2019) also reported poor acceptability for novel two-character compounds themselves, but this finding may actually support the disyllabic word bias indirectly: singleton characters may not have sufficiently defined lexical representations to create interpretable compounds, and two-character strings also have more competing lexical neighbors (existing disyllabic words) than longer ones.

By the way, prosodically motivated elasticity is yet another Chinese-typical phenomenon that is also attested in English, as seen in the expansion/contraction of words into bimoraic feet (two light syllables or one heavy one) in *John~Johnny*, *Jennifer~Jenny*, *family~fam'ly*, *laboratory~lab*, *telephone~phone*, *refrigerator~fridge*. Unlike the case in English, elasticity in Chinese interacts with morphology and syntax, as in the so-called A-not-A construction illustrated in (6a) (see Packard 2000 for the morphology and Hagstrom 2006 for the syntax). Despite these interactions, however, the operation itself seems to be phonological reduplication (Huang 1991; see Inkelas 2008 for more on this concept), and indeed, in jocular contexts monomorphemic English trochees can undergo A-not-A splitting as well, as in (6b).

(6) a. 你高不高兴？
ni__gao__bu__gao-xing
you__high__not__high-excited
Are you happy?
b. 你 hap-不-happy？
ni__hap__bu__happy
you__hap__not__happy
Are you happy?

Disyllabicity also affects how Chinese words are spoken. In a statistical model of prosodic strength in read-aloud Mandarin speech, Kochanski et al. (2003) reported a better fit with the acoustic data when they coded dissyllabic words as trochees. Perry and Zhuang (2005) found that Mandarin speakers were more likely to produce the disyllabic variant for pictures with elastic names (e.g. (大)象 *(da)xiang* '(big) elephant') when the experimenters mixed them among pictures with unambiguously disyllabic names, which served as primes.

The disyllabic bias can even make monosyllabic words harder to learn and recognize. Wang et al. (2010) found that Mandarin-speaking children were actually more accurate at repeating spoken disyllabic words than monosyllabic ones. Their performance was also influenced by lexical frequency and lexical neighbors, but again only for disyllables. Similarly, in adult readers Tsang et al. (2018) found that lexical decisions for disyllables were faster than for monosyllables, even after factoring out character frequency, word frequency, and many other variables.

Since prosody is intrinsically hierarchical, however, disyllabic feet are not the only word-like phonological units. Shih (2017) found that Taiwan Mandarin speakers, asked to read aloud digit strings, consistently split 555555 (six copies of 五 *wu*) into 55-5#55-5, where '#' represents a major phrase boundary and '-' a minor phrase boundary, rather than splitting them into three trochees (55-55-55). He interpreted these results as showing a prosodic preference for binary branching, not for binary feet per se. In another statistical model of read-aloud Mandarin speech, Tseng et al. (2005) found acoustic evidence for polysyllabic words as prosodic units, that is, for phonological words, with speakers consistently shortening word-initial syllables and lengthening word-final ones, regardless of word length.

Phonological words (prosodic constituents with only one stressed syllable) may even affect speech planning, prior to actual articulation. Chiu (2005) prompted Taiwan Mandarin speakers to construct sentences like those in (7), where b and c both have one more syllable than a, but in b this syllable is an unstressed function morpheme (clitic) and in c it is a stressed lexical

morpheme. Preparation time (from prompt to the onset of speech) was the same for a and b but longer for c, suggesting that Mandarin speakers, like the Dutch speakers tested with this method by Wheeldon and Lahiri (1997), mentally chunk their utterances into phonological words.

(7) a. 他買課本
ta__mai__keben
he__buy__ textbook
He buys a textbook.
b. 他買了課本
ta__mai__le__keben
he__buy__ASP__textbook
He bought a textbook.
c. 他買錯課本
ta__mai__cuo__keben
he__buy__wrong__textbook
He buys the wrong textbook.

Since syntactically free function morphemes can be prosodically bound, attempts to distinguish syntactically free clitics from morphological affixes invoke the same (sometimes ambiguous) wordhood tests reviewed in this chapter (see e.g. Zwicky 1985 for English, Liu 1998 and Dong and Huang 董思聪, 黄居仁 2019a for Mandarin and Chongqing, respectively). However, distinguishing "true" clitics (phrasal affixes) from ordinary prosodically bound function morphemes involves discriminating between grammatical classes and thus falls outside the scope of this chapter.

3.5 Morphological tests

One might expect morphological tests to dominate the wordhood literature, but they are usually merely the flip side of syntactic tests (see section 3.6). One that is not is the claim in Xu (2018) that the adjective in Chinese AN compounds must be monomorphemic. Disyllabic adjectives that appear in AN compounds often seem to be functionally monomorphemic, with bound characters like the first in (8a), semantic opacity as in (8b), or semantic redundancy as in (8c) (see Xu 2018: 212–3). Xu argues in detail that this generalization cannot be explained by semantic, prosodic, or syntactic constraints. While the test is not as conclusive as Xu implies (relying as it does on a fuzzy notion of monomorphemicity; see also section 3.6), it once again suggests that Chinese words are typologically "normal": as Xu observes, a similar restriction also applies in Dutch and German AN compounds.

(8) a. 聪明人
cong__ming__ren
quick-witted__bright__person
intelligent person
b. 糊涂人
hu__tu__ren
paste__mud__person
muddle-headed person
c. 重大损力
zhong__da__sunli
heavy__big__loss
heavy loss

Further morphological evidence for wordhood comes from the optional deletion of a morpheme in one compound when next to the same morpheme in another compound, as in (9) (based on examples in Dong and Huang 2020). Crucially, this haplology only happens when the compounds themselves are compounded into one, not merely syntactically adjacent. English again displays a similar phenomenon, as reflected in the glosses, though it is less obvious because, ironically, English has many more monomorphic words than Chinese (so compounding *fruit* in (9) is redundant).

(9) a. 奇異果
qiyi__guo
kiwi__fruit
kiwi(fruit)
b. 果汁
guo__zhi
fruit__juice
(fruit) juce
c. 奇異果汁
qiyiguozhi
kiwi(fruit) juice

Finally, the bit of playful cross-linguistic morphology in (10) (Dong and Huang 董思聪, 黄居仁 2019b: 15) provides yet another sign that Chinese speakers think of their words similarly to English speakers, allowing internet writers to affix English suffixes to Chinese stems, even VO compounds.

(10) 上班 ing
shang-ban__ing
attend-work__ing
going to work

3.6 Syntactic tests

While psychological, semantic, phonological, and morphological data all suggest that Chinese has words, at least to the same fuzzy extent that English does, Chinese linguists, like linguists generally, are particularly interested in syntactic evidence (e.g. Chao 1968, Duanmu 1998, Packard 2000, Li 2005). This is reflected, for example, in the working definition in Huang et al. (2017: 13) of the Chinese word "as the smallest string of character(s) that has both an independent meaning and a fixed grammatical category." Nevertheless, syntactic tests continue to reveal Chinese wordhood as no less flexible than it is in English.

Here I focus on just two of the most notorious wordhood problems in Chinese: adjective-noun (AN) and verb-object (VO) constructions (for a third notorious type, the serial verb resultative construction, see Li, this volume). Regarding the first, Duanmu (1998) and Xu (2018) review a variety of syntactic tests strongly suggesting that AN is a word and A 的 *de* N is a phrase. In contrast to A *de* N, AN is not fully productive, and consistent with the Lexicalist Hypothesis, it also does not permit A to be modified by a degree word, as shown in (11) (the same pattern can be seen with 这么 *zheme* 'such a' and 不 *bu* 'not'), disallows internal phrases, as shown in (12), and disfavors coordination of internal parts, as shown in (13) (examples taken from or based on those in Duanmu 1998).

(11) a. 新书
xin__shu
new__book
new book
b. 新的书
xin__de__shu
new__DE__book
new book
c. *很新书
hen__xin__shu
very__new__book
d. 很新的书
hen__xin__de__shu
very__new__DE__book
very new book

(12) a. *新[三本书]
xin__san__ben__shu
new__three__CL__book
b. 新的[三本书]
xin__de__san__ben__shu
new__DE__three__CL__book
three new books
c. *[有名的作者]书
youming__de__zuozhe__shu
famous__DE__author__book
d. [有名的作者]的书
youming__de__zuozhe__de__shu
famous__DE__author__DE__book
book by a famous author

(13) a. *旧跟新书
jiu__gen__xin__shu
old__and__new__book
b. 旧跟新的书
jiu__gen__xin__de__shu
old__and__new__DE__book
old and new books

As important as such evidence is for reconfirming some sort of reality for Chinese words, they do not line up perfectly with the results of other tests. For example, despite being phrasal, even A *de* N constructions can accrue idiosyncrasies: in Taiwan, (14) is lexically specified as the title for the *Mission: Impossible* movies, a connotation it retains outside of cinematic discussions.

(14) 不可能的任務
bukeneng__de__renwu
impossible__DE__mission
impossible mission

Regarding the adverbial modification test, Duanmu (1998) admits that some modifiers are permitted in certain AN constructions, like 最 *zui* 'most' in (15a), though he notes that this is less productive than *hen* 'very'. A similar restriction on productivity is seen in the contrast between the near-synonyms in (15b-c) (judgments, but not analysis, from C-.R. Huang, pc). Such observations suggest that these modifiers are compounded rather than syntactically free. However, invoking compounding to save the adverbial modification test further complicates Xu's (2018) AN monomorphemicity constraint (see section 3.5): as often happens, different wordhood tests point in somewhat different directions.

(15) a. 最高级
zui__gao__ji
most__ high__level
the most high level
b. 著名作者小说
zhuming__zuozhe__xiaoshuo
famous__author__novel
novel by a famous author
c. ?有名作者小说
youming__zuozhe__xiaoshuo
famous__author__novel
novel by a famous author

Duanmu (1998) himself finds the coordination test unreliable, but he only cites English counterexamples (quoted earlier in section 2). Indeed, Xu (2018: 240) notes that all putative Chinese counterexamples in the literature lack an overt coordinator, as in (16), which can be analyzed instead as containing a coordinative compound (while again challenging the AN monomorphemicity constraint). While this suggests a genuine typological difference, it is precisely the reverse of the stereotype: here it is Chinese that recognizes a sharper distinction between words and phrases than English.

(16) 优劣品种
you__lie__pinzhong
good__bad__strain
good and bad strains

Xu (2018) admits that anaphors can refer to AN-internal morphemes, as the null nominal head does in (17), particularly striking here because 椒 *jiao* 'pepper' is a bound root. While he argues that this is irrelevant to wordhood, given that cross-linguistically coreference is governed by pragmatics rather than syntax (see Ward et al. 1991 for experimental evidence from English), it still seems that word-internal elements, in both Chinese and English, do tend to be less accessible to pragmatics, adding yet another gradient wordhood diagnostic to our list.

(17) 我们只买青椒，不买红的。
women__zhi__mai__qing-jiao__bu__mai__hong__de
we__only__buy_green-pepper__not__buy__red__DE
We only buy green peppers, not red ones.

Turning to VO constructions, the problem here is that regardless of their semantic properties, syntactic tests show that some are consistently decomposable (e.g. like 吃饭 *chifan* 'dine'), others are consistently atomic (e.g. 出版 *chuban* 'publish', literally 'output version'), and others seem to alternate between the two statuses (e.g. 担心 *danxin* 'worry', literally 'carry heart'). This situation is illustrated in (18)-(20) using three syntactic diagnostics: phrases allow splitting, allow topicalization, and disallow a direct object, while words show the reverse pattern. The judgments here were checked in an informal survey of Mandarin speakers; I return below to the less-than perfect acceptability of (19c) (treated as grammatical in Huang 1984: 64).

(18) a. 他们吃什么饭？
tamen__chi__shenme__fan
they__eat__what__rice
What type of meal (or rice) are they eating?
b. *他们出什么版？
tamen__chu__shenme__ban
they__output__what__edition
What are they publishing?
c. 他们担什么心？
tamen__dan__shenme__xin
they__carry__what__heart
What are they worried about?

(19) a. 饭，他们一点都没吃。
fan__tamen__yidian__dou__mei__chi
rice__they__a-bit__all__not__eat
As for dining (eating rice), they did not at all.
b. *版，他们一点都没出。
ban__tamen__yidian__dou__mei__chu
edition__they__a-bit__all__not__output
c. ?心，他们一点都没担。
xin__tamen__yidian__dou__mei__dan
heart__they__a-bit__all__not__carry
As for worrying, they did not at all.

(20) a. *他们会吃饭面条。
tamen__hui__chi__fan__miantiao
they__will__eat__rice__noodles
b. 他们会出版那本书。
tamen__hui__chu__ban__yi__ben__shu
they__will_output__edition__that__CL__book
They will publish that book.
c. 他们会担心你。
tamen__hui__dan__xin__ni
they__will__carry__heart__you
They will worry about you.

Huang (1984) argues that such facts are not a problem for a syntax/morphology distinction: in any particular context, words like *danxin* either behave as syntactically composed or as

syntactic atoms, but never both at once. However, this admission still implies that word/phrase boundary is permeable, whether top down (lexicalization of phrases into words: Huang 1984) or bottom up (reanalysis of words as phrases: Packard 2000).

The boundary may be still fuzzier than that, however. Consider the example in (21a), which shows what Chao (1968: 433) called the "ionization" (splitting) of 幽默 *youmo* 'tease' (from English *humor*) in a Taiwanese newspaper. The Taiwan-Mandarin-speaking Huang (1984: 65) supplements this with the invented examples in (21b) and (21c), also claimed to be acceptable. But while all of the contemporary Taiwan Mandarin speakers I polled accepted Chao's original split construction in (21a), only some accepted the topicalized structure in (21b) and even fewer the unsplit form in (21c). This puts *youmo* in gradient contrast with *danxin*: while both can be split or remain whole, they seem to differ in the readiness with which they do so. More careful testing will be necessary to understand the actual situation (as Chen et al. forthcoming did in testing 557 Mandarin sentences in Huang et al. 2009, which unfortunately did not include cases like these).

(21) a. 還幽了他一默，說…
hai__you__le__ta__yi__mo__shuo
also__hu-__ASP__he__one__-mor__say
(I) teased him again, saying...

b. 這種默，我想你最好還是不要幽。
zhe__zhong__mo__wo__xiang__ni__zuihao__hai__buyao__you
this__kind__-mor__I__think__you__best__still__not__hu-
This way of teasing, I think you'd better not do it.

c. ?我常常幽默他。
wo__changchang__youmo__ta
I__often__tease__he
I often tease him.

To summarize all of section 3, then, the evidence for words is as strong in Chinese as in English, but as with English, we have to accept that wordhood tests do not entirely agree and that wordhood status also varies across the context of use.

4 A dynamic approach to wordhood

This shifting nature of words is just what we would expect if they arise through the dynamic interaction among distinct linguistic forces. In section 4.1 I argue that this is indeed the most promising way to look at words and sketch out how the idea may be formalized. In section 4.2 I show how this formalism captures several important aspects of Chinese wordhood.

4.1 Formalizing fuzziness

The challenge of wordhood is so notorious that it has led to two diametrically opposed attempts to eliminate words altogether: either it's syntax all the way down (e.g. Bruening 2018; Marantz 1997, 2013), or the lexicon all the way up (e.g. Baayen and Ramscar 2015; Booij 2012; Daelemans and van den Bosch 2005; Jackendoff and Audring 2016). Both extremes oversimplify, of course. Ostensibly all-syntax approaches actually shunt difficulties off to poorly described non-syntactic components. All-lexicon approaches overestimate the feasibility of generating syntactic regularity solely via analogy with memorized exemplars. Even approaches lying between the two extremes differ in whether they are more top-down (e.g., the Morphology-Syntax Mapping Hypothesis of Li 2005 retains the Lexicalist Hypothesis

but still gives all of the interesting work to syntax) or more bottom-up (e.g. the framework I am about to present here).

I personally favor working bottom up because I see that as the best way to respect words as truly multifaceted, not merely as the failed syntactic phrases posited by the Lexicalist Hypothesis. Single morphemes are too small to express the complex meanings, rhythmic prosody, syntactic architecture, and user-friendly expressions needed for effective information encoding and communication, so the semantics adds tacit meanings, phonology builds clitic groups, syntax builds trees, and processing clumps or cleaves whenever convenient.

The dynamic view of words has been operationalized in a variety of ways. The segmentation heuristics of the Sinica Corpus (Huang et al. 2017) refer not just to syntax (favoring a word analysis for a string if the component characters cannot explain the string's contextual behavior), but also semantics (non-decompositionality), phonology (disyllabicity), and psycholinguistics (frequency). Jackendoff and Audring (2016) formalize the interfacing of semantics, phonology, and morphosyntax in terms of schemas. Computational models deriving wordhood from the transition probability of lower-level units are presented in Bicknell and Levy (2010) for eye movements in reading, and independently in Huang and Xue (2012) for automated Chinese text segmentation. Geertzen et al. (2016) operationalize the word as the maximally informative unit (in the sense of mathematical information theory; see also Harris 1954). Baayen et al. (2015) show how a computational model of child language acquisition mapping phonemes to semantic units is capable of learning "words" without performing any overt segmentation procedure at all.

To keep the discussion concrete, here I adopt just the last of these models (see Baayen and Ramscar 2015 for a non-technical overview). Naive discriminative learning (NDL), motivated by general learning theory, consists of one layer of connections between "cues" (e.g. Chinese characters) and "outcomes" (e.g. meanings) for each learning "event" (e.g. a character string with a known meaning). The learning algorithm is discriminative because a cue-outcome connection is strengthened only if the cue is informative (A→X and AB→X will not generalize to B→X); it is naive because each cue-outcome connection is adjusted while ignoring all other outcomes. A schematic NDL model is shown in Figure 2 (based on (2a)); training by events would strengthen some connections more than others.

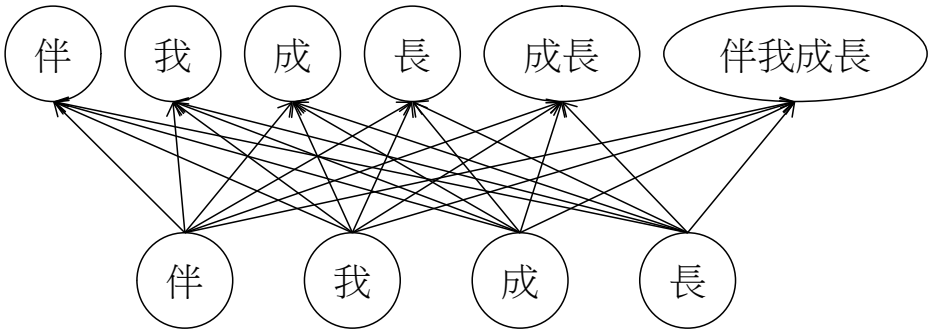

Figure 2. NDL model linking single-character cues to whole-word meaning outcomes

NDL makes an attractive formalization of dynamic wordhood for a number of reasons. As just noted, it can learn from fluent language use without overt word segmentation, and it can also model distinct morpheme-level, word-level, and phrase-level effects within the same one-layer network (Baayen et al. 2013). While working bottom-up, it also incorporates insights from top-down approaches; the network nodes may be as abstract as the modeler deems necessary, and as Marantz (2013) points out, in current practice its meaning outcomes are linguistically constrained (and not, say, universal conceptual atoms). The content-neutral nature of NDL architecture also allows it to go beyond linking form with meaning, to linking

images with meanings or even meanings with meanings (Hendrix et al. 2017). Putting these last two points together, NDL has the potential to include abstract syntactic elements as well, in case it turns out (as seems highly likely to me) that not all of syntax is reducible to analogy. Of course, by itself, NDL cannot explain where the patterns it learns come from in the first place, but perhaps it could do so by incorporating diachronic feedback loops (e.g. Kirby 2001), thereby capturing the dynamic nature of wordhood across generations as well.

4.2 Applications to Chinese

To explore how NDL could help with the Chinese wordhood question, Tsung-Ying Chen and I trained a model, via the ndl package (Arppe et al. 2018) in R (R Core Team 2020), on the written transcription of the spoken portion of the Sinica Corpus, chosen primarily for its relatively small size (linking all possible cues with all possible outcomes makes NDL a memory hog, so modeling even this half-million-word corpus required around 28 GB of RAM). Events were strings of characters in the transcription, demarcated on each end by punctuation; cues were the individual characters within an event, and outcomes were the words in the event as segmented in the corpus. The *n* most activated outcomes per event, where *n* was the "actual" number of words, were taken as the trained model's word guesses. While the model correctly identified only 75% of the "actual" word tokens, this was probably not the fault of NDL itself; to save memory space, our cues contained no sequential information (e.g. ABCD was represented the same as BDCA) and by using whole words as outcomes, our model falsely assumes that all Chinese words are semantically opaque.

Nevertheless, even this simple model managed to capture several observations made in section 3. Consistent with human word recognition, we found that the more predictable one character was from another within a two-character word, the more accurate the NDL model was at identifying this word ($\tau = .39$, $z = 58$, $p < .0001$; we used the Kendall rank correlation coefficient due to the non-normality of both variables). This result is particularly striking given that the cues were unordered characters and thus transition probability was not coded directly.

Also human-like was the positive correlation between the accuracy of our model and word frequency ($\tau = .32$, $z = 46$, $p < .0001$). The model simultaneously handled morpheme and phrase frequency as well (as Baayen et al. 2013 found for English): a multiple linear regression predicting the proportion of "actual" Chinese words detected per event showed not only a positive effect of mean word frequency ($B = 0.68$, $t = 233$, $p < .0001$) but also independent effects of mean character frequency ($B = -1.08$, $t = -151$, $p < .0001$) and whole-string frequency ($B = -0.02$, $t = -30$, $p < .0001$), the latter two effects negative due to the model's human-like tendency not to decompose these "opaque" words or common phrases.

We also explored how NDL can capture semantic tests for wordhood. Using a toy corpus, we linked cue pairs (simulating a two-character compound) with one outcome (opaque), two outcomes (transparent), or variably one or two outcomes (ambiguous). As expected, in the last case the model distributed activations across both opaque and transparent meanings (similar to *honghua*, which may mean either 'safflower' or 'red flower').

Another toy model captured the word segmentation triggered by function morphemes, a factor in some syntactic tests. We trained the events AB, BA, CD, DC, AfB, BfA, CfD, DfC, where capital letters represent content morphemes and f represents a function morpheme, with cues coded as bigrams (e.g. AB was coded as #A, AB, B#, and AfB as #A, Af, fB, B#). Outcomes assumed full semantic transparency (e.g. AB was linked with A and B, and AfB with A, f, B). When we tested the trained model on the untrained inputs AD and AfD, we found that activation of the individual content morphemes A and D was higher for AfD than for AD, just as in Chinese A *de* N is more decomposable than AN. This behavior resulted from the fact that in training, the bigrams Af and fD also became associated with A and D, respectively, boosting their activation when prompted with AfD, whereas the bigram AD had never been

encountered at all. In plain language, function morphemes trigger segmentation because they appear in more different contexts and thus have lower transition probabilities.

Since NDL cues are modality-specific form units, wordhood should not be the same for readers and listeners. While we have not modeled this in Chinese, predictions can be derived from the work of Pham and Baayen (2015) on Vietnamese, which also has monosyllabic morphemes, rampant homophony, and a predilection for compounding. They first report experimental results showing that morpheme frequency slows wordhood judgments in Vietnamese, the reverse of English. They then model this result in NDL, with Vietnamese coded in orthographic letter bigrams, and find that the lower activation of words with high-frequency morphemes is caused by homophony overloading the model's discriminative ability. NDL thus predicts that Chinese character frequency effects should generally be facilitative in reading, since like English morphemes, characters are readily discriminable, whereas for spoken Chinese, homophony should cause syllable frequency to slow responses. As we saw in Section 3.2.2, both predictions are correct.

5 Conclusions

The evidence is as strong for words in Chinese as it is in English or any other language, and indeed, both languages, like many others, prefer words to be disyllabic. All languages have wordlike units, with roughly the same processing, semantic, phonological, morphological, and syntactic behavior, since words provide an optimal solution to a universal engineering problem: linking rote memory with productive grammar. Interfacing between such fundamentally different domains, however, requires dynamic compromise. Words are thus something like tornados, twisted out of thin air by powerful competing forces, and continuing to twist just out of our grasp even once formed, but still undeniably real for all that. Capturing their full richness may require computational help, though insights from traditional linguistic analysis and experimental psycholinguistics are still needed to keep the results sufficiently human. Because words are shaped by the interfacing domains, there is no reason for the various linguistic subdisciplines to give up their favorite working definitions, but deeper insights will require deeper collaboration. In particular, Chinese and non-Chinese linguists need to consult more closely together to figure out where and why words really do differ cross-linguistically, and where and why they do not.

References

Arnon, Inbal, and Neal Snider. 2010. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62 (1): 67–82.

Arppe, Antti, Peter Hendrix, Petar Milin, R. Harald Baayen, Tino Sering and Cyrus Shaoul. 2018. ndl: Naive discriminative learning. R package version 0.2.18. https://cran.r-project.org/web/packages/ndl/index.html.

Baayen, R. Harald, and Michael Ramscar. 2015. Abstraction, storage and naive discriminative learning. In *Handbook of cognitive linguistics*, ed. Ewa Dabrowska and Dagmar Divjak, 99–120. Berlin: De Gruyter Mouton.

Baayen, R. Harald, and Antoinette Renouf. 1996. Chronicling the *Times*: Productive lexical innovations in an English newspaper. *Language* 72 (1): 69–96.

Baayen, R. Harald, Cyrus Shaoul, Jon Willits, and Michael Ramscar. 2015. Comprehension without segmentation: A proof of concept with naive discriminative learning. *Language, Cognition and Neuroscience* 31(1): 106–28.

Baayen, R. Harald, Peter Hendrix, and Michael Ramscar. 2013. Sidestepping the combinatorial explosion: An explanation of *n*-gram frequency effects based on naive discriminative learning. *Language and Speech* 56(3): 329–47.

Bai, Xuejun, Guoli Yan, Simon P. Liversedge, Chuanli Zang, and Keith Rayner. 2008. Reading spaced and unspaced Chinese text: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance* 34(5):1277–1287.

Bauer, Laurie. 1983. *English word-formation*. Cambridge, UK: Cambridge University Press.

Bicknell, Klinton and Roger Levy. 2010. A rational model of eye movement control in reading. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, 1168–78.

Bloomfield, Leonard. 1926. A set of postulates for the science of language. *Language* 2(3): 153–64.

Booij, Geert. 2012. Construction Morphology, a brief introduction. *Morphology* 22:343–46.

Bruening, Benjamin. 2018. The Lexicalist Hypothesis: Both wrong and superfluous. *Language* 94(1): 1–42.

Chao, Yuen-Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.

Chen, Zhong, Yuhang Xu, and Zhiguo Xie. Forthcoming. Assessing introspective linguistic judgments quantitatively: The case of *The Syntax of Chinese*. To appear in *Journal of East Asian Linguistics*.

Chiu, Chenhao C. 2005. Phonological words in Mandarin speech production. *Berkeley Linguistics Society: Proceedings of the 31st Annual Meeting* 31 (1): 61–72.

Chomsky, Noam. 1970. Remarks on Nominalization. *Readings in English transformational grammar*, ed. by Roderick A. Jacobs and Peter S. Rosenbaum, 11–61. Waltham, MA: Ginn and Co.

Daelemans, Walter, and Antal van den Bosch. 2005. *Memory-based language processing*. Cambridge, UK: Cambridge University Press.

DeFrancis, John. 1986. *The Chinese language: Fact and fantasy*. Honolulu: University of Hawaii Press.

Dell, François. 2004. On recent claims about stress and tone in Beijing Mandarin. *Cahiers de Linguistique Asie Orientale* 33 (1): 33–63.

Ding, Nai Ding, Lucia Melloni, Hang Zhang, Xing Tian, and David Poeppel. 2016. Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience* 19(1): 158–64.

Dong, Sicong, and Chu-Ren Huang 董思聪, 黄居仁. 2019a. Clitic *zhi* (之) in Chongqing dialect and the classification of clitics 重庆方言的语缀“之”及语缀的分类问题. In 语言学论丛第 59 辑 *Essays on Linguistics* 59. Beijing: Commercial Press.

Dong, Sicong and Chu-Ren Huang 董思聪, 黄居仁. 2019b. The limits of language innovation in special linguistic zones 语言特区中创新形式的限度. *TCSOL Studies* 华文教学与研究 76 (4): 11–20.

Dong, Sicong, and Sam Yin Wong. 2020. Haplology and lexical entries: a study based on cross-linguistic data from Sinitic languages. *Lexicography*.

Duanmu, San. 1998. Wordhood in Chinese. In *New approaches to Chinese word formation: Morphology, phonology and the lexicon in Modern and Ancient Chinese*, ed. Jerome L. Packard, 135–96. Berlin: Mouton de Gruyter.

Duanmu, San. 1999. Stress and the development of disyllabic words in Chinese. *Diachronica* 16: 1–35.

Duanmu, San. 2012. Word-length preferences in Chinese: a corpus study. *Journal of East Asian Linguistics* 21 (1): 89–114.

Duanmu, San. 2017. Word and wordhood, modern. In *Encyclopedia of Chinese language and linguistics, vol. IV*, ed. by Rint Sybesma, 543–49. Leiden: Brill.

Duanmu, this volume.

Evert, Stefan, and Marco Baroni. 2007. zipfR: Word frequency distributions in R. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Posters and Demonstrations Sessions*, pp. 29–32.

Feng, Shengli. 1998. Prosodic structure and compound words in Classical Chinese. In *New approaches to Chinese word formation: Morphology, phonology and the lexicon in Modern and Ancient Chinese*, ed. Jerome L. Packard, 197–260. Berlin: Mouton de Gruyter.

Garrett, Edward. 1999. Minimal words are not minimal feet. *UCLA Working Papers in Linguistics* 1 [Papers in Phonology 2]: 68–105.

Geertzen, Jeroen, James P. Blevins, and Petar Milin. 2016. The informativeness of linguistic unit boundaries. *Italian Journal of Linguistics* 28 (1): 25–48.

Gordon, Matthew. 2002. A factorial typology of quantity-insensitive stress. *Natural Language & Linguistic Theory* 20 (3): 491–552.

Hagstrom, Paul. 2006. A-not-A questions. In *The Blackwell companion to syntax*, ed. Martin Everaet and Henk van Riemsdijk, pp. 173–214. Malden, MA: Blackwell.

Harris, Zellig S. 1954. Distributional structure. *Word* 10: 146–62.

Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1):31–80.

Hendrix, Peter, Patrick Bolger, and Harald Baayen. 2017. Distinct ERP signatures of word frequency, phrase frequency, and prototypicality in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43(1): 128–49.

Hoosain, Rumjahn. 1992. Psychological reality of the word in Chinese. *Language processing in Chinese*, ed. by Hsuan Chih Chen and Ovid J. L. Tzeng, 111–30. Amsterdam: North-Holland.

Huang et al. In this volume.

Huang, Chu-Ren, and Nianwen Xue. 2012. Words without boundaries: Computational approaches to Chinese word segmentation. *Language and Linguistics Compass* 6(8): 494–505.

Huang, Chu-Ren, Keh-jiann Chen, Feng-yi Chen, and Li-Li Chang. 1997. Segmentation standard for Chinese natural language processing. *Computational Linguistics and Chinese Language Processing* 2(2): 47–62.

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

Huang, James C.-T. 1984. Phrase structure, lexical integrity, and Chinese compounds. *Journal of the Chinese Language Teachers Association* 19(2):53–78.

Huang, James C-T. 1991. Modularity and Chinese A-not-A questions. In *Interdisciplinary approaches to language: Essays in honor of Yuki Kuroda*, ed. Carol Georgopoulos and Roberta Ishihara, pp. 305-332. Dordrecht: Springer.

James, Huang C-T., Y-H. Audrey Li, and Yafei Li. 2009. *The syntax of Chinese*. Cambridge, UK: Cambridge University Press.

Inkelas, Sharon. 2008. The dual theory of reduplication. *Linguistics* 46: 351–402.

Jackendoff, Ray and Jenny Audring. 2016. Morphological schemas: Theoretical and psycholinguistic issues. *The Mental Lexicon* 11 (3), 467–93.

Kirby, Simon. 2001. Spontaneous evolution of linguistic structure – an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation* 5(2): 102–10.

Kochanski, Greg, Chilin Shih, and Hongyan Jing. 2003. Quantitative measurement of prosodic strength in Mandarin. Speech Communication 41 (4): 625–45.

Lenzo, Kevin. 2014. The CMU pronouncing dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

Li, Meng-Feng, Xin-Yu Gao, Tai-Li Chou, and Jei-Tun Wu. 2017. Neighborhood frequency effect in Chinese word recognition: Evidence from naming and lexical decision. *Journal of Psycholinguistic Research* 46(1): 227–45.
Li, this volume.
Li, Xingshan, Klinton Bicknell, Pingping Liu, Wei Wei, and Keith Rayner. 2014. Reading is fundamentally similar across disparate writing systems: A systematic characterization of how words and characters influence eye movements in Chinese reading. *Journal of Experimental Psychology: General* 143 (2): 895–913.
Li, Yafei. 2005. *X^0: A theory of the morphology-syntax interface*. Cambridge, MA: MIT Press.
Lin, Tzu-Jung, Richard C. Anderson, Yu-Min Ku, Kiel Christianson, and Jerome L. Packard. 2011. Chinese children's concept of word. *Writing Systems Research* 3(1): 41–57.
Liu, Fenghsi. 1998. A clitic analysis of locative particles. *Journal of Chinese Linguistics* 26 (1): 48–70.
Liu, Li. 2015. *Chinese quatra-syllabic schematic idioms: Description and acquisition*. Hong Kong: The Hong Kong Institute of Education dissertation.
Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4(2): 201–25.
Marantz, Alec. 2013. No escape from morphemes in morphological processing. *Language and Cognitive Processes* 28(7):905–16.
Matthews, Peter. H. 1991. *Morphology* (2nd edition). New York: Cambridge University Press.
Mok, Leh Woon. 2009. Word-superiority effect as a function of semantic transparency of Chinese bimorphemic compound words. *Language and Cognitive Processes* 24(7–8): 1039–81.
Monsell, Stephen. 1991. The nature and locus of word frequency effects in reading. In *Basic processes in reading: Visual word recognition*, ed. Derek Besner and Glyn W. Humphreys, 148–197. Hillsdale, NJ: Erlbaum.
Müller, Stefan. 2018. The end of lexicalism as we know it? *Language* 94 (1): e54–e66.
Myers, James. 2017. Morphological processing of compounds, behavioral studies. *Encyclopedia of Chinese language and linguistics, vol. III*, ed. by Rint Sybesma, 94–100. Leiden: Brill.
Myers, James and Jane Tsay. 2015. Trochaic feet in spontaneous spoken Southern Min. *Proceedings of the 27th North American Conference on Chinese Linguistics, Vol. 2*, Los Angeles, 368–87.
Myers, James, Yu-chi Huang, and Wenling Wang. 2006. Frequency effects in the processing of Chinese inflection. *Journal of Memory and Language* 54 (3): 300–23.
Newmeyer, Frederick J. 2009. Current challenges to the lexicalist hypothesis: An overview and a critique. *Time and again: Theoretical perspectives on formal linguistics in honor of D. Terence Langendoen*, ed. by William D. Lewis, Simin Karimi, Heidi Harley, and Scott O. Farrar, 91–117. Amsterdam: John Benjamins.
Packard, Jerome L. 1999. Lexical access in Chinese speech comprehension and production. *Brain and Language* 68(1): 89–94.
Packard, Jerome L. 2000. *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge, UK: Cambridge University Press.
Perry, Conrad, and Jie Zhuang. 2005. Prosody and lemma selection. *Memory and Cognition* 33: 862–70.
Pham, Hien, and Harald Baayen. 2015. Vietnamese compounds show an anti-frequency effect in visual lexical decision. *Language, Cognition and Neuroscience* 30(9): 1077–95.
Qin Zuxuan and San Duanmu. 2017. A judgment study of word-length preferences in Chinese NN compounds. *Lingua* 198: 1–21.

Qin, Zuxuan, and San Duanmu. 2019. An acceptability judgment study of 1+ 1 NN compounds in Chinese. *Lingua* 222: 26–38.
R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://cran.r-project.org/.
Ruan, Jia-Cing, Chiung-Wen Hsu, James Myers, and Jane S. Tsay. 2012. Development and testing of transcription software for a Southern Min spoken corpus. *International Journal of Computational Linguistics and Chinese Language Processing* 17(1):1–26.
Shih, Shu-hao. 2017. Major phrases are binary: Evidence from Taiwan Mandarin flat structure. In *Proceedings of the 34th West Coast Conference on Formal Linguistics*, ed. Aaron Kaplan, Abby Kaplan, Miranda K. McCarvel, and Edward J. Rubin, 454-461. Somerville, MA: Cascadilla Proceedings Project.
Spencer, Andrew. 1988. Bracketing paradoxes and the English lexicon. *Language* 64(4):663–82.
Taft, Marcus. 1979. Recognition of affixed words and the word frequency effect. *Memory & Cognition* 7: 263–72.
Tsang, Yiu-Kei, Jian Huang, Ming Lui, Mingfeng Xue, Yin-Wah Fiona Chan, Suiping Wang, and Hsuan-Chih Chen. 2018. MELD-SCH: A megastudy of lexical decision in simplified Chinese. *Behavior Research Methods* 50 (5): 1763–77.
Tseng, Chiu-yu, Shao-huang Pin, Yehlin Lee, Hsin-min Wang, and Yong-cheng Chen. 2005. Fluent speech prosody: Framework and modeling. *Speech Communication* 46 284–309.
Wang, Nan Mai, Che-Ming Wu, and Karen Iler Kirk. 2010. Lexical effects on spoken word recognition performance among Mandarin-speaking children with normal hearing and cochlear implants. *International Journal of Pediatric Otorhinolaryngology* 74 (8): 883–90.
Wang, Shichang, Chu-Ren Huang, Yao Yao, and Angel Chan. 2017. Word intuition agreement among Chinese speakers: a Mechanical Turk-based study. *Lingua Sinica* 3 (1): 13.
Ward, Gregory, Richard Sproat, and Gail McKoon. 1991. A pragmatic analysis of so-called anaphoric islands. *Language* 67 (3): 439–74.
Wheeldon, Linda, and Aditi Lahiri. 1997. Prosodic units in speech production. *Journal of Memory and Language* 37:356–81.
Xu, Zheng. 2018. The word status of Chinese adjective-noun combinations. *Linguistics* 56 (1): 207–56.
Yip, Po-Ching. 2000. *The Chinese lexicon: A comprehensive survey*. London: Routledge.
Zádrapa, Lukáš. 2017. Word and wordhood, premodern. *Encyclopedia of Chinese language and linguistics, vol. IV*, ed. by Rint Sybesma, 549–54. Leiden: Brill.
Zhou, Xiaolin, and William Marslen-Wilson. 1994. Words, morphemes and syllables in the Chinese mental lexicon. *Language and Cognitive Processes* 9(3):393–422.
Ziemski, Michał, Marcin Junczys-Dowmunt, and Bruno Pouliquen. 2016. The United Nations Parallel Corpus v1.0. *Language Resources and Evaluation*. Portorož, Slovenia.
Zipf, George Kingsley. 1935. *The psychobiology of language: An introduction to dynamic philology*. Boston, MA: Houghton–Mifflin.
Zou, Lijuan, Jerome Packard, Zhichao Xia, Youyi Liu, and Hua Shu. 2019. Morphological and whole-word semantic processing are distinct: ERP evidence from spoken word recognition in Chinese. *Frontiers in Human Neuroscience* 13: 133.
Zwicky, Arnold M. 1985. Clitics and particles. *Language* 61 (2): 283–305.