An acceptability judgment study of 1+1 NN compounds in Chinese*

Zuxuan Qin & San Duanmu

Abstract

Chinese 1+1 NNs refer to compounds made of two monosyllabic nouns, such as *shū-bāo* 'book-bag' and *jī-dàn* 'chicken-egg'. They occur frequently, do not violate any prosodic constraint, and appear to be highly productive. However, in Qin & Duanmu (2017), it was found that the mean acceptability ranking was quite low for randomly constructed 1+1 NNs. Since the focus of Qin & Duanmu (2017) was on a different issue, the set of 1+1 NNs was small and not of balanced design. In this study, we gathered acceptability scores from 17 Chinese native speakers on 1,000 randomly constructed 1+1 NNs. It is found that the mean acceptability score is only 22.6%, even lower than that of Qin & Duanmu (2017). In addition, while Qin & Duanmu (2017) found ambiguity, naturalness, and frequency to have significant effects on acceptability scores, the present study did not find ambiguity to be relevant. We also found that 1+1 NNs consisting of two elastic nouns are significantly less acceptable than those consisting of one or two non-elastic nouns, although the effect size is rather small. Possible reasons for the low acceptability of 1+1 NN compounds are discussed, including semantic transparency, grammatical relations between the two nouns of a compound, and modern vs. classic vocabulary.

Keywords: NN compounds, length patterns, elastic words, judgment, acceptability

1. Introduction

Word length preferences in Chinese have received much discussion in the literature, such as Lü (1963), Wu (1986, 2006), Feng (1996, 2013), Duanmu (1999, 2007, 2012), Wang (2001), Lu & Duanmu (2002), Wang (2002), Ke (2007), Zhou (2011, 2017), Zhang (2014), and Shen (2012, 2015), among others. There is a broad consensus that, in NN compounds, 1+2 (monosyllabic + disyllabic) is generally bad, whereas 1+1 (monosyllabic + monosyllabic), 2+1 (disyllabic + monosyllabic) and 2+2 (disyllabic + disyllabic) are generally good. The generalization is illustrated in (1), where the word for 'coal' and 'store' can be either monosyllabic or disyllabic (these patterns apply to many other Chinese words).

(1) Generalization on word length preferences in Chinese NN compounds

	Generalization	Example
a.	1+1 is well formed	méi-diàn
b.	*1+2 is ill formed	* méi-shāngdiàn
b.	2+1 is well formed	méitàn-diàn
d.	2+2 is well formed	méitàn-shāngdiàn
		'coal store'

There is further evidence for the generalization in (1). For example, Duanmu (2002) (1999, 2007) and Lu & Duanmu (2002) have offered a phonological explanation in terms of compound stress and foot binarity. In addition, Duanmu (2012) has shown that, in corpus data, 1+2 NN occurs infrequently while other length patterns occur frequently.

However, in a judgment study by Qin & Duanmu (2017), a problem is found with 1+1 NN. Consider the data in (2).

(2) Mean acceptability scores of the four length patterns of Chinese NN compounds (Qin & Duanmu 2017)

Length	Acceptability	St. dev	N
1+1	47.8%	18.6%	283
1+2	49.6%	17.4%	283
2+1	70.2%	16.1%	283
2+2	95.1%	7.4%	283

Two observations can be made. First, although 2+2 and 2+1 have high scores, 2+2 is clearly more acceptable than 2+1. Second, contrary to the expectation that 1+1 is highly acceptable, 1+1 and 1+2 received similar low scores and were both judged to be less acceptable than the other patterns.

The first observation has a plausible explanation. If we assume that the preferred foot is a binary one, then 2+2 forms two binary feet, but 2+1 has an unparsed syllable (or has a ternary foot), which could explain why 2+2 is better than 2+1.

However, the second observation is harder to explain. In Qin & Duanmu (2017), we reported that, while word length has a major influence on acceptability, four other factors, ambiguity, naturalness, frequency, and boundness, also have significant effects. However, in that study, 1+1 NN tokens were not balanced for all factors. For example, of the 283 tokens of 1+1 NN, just 11 consisted of free morphemes only, while 272 consisted of at least one bound morpheme. In addition, all of the compounds were made of words with 'elastic length', i.e. those that can be either monosyllabic or disyllabic in length, such as *mài(zi)* 'wheat' and *jià(gé)* 'price', where parentheses indicate a semantically redundant syllable that can be dropped in some contexts (Karlgren 1918; Guo 1938; Lü 1963; Duanmu 2002). Thus, each NN compound has four possible length patterns: 1+1, 1+2, 2+1, and 2+2. For example, the length patterns for 'wheat-price' are *mài-jià* (1+1), *mài-jiàgé* (1+2), *màizi-jià* (2+1), and *màizi-jiàgé* (2+2). However, in Duanmu et al. (2018), it was reported that 'elasticity' (i.e. the availability of length choices) has an effect on the acceptability of VO phrases. Specifically, when a word has elastic length, the preferred length ought to be used. In contrast, when a word has no elastic length, there is no length choice, and the word can still be used; this usage is referred to as "last resort".

In view of the discussion above, it is worth asking whether elastic length has an effect on the acceptability of 1+1 NNs. Given the fact that many existing 1+1 NNs, such as *zhū-ròu* 'pig-meat', *jī-ròu* 'chicken-meat', and *é-ròu* 'goose-meat', are composed of non-elastic nouns (i.e. nouns that only have a monosyllablc form), one might expect them to be more acceptable than those consisting of elastic nouns; this remains to be verified.

In summary, the low acceptability of freely combined 1+1 NN compounds is unexpected by the generalization in (1). In this study, we verify this generalization with a better design. In addition, we consider whether additional factors including ambiguity, naturalness, frequency, boundness, and elasticity can affect the low acceptability of 1+1 NN.

In section 2 we outline the methodology used in this study. In section 3 we report the results. In section 4 we first consider the effects of ambiguity, naturalness, frequency, boundness, and elasticity separately; then we consider their effects together in a linear mixed-effects model, using RStudio (2012) and lme4 (Bates et al. 2015). In section 5 we compare the results of the present study with those of Qin & Duanmu (2017) and explore other explanations for the low acceptability of 1+1 NNs, including semantic transparency (Reddy et al. 2011; Bell & Schäfer 2016; Schäfer 2018), grammatical relations between the two nouns of a compound (Bisetto & Scalise 2005; Ceccagno & Scalise 2006; Ceccagno & Basciano 2009), and modern vs. classic vocabulary. In section 6 we offer concluding remarks.

2. Method

2.1. Test materials

To ensure a balanced representation of 1+1 NN compounds, we used random sampling and free combination to construct stimuli. The steps were as follows.

(i) Gathering an initial list of nominal monosyllabic morphemes (characters). We gathered all commonly used characters from nominal monosyllabic entries in XDHYCD (2005), based on the following two criteria. First, we excluded characters which only occur in restricted contexts, such as those labeled as <口> (colloquial), <书> (literary), <方> (dialectal), and <古> (archaic). Second, we excluded characters with a frequency lower than 1,000 according to Modern Chinese Character Frequency List (Da 2004). After these exclusions, we were left with 3,592 characters.

(ii) Gathering the final list of characters. From the 3,592 characters obtained in (i), we randomly selected 300. From this list, we excluded those whose nominal meaning is seldom used. For example, we excluded the character 举 *jǔ*, because although it occurs frequently, most of its frequency comes from its verb form (‘to raise’ or ‘to lift’) and not from its noun form (‘a successful candidate in the imperial examinations at the provincial level in ancient China’). This yielded a final list of 257 characters.

(iii) Generating test 1+1 NNs. By exhaustively combining the list of 257 characters with each other, we obtained 257 x 257 = 66,049 1+1 NNs. From this list, we randomly selected 1,000 1+1 NNs to be used for our acceptability judgment experiment.

2.2. Experimental procedure

The 1,000 1+1 NNs, written in Chinese, were made into a randomized list, each on a separate line in a spreadsheet file. Since some characters have more than one meaning, the intended meaning of every character was indicated on the answer sheet. Subjects were asked to rate the acceptability of the 1,000 test NNs on a 7-point scale, with 1 representing ‘completely unacceptable’ and 7 ‘completely acceptable’. Some examples are shown in (3) (in the original) and (4) (English translation).

(3) Sample items in the test sheet

字组	字组中两字意义的说明	得分	评分说明
道朝	道=水流通行的途径, 如“河道”的“道”; 朝=朝代		(1=完全不可以) 1→2→3→4→5→6→7 (7=完全可以)
庙樱	庙=宗庙; 樱=樱花		(1=完全不可以) 1→2→3→4→5→6→7 (7=完全可以)
狱衫	狱=监狱; 衫=单上衣		(1=完全不可以) 1→2→3→4→5→6→7 (7=完全可以)

(4) English translation of (3)

NN	Intended meanings of each N	Score	Explanation of score scale
dào-cháo	*dào* means ‘channel through which water flows’, e.g. ‘channel’ in ‘river channel’; *cháo* means ‘dynasty’		(1=completely unacceptable) 1→2→3→4→5→6→7 (7=completely acceptable)
miào-yīng	*miào* means ‘temple’; *yīng* means ‘sakura’		(1=completely unacceptable) 1→2→3→4→5→6→7 (7=completely acceptable)
yù-shān	*yù* means ‘prison’; *shān* means ‘thin upper garment’		(1=completely unacceptable) 1→2→3→4→5→6→7 (7=completely acceptable)

Seventeen native speakers of Chinese, all unfamiliar with the field of linguistics, were paid to participate in the judgment experiment. They were all university students in mainland China, ranging in age (mean = 21.6, standard deviation = 2.6) from 19 to 27. All of them showed great willingness to participate in the experiment. As is the normal case in China, all of them mainly speak Chinese in their daily life. Nine of them were pursuing a bachelor’s degree and eight of them were pursuing a master’s degree. They were instructed to rate the acceptability of each 1+1 NN according to the specified meanings of its component Ns. The tests took place in a relaxed and comfortable setting. To avoid fatigue, we divided the list into four sessions, with 250 1+1 NNs per session. There was an interval from two to three days between sessions. In each session, there was no time limit; subjects were allowed as much time as needed to rate each test compound. However, feedback from the subjects showed that it normally took 120 minutes or so to finish each session. All the questionnaires were distributed and collected through email. 【以上3-5行绿色两句话显得多余】

3. Results

For better comparison with Qin & Duanmu (2017), we converted the seven-point acceptability scores into percentage values, with 1 = 0%, 2 = 16.7%, 3 = 33.3%, 4 = 50%, 5 = 66.7%, 6 = 83.3%, and 7 = 100%. It can be shown that the conversion has little effect on the statistical results. The mean acceptability score is shown in (5), at merely 22.6%, less than half of the mean acceptability score of 1+1 NNs (47.8%) reported in Qin & Duanmu (2017).

(5) Mean acceptability score of 1+1 NNs for all 17 subjects

Acceptability	St. dev	N
22.6%	17.3%	1,000

Next, we consider mean acceptability scores by individual subjects, shown in (6) and Fig. 1. We offer several comments about these results. First, even the highest mean acceptability score is rather low, at just 36.6% (by subject C). Second, there is large variation among the subjects. For instance, the highest score, at 36.6% by subject C, is over 15 times the lowest, at 2.3% by subject M. Tukey multiple comparisons of means (at 95% family-wise confidence level) show significant differences among the majority of the subjects (102/136 = 75%). This is shown in (7) and (8). In addition, the results of ANOVA (analysis of variance), shown in (9), also suggest a significant subject effect on the acceptability scores (p<2e-16 ***). Third, among the scores given by each subject, there is a large standard deviation, suggesting a large variation in acceptability among the 1+1 NNs. Finally, the mean acceptability scores of subjects L, M, O, and P are all below 10%, far lower than those of the others. It is also worth noting that for L, M, O, and P the majority of the 1+1 NNs received a score of 0%, hence there is no corresponding box for these subjects in Fig. 1. 【采取 Mike 意见，将四个小段并成一个大段】

(6) Acceptability scores by subjects

Subject	Mean score	Standard deviation	N
A	29.9%	33.1%	1000
B	16.8%	22.3%	1000
C	36.6%	29.8%	1000
D	31.7%	31.5%	1000
E	24.4%	30.5%	1000
F	27.9%	37.8%	1000
G	29.6%	32.4%	1000
H	21.7%	28.6%	1000
I	28.7%	29.5%	1000
J	26.8%	35.2%	1000
K	33.3%	35.5%	1000
L	8.9%	21.2%	1000
M	2.3%	13.7%	1000
N	32.5%	40.9%	1000
O	6.6%	18.1%	1000
P	7.8%	18.8%	1000
Q	18.4%	29.6%	1000
All	29.9%	33.1%	1000

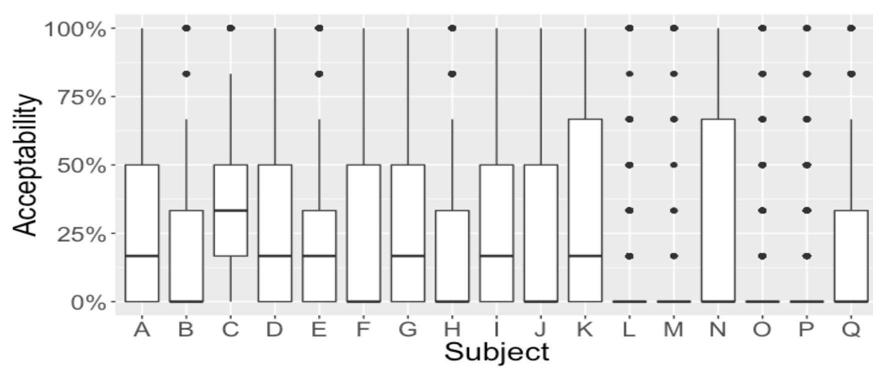

Fig. 1：Boxplot of acceptability score of 1+1 NNs by subjects

(7) Pairwise comparisons between subjects (---, *, **, and *** indicates p-values >0.05, between 0.01 and 0.05, between 0.001 and 0.01, and <0.001 respectively)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
B	***															
C	***	***														
D	---	***	*													
E	**	***	***	***												
F	---	***	***	---	---											
G	---	***	***	---	**	---										
H	***	*	***	***	---	***	***									
I	---	***	***	---	---	---	---	***								
J	---	***	***	*	---	---	---	*	---							
K	---	***	---	---	***	**	---	***	---	***						
L	***	***	***	***	***	***	***	***	***	***	***					
M	***	***	***	***	***	***	***	***	***	***	***	***				
N	---	***	---	---	***	*	---	***	---	**	---	***	***			
O	***	***	***	***	***	***	***	***	***	***	***	---	---	***		
P	***	***	***	***	***	***	***	***	***	***	***	---	**	***	---	
Q	***	---	***	***	***	***	***	---	***	***	***	***	***	***	***	***

(8) Summary of pairwise comparisons between subjects

Difference	---	*	**	***	All
Pairs	34	6	5	91	136

(9) Subject effect on acceptability scores

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Subject	16	180.7	11.297	129.8	<2e-16 ***
Residuals	16983	1478.6	0.087		

4. Other factors

In Qin & Duanmu (2017), it was found that, while length patterns had the most significant effect on the acceptability of constructed NN compounds (in agreement with previous work such as Lü 1963; Wu 1986, 2006; Feng 1996, 2013; Duanmu 1999, 2007, 2012; Wang 2001; Lu & Duanmu 2002), several other factors, ambiguity, boundness, naturalness, and frequency, also played a significant role. In this section, we consider whether these factors have similar effects on the acceptability of 1+1 NNs with a much larger dataset. In addition, we consider the effect of elasticity. We discuss each of the factors below.

4.1. Frequency

Among the 1,000 1+1 NNs in this study, some are frequently used (such as *yào-qián* 'medicine money'), some are less frequently used (such as *yuè-hún* 'moon soul'), and some seem to be entirely new (such as *gōng-fēng* 'palace wind'). We ask whether compounds that are more frequent tend to receive higher acceptability scores. Following Qin & Duanmu (2017), we coded the frequency value of each 1+1 NN according to the BCC corpus [Beijing Language and Culture University Chinese Corpus] (Xun et al., 2016), which contains about 1.5 billon characters. The frequency values of some sample 1+1 NNs are shown in (10).

(10) Frequency values of some 1+1 NNs in the BCC corpus

1+1 NN	Gloss	Frequency
yào-qián	'medicine money'	52
yuè-hún	'moon soul'	17
gōng-fēng	'palace wind'	0
tù-zuì	'rabbit fault'	0
pán-yī	'plate clothes'	0

The results of Pearson correlation analysis show a significantly positive correlation between the frequency and acceptability scores ($r = 0.072$, $t = 2.2798$, $df = 998$, $p = 0.023$), confirming the generalization that the higher the frequency of the word, the higher the acceptability score.

If frequency is treated as a binary variable, i.e. either 'Existing' (with a frequency greater than or equal to 1) or 'Non-existing' (with zero frequency), the results (Welch Two Sample t-test) are similar, as seen in (11). The mean acceptability score of Existing 1+1 NNs is significantly higher than that of Non-existing ones, their difference being 32.2% (a strikingly large value). It is worth noting that among the 1,000 compounds, only 35 actually exist. This means that only a very small portion of freely generated NN compounds are actually used and that the majority of the compounds in our study have never been seen before.

(11) The frequency effect on the acceptability of 1+1 NNs

1+1 NN	Acceptability	St. dev	N	p-value
Existing	53.6%	19.7%	35	2.299e-11
Non-existing	21.4%	16.2%	965	***

4.2. Naturalness

We follow Qin & Duanmu (2017) and use 'naturalness' to refer to whether the referent of a compound is common or natural. In Qin & Duanmu (2017), it was found that compounds whose referents are uncommon tend to be ranked lower; for example, some subjects ranked *xuè-chí* 'blood pond' less acceptable than *shuǐ-chí* 'water pond', apparently because the former is less common (or harder to conceive) than the latter. The concept of naturalness is partly subjective in that it is likely based on one's own experience or perspective. In order to minimize the subjectivity, we used the majority opinion of three native speakers of Chinese who were not the authors of this study, to annotate naturalness. Each annotator independently decided on whether each 1+1 NN is 'natural' or 'unnatural'. Then, a compound was classified as 'natural' if at least two annotators thought so; otherwise, it was classified as 'unnatural'. Some examples are shown in (12).

(12) Examples of natural and unnatural 1+1 NNs

1+1 NN	Gloss	'Natural' scores	Annotation
lù-gōng	'land palace'	3/3	natural
chá-yuàn	'tea yard'	2/3	natural
bì-fēng	'arm wind'	1/3	unnatural
yù-xiǎo	'prison dawn'	0/3	unnatural

The statistical results (Welch Two Sample t-test), given in (13), show that the mean acceptability score of natural 1+1 NNs is significantly higher than that of unnatural ones (p<2.2e-16), with a difference of 25.6%. We can also see that, in our data, unnatural compounds (totaling 838) far exceed natural ones (totaling 162). Moreover, since just 35 of the 1000 compounds have been found in the large BCC corpus (Xun et al., 2016), most of the 162 natural compounds have not occurred before.

(13) The naturalness effect on the acceptability of 1+1 NNs

1+1 NN	Acceptability	St. dev	N	p-value
Natural	44.0%	19.8%	162	< 2.2e-16
Unnatural	18.4%	13.3%	838	***

4.3. Ambiguity

We follow Qin & Duanmu (2017) in defining and annotating 'ambiguity'. The factor refers to whether a written word has a more common meaning, other than the intended one in a compound. The most common meaning of a word is determined by consulting the word frequency list of Guojia Yuwei (2012) and the BCC Corpus (Xun et al., 2016). Some examples are shown in (14).

(14) Sample annotations of ambiguity in 1+1 NNs

1+1 NN	Target N	Intended meaning	Competitor	Annotation
gōng-fēng	gōng	'palace'	(none)	unambiguous
	fēng	'wind'	(none)	unambiguous
měi-yuè	měi	'America'	'beautiful'	ambiguous
	yuè	'music'	(none)	unambiguous

In (14), the most common meaning of *gōng* is its intended meaning 'palace' and is thus classified as 'unambiguous'. The most common meaning of *fēng* is its intended meaning 'wind' and is also classified as 'unambiguous'. The most common meaning of *yuè* is its intended meaning 'music' and is again, classified as 'unambiguous'. However, the most common meaning of *měi* 'beautiful' is not its intended meaning 'America' and is thus classified as 'ambiguous'. We then classified a compound to be unambiguous if both of its nouns are unambiguous; otherwise, we classified a compound to be 'ambiguous'.

The statistical results (Welch Two Sample t-test) are given in (15), which show that the acceptability of unambiguous 1+1 NNs is just 2% higher than that of ambiguous ones, and the difference is not statistically significant (p=0.0706).

(15) The ambiguity effect on the acceptability of 1+1 NNs

1+1 NN	Acceptability	St. dev	N	p-value
Ambiguous	21.8%	17.0%	614	0.0706
Unambiguous	23.8%	17.8%	386	

It can be shown that a four-way ambiguity classification (UU, AU, UA, and AA, where U is an unambiguous noun and A an ambiguous one) would yield the same result, i.e. ambiguity does not have a significant effect on acceptability.

4.4. Boundness

The boundness factor refers to whether a morpheme is free or bound. According to Bloomfield (1933), a morpheme is free if (i) it can occur alone, or (ii) if it can serve as an independent syntactic constituent; otherwise, it is bound. XDHYCD (2005) uses the same definition, where a free morpheme is annotated with a part of speech and a bound form is not. If we follow this definition and label a 1+1 NN as ‘Free’ when both its nouns are free in XDHYCD (2005), and label the compound as ‘Bound’ otherwise, we obtained the statistical results in (16), which show that Free and Bound 1+1 NNs have similar acceptability values; their difference is not statistically significant (p=0.451). In other words, boundness has no effect on the acceptability of 1+1 NNs.

(16) The boundness effect on the acceptability of 1+1 NNs

1+1 NN	Acceptability	St. dev	N	p-value
Free	23.5%	20.0%	189	0.451
Bound	22.3%	16.7%	811	

There is, however, another criterion to determine whether or not a morpheme is free. Following Sproat & Shih (1996: 67), Qin & Duanmu (2017) used the criterion in (17).

(17) Another criterion to determine free morphemes (Sproat & Shih 1996: 67; Qin & Duanmu 2017):
A monosyllabic item is free if it can serve as A in [A *de* B]; otherwise it is bound.

[A *de* B] is a possessive construction in which A is the possessor, B the possessed, and *de* the possessive marker. According to (17), *shuǐ* ‘water’ is free, since it can serve as A in [A *de* B], e.g. *shuǐ de zhòngliàng* ‘water’s weight’, whereas *huá* ‘China’ is bound, since it cannot, e.g. **huá de rénkǒu* ‘China’s population’. It can be seen that the criterion in (17) is more stringent than what is assumed in XDHYCD, because XDHYCD considers both *shuǐ* ‘water’ and *huá* ‘China’ to be free. To determine the effect of (17), we added an analysis, in which we annotated morphemes according to three boundness categories: ‘Bound’, ‘Free1’, and ‘Free2’. A morpheme is ‘Bound’ if it is annotated as bound in XDHYCD (2005); a morpheme is ‘Free1’ if it is annotated as free in XDHYCD (2005) but fails the test in (17); a morpheme is ‘Free2’ if it is annotated as free in XDHYCD (2005) and passes the test in (17). Whether a morpheme passes the criterion in (17) is based on the majority judgment of three native speakers. Some examples are in (18).

(18) Examples of three-way Boundness annotation

Morpheme	Gloss	By XDHYCD	By (17)	Annotation
bāo	‘bag’	free	3/3	Free2
dōng	‘winter’	free	2/3	Free2
wài	‘outside’	free	1/3	Free1
měi	‘America’	free	0/3	Free1
fēng	‘maple’	bound	0/3	Bound
hé	‘lotus’	bound	0/3	Bound

The new annotation creates nine boundness types of 1+1 NNs, shown in (19), along with their acceptability statistics.

(19) Acceptability of nine boundness types of 1+1 NNs

1+1 NN	Acceptability	St. dev	N
BB	0.225	0.167	326
BF1	0.199	0.155	93
BF2	0.233	0.158	133
F1B	0.240	0.190	103
F1F1	0.261	0.250	38
F1F2	0.197	0.147	42
F2B	0.214	0.164	156
F2F1	0.206	0.178	40
F2F2	0.261	0.207	69

The results (one-way ANOVA) show, once again, that there is no significant boundness effect on the acceptability of 1+1 NN compounds ($p=0.268$).

4.5 Elasticity

Elasticity refers to whether a word has elastic length, i.e. having both a monosyllabic form and a disyllabic form. In Qin & Duanmu (2017), the overall acceptability of 1+1 NNs is 47.8%, which is far lower than that of 2+2 or 2+1 NNs, contrary to the widely assumed generalization in (1) that 1+1 NNs are just as good as 2+2 and 2+1 NNs. It is worth noting that all the compounds in Qin & Duanmu (2017) were made of nouns with elastic length (or ‘elastic nouns’). In contrast, many commonly used 1+1 NNs are made of non-elastic nouns, such as *jī-ròu* ‘chicken meat’, where neither *jī* ‘chicken’ or *ròu* ‘meat’ has elastic length. It is reasonable, therefore, to ask whether the lack of bound nouns in Qin & Duanmu (2017) has contributed to the low acceptability score of 1+1 NNs. In other words, we would like to find out whether elasticity has an effect on the acceptability of 1+1 NNs.

To discover the effect of this phenomenon, we annotated the elasticity of all monosyllabic nouns, according to the list of elastic words in Dong (2015). Our annotation yielded four compound types: EE, EN, NE, and NN, where E is a noun with elastic length and N is a noun with non-elastic length.

The results (one-way ANOVA) show that elasticity has a significant effect on the acceptability of 1+1 NNs (p=0.00171). Mean acceptability scores of the four elasticity types are shown in (20) and (21). Among the four types, EE is the least acceptable, although differences between each pair of types are statistically insignificant, except for the difference between EE and EN.

(20) Acceptability of the four elasticity types of 1+1 NNs

Elasticity type	Acceptability	St. dev	N
EE	20.5%	16.5%	484
EN	25.3%	18.2%	195
NE	23.6%	17.9%	220
NN	25.2%	17.1%	101

(21) Tukey multiple comparisons of means (at 95% family-wise confidence level)

	difference	lower range	upper range	p value
EN-EE	0.048	0.011	0.086	0.006
NE-EE	0.031	-0.005	0.067	0.119
NN-EE	0.048	-0.001	0.096	0.055
NE-EN	-0.017	-0.061	0.027	0.745
NN-EN	0.000	-0.055	0.054	1.000
NN-NE	0.017	-0.036	0.070	0.849

The low acceptability of EE is confirmed by an analysis that compares EE with non-EE compounds (Welch Two Sample t-test), given in (22). The result shows that EE 1+1 NNs are significantly less acceptable than non-EE 1+1 NNs (p=0.0001839).

(22) The elasticity effect on the acceptability of 1+1 NNs

Elasticity type	Acceptability	St. dev	N	p-value
EE	20.5%	16.5%	484	0.0001839
Non-EE	24.5%	17.9%	516	***

4.6. A Linear Mixed-Effects Model analysis of different factors

In the discussions above, we have separately examined the effects of frequency, naturalness, ambiguity, boundness, and elasticity on the acceptability of 1+1 NNs. In this section, we consider their effects together by fitting a Linear Mixed-Effects Model of the relationship between acceptability scores and all the relevant factors, using RStudio (2012) and lme4 (Bates et al. 2015). Specifically, we included naturalness, ambiguity, boundness, frequency, and elasticity as fixed effects, and subjects and items as random effects. We used both frequency and boundness as binary factors: for the former, we distinguish 'Existing' vs. 'Non-existing' 1+1NNs, and for the latter, Free (consisting of free morphemes only) vs. Bound (consisting of one or two bound morphemes) 1+1 NNs. The fitted model is shown in (23).

(23) Score ~ Frequency + Naturalness + Boundness + Ambiguity + Elasticity + (1 | Subject) + (1 | Item)

The main results are shown in (24). It can be seen that naturalness, frequency, and elasticity are significant predictors of acceptability scores, while ambiguity and boundness are not. Specifically, all else being held constant, the acceptability of non-existing 1+1 NNs is 21.9% lower than that of existing ones; the acceptability of natural 1+1 NNs is 22.9% higher than that of unnatural ones; and the acceptability of EE 1+1 NNs is 2.51% lower than that of non-EE ones (EN, NE, and NN compounds as a group). We also see that naturalness and frequency have much greater effect sizes than elasticity.

(24) The results of fixed effects in a Linear Mixed-Effects model

	Estimate	Std. error	df	t.value	Pr(>|t|)
(Intercept)	3.93e-01	3.61e-02	7.20e+01	10.886	<2e-16***
FrequencyN	-2.19e-01	2.47e-02	9.99e+02	-8.872	<2e-16***
NaturalnessY	2.29e-01	1.23e-02	9.99e+02	18.685	<2e-16***
BoundnessF	4.03e-04	1.14e-02	9.99e+02	0.036	0.97167
AmbiguityY	-9.87e-03	9.15e-03	9.99e+02	-1.079	0.28103
ElasticityN	2.51e-02	8.90e-03	9.99e+02	2.823	0.00485**

5. Explaining the low acceptability of 1+1 NN compounds

In this section, we explore explanations for the low acceptability of 1+1 NN compounds. We start with a comparison between the present results and those of Qin & Duanmu (2017). Then we consider three additional parameters: semantic transparency (Reddy et al. 2011; Bell & Schäfer 2016; Schäfer 2018), grammatical relations between the two constituents of a compound (Bisetto & Scalise 2005; Ceccagno & Scalise 2006; Ceccagno & Basciano 2009), and modern vs. classic vocabulary.

5.1. Comparison between the present results and those of Qin & Duanmu (2017)

There is a stark difference in the acceptability rate of 1+1 NN compounds in the present study and that of Qin & Duanmu (2017); this is summarized in (25).

(25) Acceptability of 1+1 NNs as found in two studies

	Acceptability	St. dev	N
Qin & Duanmu (2017)	47.8%	18.6%	283
Present study	22.6%	17.3%	1,000

The main reason, we suggest, lies in a difference in the methods. In Qin & Duanmu (2017), the focus was on the word length effect. Therefore, after compounds were created through free combinations of NN, a pre-screening process was applied, in which compounds that were ‘clearly uninterpretable’ were excluded (Qin & Duanmu 2017). For example, ‘wheat price’ was kept in but ‘price wheat’ was screened out. The actual test list, therefore, consisted of only those compounds that have conceivable interpretations. In contrast, in the present study, there was no pre-screening, and the test list consisted of a random selection from all freely constructed compounds, nearly half of which were uninterpretable. As we shall see in section 5.2, unintelligible compounds have very low mean acceptability ranking. This seems to explain most of the difference in (25).

5.2. Semantic transparency

Reddy et al. (2011), Bell & Schäfer (2016), and Schäfer (2018) have proposed a gradient notion of semantic 'transparency', which is equated with literality, i.e. how literal a compound and its constituents are to be understood in this compound. We ask, therefore, whether semantic transparency has an effect on acceptability. It is possible that the more transparent a compound is, the more acceptable it is, because it takes more efforts to interpret an opaque compound than a transparent one on the part of the subjects, who may, in turn, give a lower rating score for the former than for the latter.

However, previous discussions of semantic transparency focused on interpretable compounds, whereas in our study, many compounds are uninterpretable. Therefore, we decided to verify transparency in two steps: (i) we consider the effect of interpretability on acceptability, and (ii) among compounds that are interpretable, we consider the effect of transparency on acceptability.

We randomly selected 200 1+1 NNs from the 1,000 used in this study. Then, we annotated each compound for its interpretability and its transparency, using three independent Chinese native speakers (excluding the authors). We labeled a compound 'uninterpretable' if all three annotators labeled it so; otherwise, we labeled a compound 'interpretable'. Next, for each interpretable compound, we asked the same three native speakers to offer its most natural/likely definition. Finally, we asked the three native speakers to judge whether each compound is to be understood literally, given the most natural/likely definition associated with it. We labeled a compound to be transparent if two or three subjects determine that it is to be understood literally, otherwise we labeled it to be 'opaque'.

Of the 200 randomly selected 1+1 NNs, 112 are interpretable. The statistical results (Welch Two Sample t-test) given in (26), suggest that there is a significant effect of interpretability on the acceptability.

(26) The interpretability effect on the acceptability of 1+1 NNs

Interpretability type	Acceptability	St. dev	N	p-value
Interpretable	30%	16.2%	112	< 2.2e-16 ***
Uninterpretable	12.4%	9%	88	

Next, we consider transparency. The statistical results (Welch Two Sample t-test) given in (27), show no significant effect of transparency on the acceptability of 1+1 NNs.

(27) The transparency effect on the acceptability of 1+1 NNs

Transparency type	Acceptability	St. dev	N	p-value
Transparent	30.4%	16%	107	0.3496
Opaque	21%	19.8%	5	

It can be shown that a four-way categorization distinguished by the number of 'literal' judgment for a compound (0, 1, 2, or 3) would yield the same result, i.e. literality does not have a significant effect on acceptability. It is worth noting that 107 of the 112 interpretable 1+1 NNs are transparent and only 5 of them are opaque. Thus, given the unbalanced sample sizes, the result does not seem to be compelling. ['result does']

5.3. Grammatical relations

A reviewer suggested that we consider the grammatical relations between the two nouns of a compound and its effect on acceptability. According to Bisetto & Scalise (2005), Ceccagno & Scalise (2006), and Ceccagno & Basciano (2009), NN compounds can be classified into one of the three types: subordination, coordination, and attribution. The present authors classified each interpretable compound into one of such three types based on its most natural/likely definition. Only subordinate words, e.g. 'land palace' (a palace on land), and attributive words, e.g. 'poison seedling' (poisonous seedling), were found; no coordinate compound was found in our data. The statistical results (Welch Two Sample t-test) are given in (28), which show that there is no significant effect of compound type on the acceptability of 1+1 NNs.

---

(28) The compound type effect on the acceptability of 1+1 NNs

Compound type	Acceptability	St. dev	N	p-value
Attributive	31.5%	16.7%	13	0.7268
Subordinate	29.8%	16.2%	99	

It is worth noting that 99 of the 112 interpretable 1+1 NNs are subordinate compounds and just 13 are attributive compounds. Given the unbalanced sample sizes, the results once again need to be taken with caution.

5.4. Classic vs. modern vocabulary

A reviewer suggests that it is possible that 1+1 NNs are acceptable or productive in classical or technical language, such as *yǐ-wáng* 'ant king' (queen ant) and *gū-sǎn* 'pileus' (mushroom umbrella) as cited in Sproat & Shih (1996), but not in ordinary/colloquial speech. In contrast, our judgment experiments seem to address people's intuitions about ordinary/colloquial speech. This mismatch of styles might explain why the acceptability rating of 1+1 NNs is so low.

As discussed in section 2 above, our data did not include compounds made of nouns of restricted styles, such as those labeled by XDHYCD as 'archaic', 'dialectal', or 'colloquial'. Therefore, we do not have enough data to support a statistical analysis of the style effect. Still, we would like to offer some speculations. It seems unlikely that style is a major factor in explaining the low mean acceptability ranking of our data. For example, a close look at the test compounds shows that many of them can occur in technical speech, such as *yuan-gū* 'garden mushroom' and *qiè-zé* 'concubine responsibility', yet the mean acceptability rating of them does not seem to be high. In addition, there are many non-classic and non-technical compounds, such as *méi-diàn* 'coal store' and *jì-gōng* 'skill worker, which are highly acceptable. Therefore, we leave the verification of this suggestion for future research.

6. Concluding remarks

In this study, we have found that the acceptability of Chinese compounds made by freely combining two monosyllabic nouns (1+1 NNs) is very low, at a mean of just 22.6%, confirming the preliminary report in Qin & Duanmu (2017). The result contradicts a popular generalization in (1), according to which 1+1 NNs are well formed in Chinese (Wu 1986, 2006; Feng 1996, 2013; Duanmu 1999, 2007, 2012; Wang 2001; Lu & Duanmu 2002; Zhou 2011, 2017; Zhang 2014; among others).

Given the fact that 1+1 NNs are the most frequent compound type among all occurring compounds (Duanmu 2012), it remains to be explained why the mean acceptability of freely combined 1+1 NNs is so low. Three factors are found to have a significant effect on acceptability: naturalness of the referent, frequency of the compound, and boundness of the component nouns. However, the most important explanation seems to be that most freely combined 1+1 NNs are simply not interpretable. In other words, despite the fact that 1+1 NNs are the most common form of Chinese compounds, they still just constitute a very small fraction of all possible combinations of 1+1 NNs, most of which are uninterpretable.

We have also considered three other explanations of the low rating of 1+1 NNs: semantic transparency, grammatical relation between the two nouns of a compound, and classic vs. modern vocabulary. None of them seems to be a better explanation than the fact that many of the freely combined 1+1 NNs in our study are hardly interpretable.

In summary, the generalization on word length effect, shown in (1), remains true among occurring compounds in Chinese. On the other hand, for freely combined 1+1 NN compounds, interpretability seems to be the most important factor in determining their acceptability. Furthermore, most of these freely combined 1+1 NN compounds are judged as unintelligible, and hence have low acceptability rankings. This means that, as far as compounds are concerned, high occurrence frequency (among all compound length types) does not guarantee high productivity (or high acceptance rate for freely created compounds).

***Acknowledgments:**

This study is supported by the Fundamental Research Funds for the Central Universities [Grant No. 2018SQN17] and Sichuan Social Sciences Planning Project [Grant No. SC17B107]. Part of this paper was presented at The 5th International Conference on Chinese Prosodic Grammar at Fudan University. We would like to thank the audience for their comments. We would also like to thank the editor of *Lingua* and anonymous reviewers for their comments, and Dr. Michael Opper for proofreading the final manuscript of this paper.

References

Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. lme4: Linear Mixed-Effects Models Using Eigen and S4. R Package Version 1.1-12.

Bell, M., Schäfer M., 2016. Modelling semantic transparency. Morphology 26(2): 157-199.

Bisetto, A., Scalise, S., 2005. Classification of compounds. Lingue e linguaggio 4: 319-332.

Bloomfield, L., 1933. Language. New York: Holt, Rinehart & Winston.

Ceccagno, A., Basciano, B., 2009. Sino-Tibetan: Mandarin Chinese, in: Lieber R., Štekauer P. (eds), The Oxford Handbook of Compounds. Oxford: Oxford University Press, 478-490.

Ceccagno, A., Scalise, S., 2006. Classification, structure and headedness of Chinese compounds. Lingue e linguaggio 2: 233-260.

Da, J., 2004. Xiandai hanyu danzi pinlü biao (Character frequency list of Modern Chinese). http://lingua.mtsu.edu/chinese-computing/statistics/ (accessed on Feb 22, 2018)

Dong, Y., 2015. The Prosody and Morphology of Elastic Words in Chinese: Annotations and Analyses (doctoral dissertation). University of Michigan, Ann Arbor.

Duanmu, S., 1999. Zhongyin lilun he Hanyu de cichang xuanze (Metrical theory and word length choices in Chinese). Zhongguo Yuwen 4, 246-254.

Duanmu, S., 2007. The Phonology of Standard Chinese, 2nd edition. Oxford: Oxford University Press.

Duanmu, S., 2012. Word-length preferences in Chinese: a corpus study. J. East Asian Linguist. 21 (1), 89-114.

Duanmu, S., Feng, S., Dong, Y., & Zhang, Y., 2018. A judgment study of length patterns in Chinese: prosody, last resort, and other factors. Journal of Chinese Linguistics, 46(1), 42-68.

Feng, S., 1996. Lun Hanyu de yunlüci (On the prosodic word in Chinese). *Zhongguo Shehui Kexue* 1, 161-176.

Feng, S., 2013. Hanyu yunlü jufaxue (zengding ben), (Chinese Prosodic Syntax). expanded edition. Shangwu Yinshuguan, Beijing.

Guo, S., 1938. Zhongguo yuci zhi tanxing zuoyong (The function of elastic word length in Chinese). *Yanjing Xuebao* 24, 1-34.

Guojia Yuwei [State Language Commission], 2012. Xiandai Hanyu Yuliaoku Ciyu Fencilei Pinlü Biao (A Corpus-Based Frequency List of POS-Distinguished Words in Modern Chinese). www.cncorpus.org (accessed on Oct 5, 2016).

Karlgren, B., 1918. Ordet och pennan i Mittens Rike. Stockholm: Svenska Andelsförlaget.

Ke, H., 2007. A Study of Monosyllabic and Disyllabic Usage in Modern Chinese (doctoral dissertation). Institute of Linguistics, Chinese Academy of Social Sciences, Beijing.

Lu, B., Duanmu, S., 2002. Rhythm and syntax in Chinese: a case study. J. Chin. Lang. Teach. Assoc. 37 (2), 123-136.

Lü, S., 1963. Xiandai Hanyu dan shuang yinjie wenti chu tan (A preliminary study of the problem of monosyllabism and disyllabism in modern Chinese). *Zhongguo Yuwen* 1, 10-22.

Qin, Z., Duanmu, S., 2017. A judgment study of word-length preferences in Chinese NN compounds. Lingua 198:1-21.

Reddy, S., McCarthy, D., & Manandhar, S., 2011. An empirical study on compositionality in compound nouns. In *Proceedings of the 5th international conference on natural language processing*, AFNLP, Chiang Mai, Thailand (pp. 210–218).

RStudio, 2012. RStudio: Integrated Development Environment for R (Version 0.99.902) [Computer Software]. Boston, MA. Retrieved May 13, 2016. Available from http://www.rstudio.org/.

Schäfer, M., 2018. The semantic transparency of English compound nouns. Berlin: Language Science Press.

Shen, J., 2012. 'Mingdongci' de fansi: wenti he duice (Reflections on 'Nouny Verbs': Problems and Solutions). Shijie Hanyu Jiaoxue 26(01):3-17.

Shen, J., 2015. Hanyu cilei de zhuguanxing (On subjectivity of Chinese word classes). Waiyu Jiaoxue Yu Yanjiu 47(05):643-658.

Sproat, R., Shih, C., 1996. A corpus-based analysis of Mandarin nominal root compound. J. East Asian Linguist. 5 (1), 49-71.

Wang, H., 2001. Yinjie danshuang yinyu zhanlian (zhongyin) yu yufa leixing jiegou he chengfen cixu (The relations between the number of syllable, the tonal range of pitch and the grammatical structure in Chinese). Dangdai Yuyanxue 4, 241-252.

Wang, C., 2002. In Jufa zuhe zhong de danshuang yinjie xuanzhe de rengzhi jieshi (A cognitive account of the choice between monosyllabic and disyllabic words in syntactic combination). In: Zhongguo Yuwen Zazhi She (Eds.), Yufa yanjiu he tansuo, (Studies on Grammar).vol. 11.Shangwu Yinshuguan, Beijing, pp. 151-168.

Wu, W., 1986. Xiandai hanyu sanyinjie zuhe guilü chutan Hanyu yunlü (A preliminary study of collocation rules in trisyllabic sequences in Chinese). Shanghai: Xuelin Chubanshe.

Wu, W., 2006. Hanyu yunlü jufa tansuo (A Study of Prosodic Grammar). Shanghai: Xuelin Chubanshe.

XDHYCD., 2005. Xiandai Hanyu Cidian (Modern Chinese Dictionary. Compiled by the Institute of Linguistics, Chinese Academy of Social Sciences). 5th edition. Beijing: Shangwu Yinshuguan.

Xun, E., Rao, G., Xiao, X., Zang, J., 2016. Da shuju beijing xia BCC yuliaoku de yanzhi (The construction of the BCC Corpus in the age of Big Data). Yuliaoku Yuyanxue 3 (1), 93-118.

Zhang, H., 2014. Yunlü yinxixue yu hanyu yunlü yanjiu zhong de ruogan wenti (Issues on prosodic phonology and Chinese prosodic grammar). Dangdai Yuyanxue 16(03): 303-327.

Zhou, R., 2011. Xiandai hanyu yunlü yu yufa de hudong guanxi yanjiu (A study of the interaction between prosody and grammar in Chinese). Beijing: Shangwu Yinshuguan.

Zhou, R., 2017. Hanyu yunlü yufa yanjiuzhong de qingzhong xiangsi, songjin xiangsi he duoshao xiangsi (The principles of weight, elasticity, and quantity iconicity in Chinese prosodic grammar). Zhongguo Yuwen 5, 536-552.