Adjective-Noun Compounds in Mandarin: A Study on Productivity

Shen Tian and R. Harald Baayen

February 17, 2021

Abstract

In structuralist linguistics, compounds are argued not to constitute morphological categories, due to the absence of systematic form-meaning correspondences. This study investigates subsets of compounds for which systematic form-meaning correspondences are present: adjective-noun compounds in Mandarin. We show that there are substantial differences in the productivity of these compounds. One set of productivity measures (the count of types, the count of hapax legomena, and the estimated count of unseen types) reflect compounds' profitability. By contrast, the category-conditioned degree of productivity is found to correlate with the internal semantic transparency of the words belonging to a morphological category. Greater semantic transparency, gauged by distributional semantics, predicts greater category-conditioned productivity. This dovetails well with the hypothesis that semantic transparency is a prerequisite for a word formation process to be productive.

keywords
morphological category, morphological productivity, distributional semantics, Mandarin adjective-noun compounds, semantic transparency

1 Introduction

This study presents a quantitative investigation of the productivity of adjective-noun compounds in Mandarin and addresses the question of how a word formation process that subserves the creation of names for things and events in an idiosyncratic, non-compositional world can be productive.

The present study focuses on the productivity of compounding. From a classical structuralist perspective, compounding is special in that compounds do not form a morphological category (Schultink 1961), as compounds are taken not to have in common shared aspects of form and meaning. Furthermore, the classical tools with which degrees of productivity are explained, namely, phonological, morphological, semantic, and pragmatic restrictions on base words (Booij 1977; Aronoff and Fudeman 2011) as well as output restrictions (Plag 2003), are not straightforwardly applicable to compounding. A survey of five textbooks (Lieber 2010; Plag 2003; Aronoff and Fudeman 2011; Booij 2010; Bauer 2001) reveals that differences in the productivity of derivational affixes are discussed, but that no theoretical reflection is provided for differences in productivity for compounding. In general, it might be argued that compounds straddle the boundary between morphology and syntax, and that they partake in the productivity of the syntax. Given the definition of Plag (2010) of a compound, namely as a word that consists of two elements, of which is either a root, a word, or a phrase, it would seem that there are no principled constraints on the productivity of compounding.

The aim of the present study is to show the productivity of Mandarin adjective-noun compounding is driven by semantic transparency. In Mandarin, adjective-noun constructions typically subserve what Kastovsky (1986) calls the labeling function, and what we henceforth refer to as the naming function. Many of the adjective-noun compound names have idiosyncratic meanings, simply because the things in the world they name for do not fall into straightforward logical ontology. Nevertheless, using distributional semantics, we are able to show that Mandarin adjectives are more productive in adjective-noun compounds when the semantic similarity of the adjective to its compounds is greater. In other words, we will argue that adjective-noun compounds in Mandarin constitute morphological categories, and that these morphological categories have their own degrees of productivity that vary with their semantic properties.

In English, compounds account for some 20% of the lexemes. A count based on the CELEX database (Baayen et al. 1996) shows that 10,726 out of 52,447 lexemes contain at least two stems. In Mandarin, compounding accounts for some 70-80% of all words (Institute of Language Teaching and Research). In a study of neologisms, compounds take some 95%, of which nearly three quarters are adjective-noun formations (Ceccagno and Basciano 2007).

A preliminary question that needs to be addressed is whether adjective-noun formations are compounds or rather phrases (see, e.g., Booij 2010, for detailed discussion). In Mandarin, phrases consisting of an adjective and a noun require insertion of the possessive marker 的 *de* (see Table 1 in the supplementary materials for examples[1]). Such phrases are always fully transparent, they are not listed in dictionaries, and they do not serve the naming function. Most adjective-noun formations that do not contain 的 *de* tend to be listed in dictionaries. More often, they function as labels or names in which case they may differ substantially in semantic transparency (e.g., 大家, *da4jia1*, big family, 'everyone'; 大碗, *da4wan3*, big bowl, 'big bowl'). In addition, the adjective in adjective-noun compounds cannot undergo further modification, whereas this is unproblematic for adjective-noun phrases with *de*. For further evidence that adjective-noun combinations in Mandarin are words, see Xu (2018).

When the noun in an adjective-noun compound is monosyllabic, the compound typically serves the naming function. For the base nouns, which typically have many different meanings when used in isolation, the adjective modifies only one of these meanings. For instance, the word 象 (*xiang4)* has meanings varying from 'image', 'figure' to 'elephant'. When it combines with 大 (*da4*, 'big'), the adjective-noun construction 大象 (*da4 xiang4*) singles out the meaning of elephant, without however specifying the size of the elephant. To specify that a particular elephant is big or small, further adjectival modification is required using the phrasal construction: 大的大象, (*da4 de da4xiang4*, 'big elephant'); 小的大象, (*xiao3 de da4xiang4*, 'small elephant'). Thus, the adjective functions as a discriminative device to single out one of the senses of 象, resulting in a compound

[1] https://osf.io/wq57p/?view_only=181bef0617324fd5a3bdbe01e168e45b

that is unambiguous (see Table 2 in supplementary materials for further discussion). Moreover, according to Huang et al. (2017), polar adjectives are also regarded as state intransitive verbs, therefore, adjective-noun compounds may be interpreted as intransitive verb-noun compounds as well.

Informed by the Chinese High Frequency Dictionary (Institute of Language Teaching, Beijing Language and Culture Institute 1986) and the Chinese National Corpus (`http://corpus.zhonghuayuwen.org/`), in the present study we investigated 56 most frequent monosyllabic polar adjectives in Mandarin that occur in nominal adjective-noun compounds. Since the overwhelming majority (i.e. 73.6 %) of Modern Chinese are disyllabic words (Institute of Language Teaching and Research 1986), the focus of the present study is two-character nominal adjective-noun compounds. Figure 1 compares the Mandarin adjectives with their approximate English translation equivalents. The horizontal axis represents the number of adjective-noun compounds in English listed in the CELEX lexical database, and the vertical axis represents the counts of nominal adjective-noun compounds that occur in the Chinese National Corpus. There is a positive correlation between type frequency in English and in Mandarin, highlighted by the red regression line ($\hat{\beta} = 0.39, p < 0.0001$). The blue line in Figure 1 represents the line $y = x$. Most of the data points are located above this line, indicating that type counts are larger for Mandarin (in total 2,046 types) as compared to English (in all 232 types), an observation that dovetails well with the greater overall productivity of compounding in Mandarin.

In the remainder of this study, we address two clusters of research questions. First, does a quantitative evaluation of the productivity of nominal adjective-noun compounds support significant differences in productivity between the adjectives? In other words, do adjectives in Mandarin take on a role similar to that of prefixes in English such as *un-*? This set of interrelated questions is addressed in section 2 and section 3. Second, given differences in productivity, is it helpful to characterize Mandarin nominal adjective-noun compounds as morphological categories? What makes these morphological categories productive? Classical analyses of the productivity of derivational affixes have focused on restrictions on the input and output of the corresponding word formation rules. However, for Mandarin, we do not know of any phonological, morphological, semantic, or pragmatic restrictions on adjective-noun compounding. Are the differences in productivity simply reflecting differences in cultural popularity, or is there a language-internal factor at play? This cluster of questions is addressed in section 4.

2 Measuring productivity for Mandarin adjective-noun compounds

The morphological productivity of a word formation rule, or equivalently, a morphological category, is generally understood as the rule being able to give rise to new words (Plag 2003). Kruisinga (1932) distinguished between productive and unproductive morphological categories, and characterized in terms of "living" and "dead" suffixes. Bolinger (1948) proposed a more gradient approach to productivity, according to which productivity is the "statistical readiness with which an element enters into new combinations". Aronoff (1976)'s definition of morphological productivity as the ratio of possible words (i.e. a word which can be created in accordance with certain phonological,

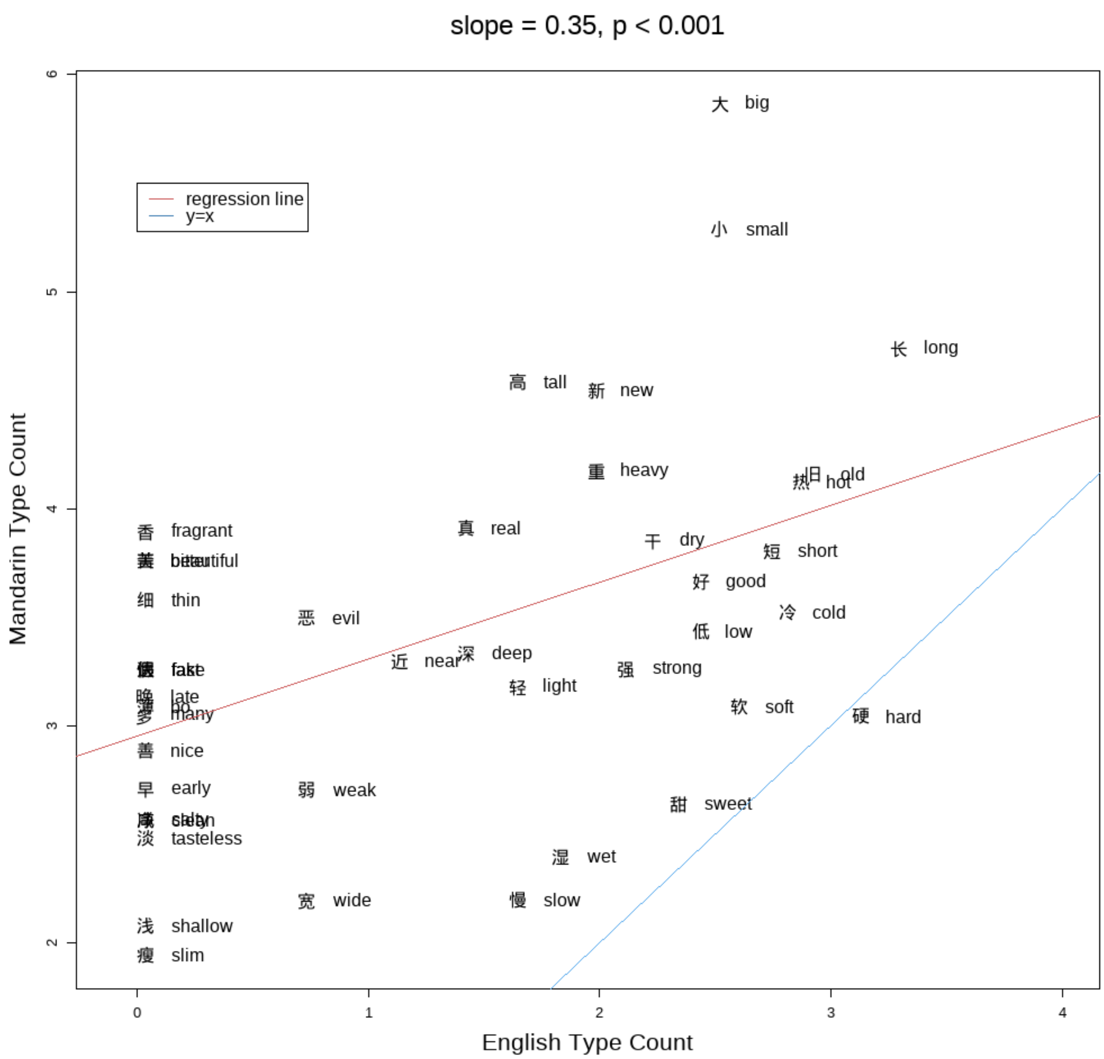

Figure 1: Comparison of type counts for Mandarin nominal adjective-noun compounds and their English counterparts. In parenthesis, the slope of the regression line and the corresponding p-value are listed.

semantic, and morphological rules) to actual words stands in this tradition. In the present study, we adopt statistical formalization of aspects of productivity, conceptualized as an intrinsically gradient phenomenon, reviewed in Baayen (2009).

The first aspect of productivity, henceforth "realized productivity", captures one aspect of what Corbin (1987) and Bauer (2001) refer to as "profitability", i.e., the quantitative aspect of productivity, namely, the extent to which a word formation process has already been used. Realized productivity quantifies past usage by means of the count of types $V(A, N_A)$ observed for adjective-noun compounds with adjective A in a corpus with N word tokens.

As can be seen in Table 1 (for the GIGP and FZM measures, see section 3), there are substantial differences in realized productivity for adjectives, with counts ranging from 2 (忙, *mang2*, 'busy') to 351 for 大 (*da4*, 'big'). A chi-squared test on the type counts provided strong support for between-adjective differences in realized productivity ($\chi^2_{(44)} = 3455.4, p < 0.0001$). Thus, even though we are studying compounds rather than derivational affixes, we nevertheless observe clear differences in realized productivity.

However, the problem with a simple type count is that it only presents what words are in use, without providing insight into the probability of coinages in the future. For example, the English suffix *-ment* has many derivatives that were created centuries ago and that are still in use. Nevertheless, speakers now hardly create new forms with it. In spite of its fairly high type count, *-ment* is generally regarded as unproductive (Bauer 2001).

In order to capture the degree to which a rule is productive in the sense that it is available for creating new words, several quantitative measures are available. Two of these measures make use of the count of the lowest-frequency words, the words which occur only once in a corpus, the so-called hapax legomena (henceforth hapaxes). The number of hapaxes for adjective A in a corpus with N tokens, $V(1, A, N_A)$, is proportional to the extent to which adjective-noun compounds with A contribute to the rate at which the total vocabulary increases as word tokens are sampled. The hapax-conditioned degree of productivity $\mathcal{P}^* = V(1, A, N_A)/H(N)$ which takes the number of hapaxes found for A in a corpus of size N, and divides this count by the total number of hapaxes $H(N)$ for all types in the corpus. As for a given N, $H(N)$ is a constant across morphological categories, so this productivity measure simply compares the number of hapaxes observed for the different adjectives. As can be seen in Table 1, counts of hapaxes range from 51 for 大 (*da4*, 'big') to 0 for 饱 (*bao3*, 'full'). Figure 2 illustrates that, on a logarithm scale, the number of types $V(A, N)$ and hapax $V(1, A, N_A)$ have a strong positive correlation for our data ($\hat{\beta} = 0.9, p < 0.0001$). Nevertheless, when specifically zooming in on what information $V(1, A, N_A)$ provides over and above $V(A, N)$, we see that, for instance, for roughly the same realized productivity $V(A, N)$, 硬 (*ying4*, 'hard') has a rather low value of hapaxes *V(1, 硬, N)* than *V(1, 厚, N)*. Thus, when a token is sampled and added to the corpus, and given that this token represents a hitherto unseen type, it is more likely that this token is an adjective-noun compound with 厚 (*hou4*, 'thick') than that it is an adjective-noun compound with 硬 (*ying4*, 'hard') .

It is possible that a morphological category contributes little to the growth rate of the vocabulary as a whole, but nevertheless is well available for further word formation. For example, the Dutch suffix *-ster* is not used very often, but new words with it denoting female agents are unproblematic (Baayen 1994). To assess productivity with respect to the morphological category itself, we can

Table 1: Overview of productivity measures on adjectives in Mandarin nominal adjective-noun compounds.

English Translation	Chinese	Hapax ($V1, A, N_A$)	Type (V, A, N_A)	Token N	$\mathcal{P}(A, N_A)$	log$\mathcal{P}(N, N_A)$	GIGP $V(0, A, N_A)$	FZM $V(0, A, N_A)$
big	大	51	351	24668	0.00	-6.18	47	46
small	小	28	197	8404	0.00	-5.70	27	28
long	长	17	113	3148	0.01	-5.22	26	21
tall	高	20	97	5547	0.00	-5.63	886	1347
new	新	22	93	3692	0.01	-5.12	51	37
heavy	重	19	64	2455	0.01	-4.86	8470	25301
old	旧	17	63	428	0.04	-3.23	26	63
hot	热	15	61	2023	0.01	-4.90	33	22
fragrant	香	19	48	785	0.02	-3.72	40264	283916
real	真	11	49	1855	0.01	-5.13	22	15
dry	干	14	46	3845	0.00	-5.62	4913	15962
short	短	4	44	601	0.01	-5.01	3	3
beautiful	美	11	42	2678	0.00	-5.49	1784	5283
bitter	苦	10	42	629	0.02	-4.14	24	179
good	好	7	38	1782	0.00	-5.54	9	100
thin*	细 *	9	35	3573	0.00	-5.98	4676	6549
cold	冷	6	33	546	0.01	-4.51	5	98
evil	恶	3	32	458	0.01	-5.03	1	154
low	低	4	30	1007	0.00	-5.53	3	2
deep	深	6	27	1434	0.00	-5.48	16	121
fake	假	5	25	726	0.01	-4.98	7	5
far	远	3	25	798	0.00	-5.58	NA	NA
near	近	4	26	324	0.01	-4.39	3	2
fast	快	5	25	444	0.01	-4.49	5	9
young	少	7	25	2108	0.00	-5.71	3822	56
strong	强	5	25	1325	0.00	-5.58	394	44
light	轻	4	23	406	0.01	-4.62	4	123
bo*	薄 *	4	21	255	0.02	-4.15	6	4
late	晚	3	22	600	0.00	-5.30	6	4
hou*	厚 *	9	20	276	0.03	-3.42	535054	51
many	多	4	20	853	0.00	-5.36	2	2
soft	软	5	21	267	0.02	-3.98	7	87
hard	硬	2	20	261	0.01	-4.87	518	1
nice	善	9	17	233	0.04	-3.25	46730	426141
thick*	粗 *	1	18	217	0.00	-5.38	NA	NA
idle	闲	4	16	179	0.02	-3.80	NA	NA
early	早	5	14	296	0.02	-4.08	1126	9380
weak	弱	6	14	244	0.02	-3.71	12	9
sweet	甜	2	13	163	0.01	-4.40	1	1
clean	净	5	12	55	0.09	-2.40	18	11
salty	咸	5	12	60	0.08	-2.48	433068	11673065
bad	坏	1	11	417	0.00	-6.03	NA	NA
wet	湿	2	10	255	0.01	-4.85	2	2
tasteless	淡	5	11	166	0.03	-3.50	7	NA
wide	宽	2	8	134	0.01	-4.20	1261	1089
dirty	脏	1	8	157	0.01	-5.06	NA	NA
smelly	臭	1	8	171	0.01	-5.14	NA	NA
ugly	丑	0	7	63	0.00	NA	NA	NA
slow	慢	2	8	50	0.04	-3.22	1	51
shallow	浅	1	7	149	0.01	-5.00	1	30
slim	瘦	1	6	48	0.02	-3.87	1	1
narrow	窄	2	5	17	0.12	-2.14	NA	NA
hungry	饿	1	5	15	0.07	-2.71	NA	NA
fat	胖	0	4	130	0.00	NA	NA	NA
full	饱	0	3	35	0.00	NA	NA	NA
busy	忙	1	2	6	0.17	-1.79	NA	NA

* 薄 (*bo2*) and 厚 (*hou4*) are used for attributing 'thinness' and 'thickness' only for flat objects such as books and sheets of paper, whereas, 细 (*xi4*) and 粗 (*cu1*) are used to describe the 'thinness' and 'thickness' of cylinder-like objects such as tree trunks or ropes. NA: estimates not available.

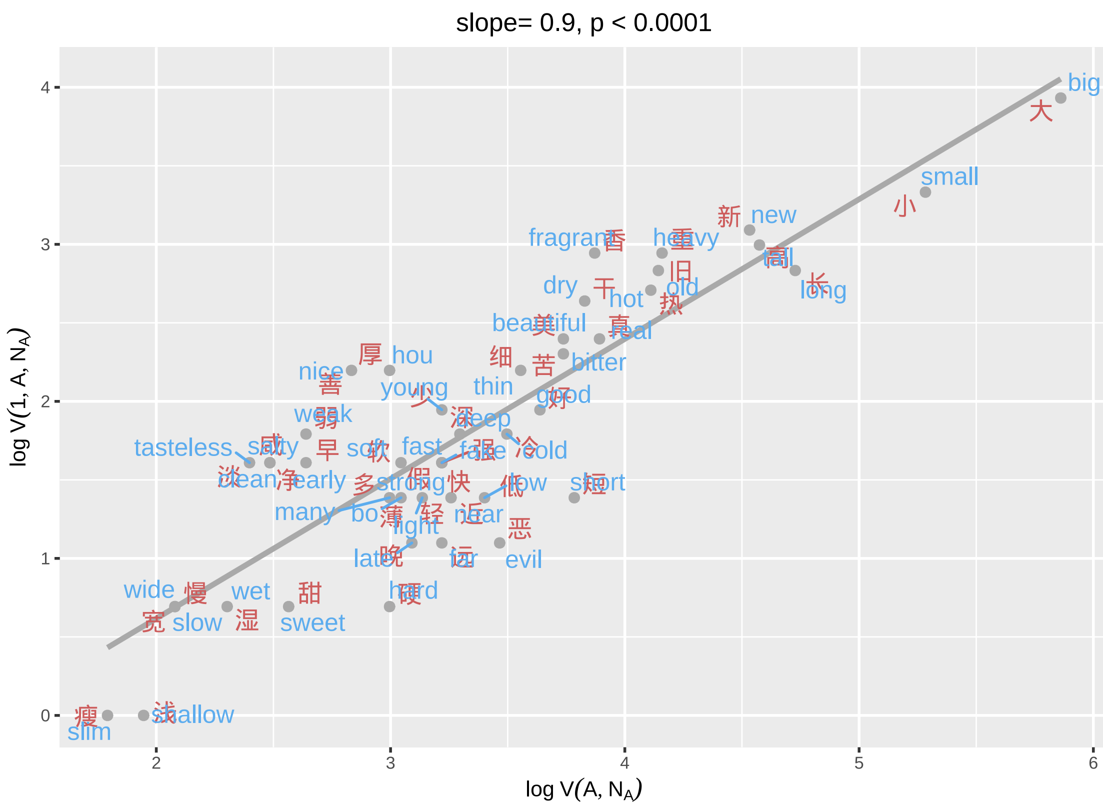

Figure 2: Scatterplot of adjectives in the plane of $\log V(A, N_A)$ and $\log V(1, A, N_A)$. The grey line is the least squares regression line.

calculate the probability $\mathcal{P}(A, N_A)$ that a token containing adjective A is an adjective-noun type that has not been seen before given that we have seen N_A tokens of this kind of compound:

$$\mathcal{P}(A, N_A) = V(1, A, N_A)/N_A.$$

This category-conditioned degree of productivity focuses exclusively on the types and tokens of adjective A. It captures the potential probability of encountering a new occurrence with a given adjective among all the words containing that adjective. Figure 3 illustrates the negative correlation ($\hat{\beta} = -0.59, p < 0.0001$) between type counts $V(A, N_A)$ and $\mathcal{P}(A, N_A)$ on a logarithm scale for Mandarin nominal adjective-noun compounds. What is at first sight surprising is that the adjective 大 (*da4*, 'big'), which is characterized by large numbers of types and hapaxes, shows up with the smallest value of $\mathcal{P}(A, N_A)$. Conversely, 瘦 (*shou4*, 'slim'), that shows up with few types and hapaxes, reveals high value of $\mathcal{P}(A, N_A)$. There are several possible explanations for this negative correlation between $V(A, N_A)$ and $\mathcal{P}(A, N_A)$.

First, from a statistical perspective, this negative correlation is unsurprising. The number of compound types $V(A, N_A)$ observed for N_A tokens of compounds with A is a monotonically increasing function of N_A. As we sample more tokens, we observe more types; but the rate at which new types are observed decreases. Since $\mathcal{P}(A, N_A)$ quantifies the rate at which the vocabulary size $V(A, N_A)$ increases, it necessarily decreases as N_A and $V(A, N_A)$ increase. What is remarkable for the present set of Mandarin adjective-noun compounds is that across adjectives, we have a truly negative linear relation between $\log \mathcal{P}(A, N_A)$ and $\log V(A, N_A)$ — an analysis with a generalized additive model in which the linearity assumption is relaxed. In other words, Figure 3 reveals a relation that holds across all adjectives such that a given increase in 1 unit of type counts goes hand in hand with a fixed decrease of 0.59 units of category-conditioned productivity on a logarithm scale. This law-like behavior is likely to be due to the fact that all the compound types involve adjectival modification, with the adjectives restricted to adjectives of degree.

Although 大 (*da4*, 'big') may have small $\mathcal{P}(A, N_A)$ value for purely statistical reasons, it is conceivable that irregularity may play a part. If many of these compounds are phonologically irregular, the productivity of the adjective should suffer. On the other hand, if phonologically opaque forms are rare, as is the case for English *-ness*, where *business* is an outlier both with respect to form and meaning, then their influence on productivity would be negligible. In Mandarin, one kind of form irregularity that occurs in adjective-noun compounds is that the fixed tone of the noun can be lost, changing into what is called a floating tone. For instance, the noun 人 (*ren2*, 'person') appears in the compound 大人 with a floating tone (*da4ren*, 'officer'). In cases like this, there is a loss of suprasegmental transparency. In addition to the second syllable losing its original tone, the segmental make-up of the initial syllable can also change. For instance, in 大夫, *dai4 fu* ('doctor'), *da4* has changed into *dai4*. Table 2 lists the words in our data set for which such changes are attested. The segmental change is extremely rare, and is found for only 2 compounds (大夫 *dai4fu* ; 大王 *dai4wang*). Compounds with a floating tone for the noun are also small in number (12). The compounds showing these changes typically have high token frequency counts, the only exception being 小的 (*xiao3de0*, 'humble title for oneself'), with a frequency of 1. Given that the proportion of irregular forms is quite small (12/1482 = 0.008), it is unlikely that phonological opacity is a significant contributor to the negative correlation between $\log \mathcal{P}(A, N_A)$ and $\log V(A, N_A)$.

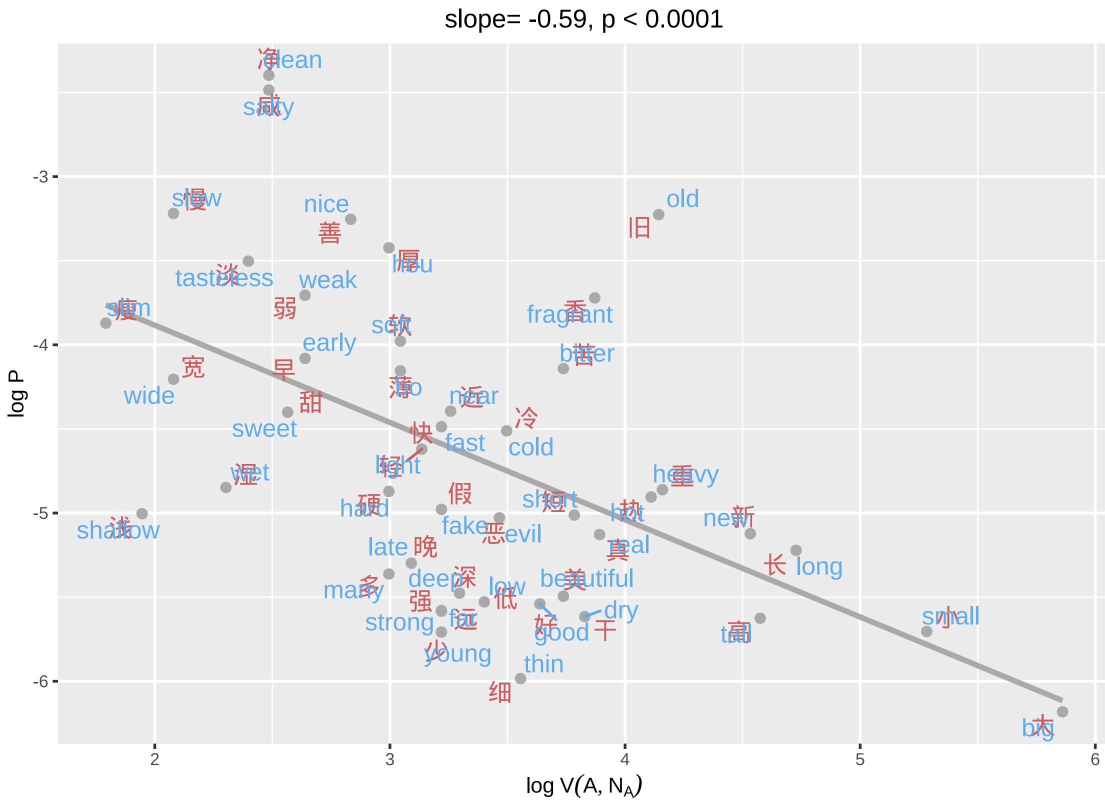

Figure 3: Scatterplot of adjectives in the plane spanned by log $V(A, N_A)$ and log $\mathcal{P}(A, N_A)$. The grey line represents the least squares linear regression line.

Table 2: Overview of phonological irregularities in Mandarin nominal adjective-noun compounds. Adjective and noun are denoted by C1 and C2, respectively.

Modifier	Word	Token frequency	C1pinyin	C1tone	C2pinyin	C2tone	C1change	C2floating
大 'big'	大人 'officer'	336	da4	4	ren2/ren0	2/0	no	yes
大 'big'	大夫 'title/doctor'	479	da4/dai4	4	fu1/fu0	1/0	yes	yes
大 'big'	大王 'king'	126	da4/dai4	4	wang2/wang	2/0	no/yes	yes
好 'good'	好儿 'benefit'	4	hao3	3	er2/er0	0	no	yes
冷 'cold'	冷颤 'chill'	5	leng3	3	zhan0	0	no	yes
少 'few'	少爷 'young master'	178	shao4	4	ye0	0	no	yes
瘦 'slim'	瘦子 'slim person'	8	shou4	4	zi0	0	no	yes
甜 'sweet'	甜头 'goodies'	19	tian2	2	tou0	0	no	yes
小 'small'	小子 'little boy'	249	xiao3	3	zi0	0	no	yes
小 'small'	小的 'humble title of oneself'	1	xiao3	3	de0	0	no	yes
小 'small'	小儿 'little son/smallness'	145	xiao3	3	er2/ er0	0	no	yes
真 'real'	真儿 'seriousness'	3	zhen1	1	er0	0	no	yes

Another factor that might contribute to this negative correlation is that adjectives with larger numbers of types would have few nouns left to which they could attach in a meaningful way. To address whether "onomasiological exhaustion" is at issue for Mandarin, we used Large Number of Rare Events (LNRE Baayen 2001) models to estimate how many word types are in the population but that have not been sampled in our corpus. In section 3, we will show that the estimated numbers of unseen types $V(0, A, N_A)$ have no systematic relation with type $V(A, N_A)$. Consequently, the negative correlation between the number of types and the category-conditioned degree of productivity can't be due to onomasiological exhaustion.

3 Estimating the numbers of unseen words with LNRE models

Recall that Aronoff (1976) suggested to quantify productivity in terms of the ratio of possible to actual words. We can decompose the number of possible words $S(A)$ into two parts, the number of words observed in the corpus $V(A, N_A)$, and the number of words that are possible in the population but that have not appeared in the corpus, henceforth $V(0, A, N_A)$.

In order to estimate the number of unseen adjective-noun compounds, we need the frequency spectrum of the observed words. The frequency spectrum tabulates, for each word frequency m, the number of words $V(m)$ there are with that frequency (see Table 3 in the supplementary materials for examples). In Figure 4, the black bars represent the observed spectrum for adjectives with 大 ('big'). Our goal is to estimate the number of unseen types $V(0)$ with this adjective. When going from right ($m = 15$) to left ($m = 1$), the height of the black bars increases, with greater increments as m becomes smaller. The number of unseen types $V(0)$ can be thought of as a bar positioned above $m = 0$, and its height is expected to be greater, the larger the increment between $V(1)$ and $V(2)$ is. To make this more precise, we made use of two statistical models for Large Number of Rare Event (LNRE) distributions (Baayen 2001) available in the **ZipfR** package (Evert and Baroni 2006) for R (R Core Team 2013). The first model, proposed by Sichel (1986), and known as the Generalized Inverse Gauss-Poisson (GIGP) model, provides the fit shown by the red bars in the left panel of Figure 4. The fitted values are a negatively decelerating function of m, and the estimated number of unseen types $\hat{V}(0, A, N_A)$ is 47. The second model, the finite Zipf-Mandelbrot model, estimates $\hat{V}(0, A, N_A)$ at 46. Table 1 lists the counts of unseen types for all adjectives where there were sufficient data to be able to fit the models. In what follows, we work with the estimates of the GIGP model, which tended to be somewhat more precise than those of the finite Zipf-Mandelbrot model, while mostly generating less extreme predictions for the counts of unseen types.

The number of unseen types $\hat{V}(0, A, N_A)$ turns out not to be correlated with the number of types $V(A, N_A)$ ($r = 0.038, p = 0.289$) but $\hat{V}(0, A, N_A)$ tends to increase with $V(1, A, N_A)$, as shown in Figure 5. The regression line in this figure was obtained with a Gaussian Location-Scale Generalized Additive Model (Wood 2017), which models mean and variance of $\hat{V}(0, A, N_A)$ as a function of $V(1, A, N_A)$. Both mean and varience of $\hat{V}(0, A, N_A)$ turned out to be linear in $V(1, A, N_A)$ (see Table 4 in supplementary materials). As can be seen in Figure 5, there is a scatter of adjectives where large numbers of unseen types are estimated. Due to the insufficiency of type counts for some adjectives, the estimate of $\hat{V}(0, A, N_A)$ is necessarily imprecise. Furthermore, for the present data, the frequency spectrum is often rather irregular, which can also give rise to overestimation

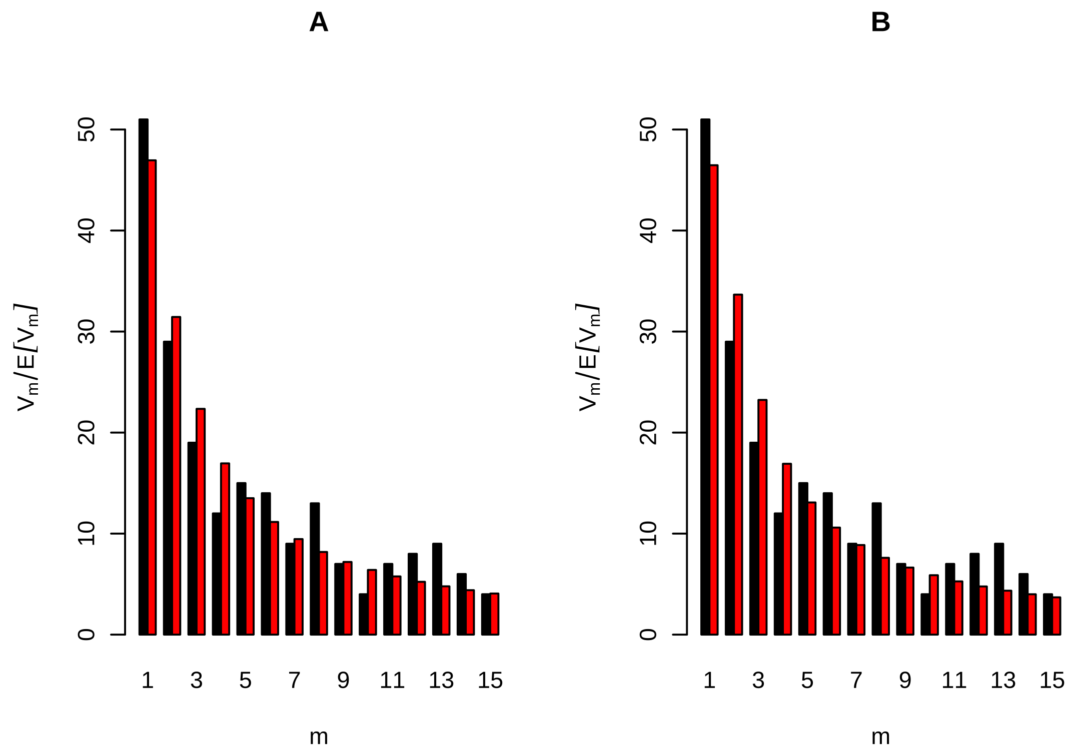

Figure 4: Frequency spectrum of adjective-noun compounds initiating with 大 (*da4*, 'big') based on the GIGP model (panel A) and the Zipf-Mandelbrot model (panel B). Black bars represent the observed spectrum elements $V(m)$ and red bars represent estimated counts E[$V(m)$].

of $\hat{V}(0, A, N_A)$. Therefore, the regression line predicted by the Gaussian Location-Scale Generalized Additive model is our most reliable summary of the relation between log $\hat{V}(0, A, N_A)$ and log $V(1, A, N_A)$. The absence of a correlation of $\hat{V}(0, A, N_A)$ with $V(A, N_A)$ and the presence of a correlation with $V(1, A, N_A)$, even though $V(1, A, N_A)$ and $V(A, N_A)$ are themselves correlated, points to the prime importance of $V(1, A, N_A)$ for estimating unseen types. Returning to Figure 4, the number of unseen types for $m = 0$ clearly depends more on the number of hapaxes than on the total number of types $V(A, N_A)$, which is the sum overall spectrum elements $V(m)$ including many high values of m that contribute little or no information to the shape of the spectrum for low m and $m = 0$.

Importantly, Figure 5 clarifies that even though 大 (*da4*, 'big') has the highest count of types, it has not exhausted the number of words it could give rise to. As a consequence, onomasiological exhaustion is clearly not a reason for the negative correlation between $V(A, N_A)$ and $\mathcal{P}(A, N_A)$ in Figure 3. This leaves us with the question whether the category-conditioned productivity $\mathcal{P}(A, N_A)$ of Mandarin adjective-noun compounds is merely a statistical measure that necessarily decreases as the number of types $V(A, N_A)$ increases. In section 4, we show that the semantic transparency of the adjective-noun constructions is correlated with $\mathcal{P}(A, N_A)$, such that more transparent constructions have greater category-conditioned degrees of productivity.

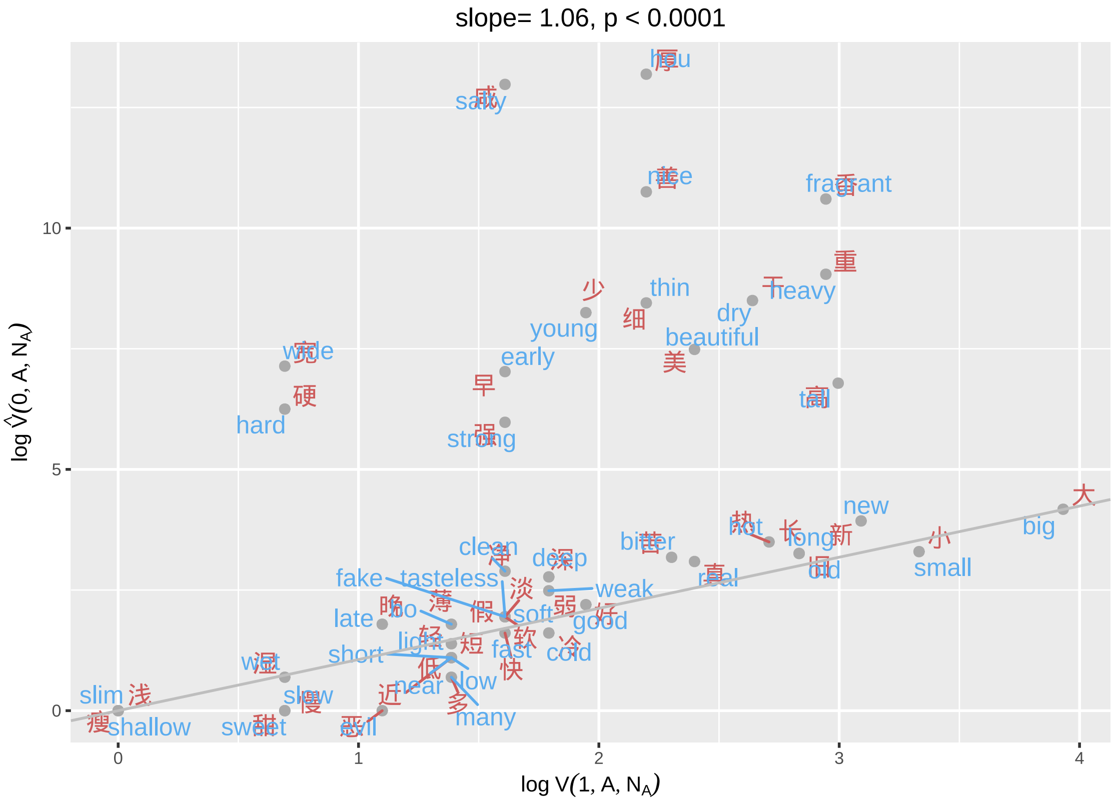

Figure 5: Scatterplot of the estimated count of unseen types $\hat{V}(0, A, N_A)$ and the observed count of hapaxes $V(1, A, N_A)$ for those adjectives with sufficient data for a GIGP model to be fitted. The regression line was obtained with a Gaussian Location-Scale Generalized Additive Model.

4 Productivity and semantic transparency

In order to clarify how the category-conditioned degree of productivity of Mandarin adjective-noun compounds varies with semantic transparency, we make use of distributional semantics. The basic idea of distributional semantics holds that words with similar distributions have similar meanings (Harris 1954; Firth 1957; Sahlgren 2001; McDonald and Shillcock 2001). Representing words by numeric vectors, the semantic similarity between words can be assessed either by the correlation between the vectors, or by the cosine of the angle between the vectors. The closer in meaning the two words are, the larger their cosine similarity is. Alternatively, the correlation of the two vectors can be observed. Methods of distributional semantics have proved useful in several previous studies addressing productivity and historical change (Goldberg 2016; Perek and Hilpert 2017; Perek 2018).

There are many ways in which words' semantic vectors can be calculated (see, e.g., Landauer and Dumais 1997; Shaoul and Westbury 2010; Mikolov et al. 2013). Given words' semantic vectors, known as "embedding" in distributional semantics, the semantic lexicon can be conceptualized as a high-dimensional space in which words are points (Boleda 2020). In the present study, we take semantic vectors of the target Mandarin adjectives, nouns, and adjective-noun compounds from Tencent AI Lab Embedding Corpus for Chinese Words and Phrases, which provides 200-dimensional vector representations for 8 million Chinese words and phrases (Song et al. 2018) obtained with a directional skip-gram model (`https://ai.tencent.com/ailab/nlp/en/embedding.html`). Semantic vectors (2262 unique in total) are available for all 56 adjectives, all 751 nouns, and for 1482 (out of 2055) adjective-noun compounds.

The accuracy of semantic vectors depends on whether words are properly semantically disambiguated. For instance, if the two meanings of English *bear* are not properly distinguished, the semantic vector for this word will be somewhere in between the true semantic vectors of the animal 'bear' and the verb 'to bear'. This raises the question of whether we can expect the semantic vectors for the adjectives and adjective-noun compounds in our study to be reasonably accurate.

To address this question, we consulted the Chinese Wordnet at `http://lope.linguistics.ntu.edu.tw/cwn/` and extracted the number of lemma for the adjectives and nouns occurring in our dataset. The mean number of lemma for both the adjectives and nouns is 1, but for both adjectives and nouns there are words with more lemma (up to 6 for adjectives, and up to 13 for nouns). Within our set of adjective-noun compounds, the adjectives tend to have one prototypical meaning only. Thus, of the 56 adjectives for which we have semantic vectors, only one adjective is written with a character (香, *xiang1*) that is truly polysemous, meaning either 'fragrant' or 'incense'. The lower part of Table 3 clarifies that within our set of adjective-noun compounds, the adjectival meaning of 香 is attested most often. The semantically related nominal meaning is not attested in the present study. The remaining 55 adjectives are written with a character that is unambiguous for 46 adjectives, but that is ambiguous in 9 cases. These cases are listed in the upper half of Table 3. For instance, the character 少 represents two words, *shao4* meaning 'young' and *shao3* meaning 'few'. In other words, *shao4* and *shao3* are homographs. For the nine homographic characters, one adjectival meaning is dominant in our dataset, with only one exception: 薄 (*bo2, bao2*). Its two meanings are rather more similar and the distribution of its semantic usages are almost balanced. Given the strong leftward skew of the distributions of the WordNet lemma of the adjectives and nouns, and the scarcity of ambiguity in the compounds, the overall accuracy of the semantic vectors may be expected to be reasonable.

Table 3: Characters that have homographs (upper part) and that are polysemous (lower part). The column labeled "Number of meanings" lists the counts of adjective-noun compounds in our database in which a given meaning is realized. POS: Part of speech.

Homographs								
Adjectival meaning in this study					Other meaning			
Character	Pronunciation	Meaning	Number of meanings	POS	Pronunciation	Meaning	Number of meanings	POS
少	shao4	young	16	adjective	shao3	few	2	adjective
干	gan1	dry	20	adjective	gan4	to do/major	0/5	verb/adjective
重	zhong4	heavy, serious	42	adjective	chong2	to repeat/repetitive	0/2	verb/adjective
长	chang2	long	73	adjective	zhang3	elder	9	adjective
好	hao3	good	34	adjective	hao4	be fond of/preference	0/0	verb/ noun
薄	bo2	slight, fickle, infertile	6	adjective	bao2	thin	7	adjective
大	da4	big	285	adjective	dai4	lexicalized	2	adjective
恶	e4	evil	22	adjective	wu4/e2	to hate	0	verb
假	jia3	fake	22	adjective	jia4	holiday	0	noun
Polysemous character								
Adjectival meaning in this study					Other meaning			
Character	Pronunciation	Meaning	Number of meanings	POS	Pronunciation	Meaning	Number of meanings	POS
香	xiang1	fragrant	27	adjective	xiang1	incense	0	noun

In order to explore productivity via means of semantic vectors, for each compound word, we calculated the correlation between the semantic vectors of the adjective and the compound ($r_{\text{A-AN}}$), the correlation between the noun and the compound ($r_{\text{N-AN}}$), and the correlation between the adjective and the noun ($r_{\text{A-N}}$). For each adjective, we also calculated the average of all pairwise correlations of compounds ($\bar{r}_{\text{AN}}$). This last measure captures the semantic coherence (Aronoff 1976) of an adjective-noun morphological category.

In addition to these quantitative measures, we also evaluated manually each compound on three dimensions. First, we checked whether the compound is listed in a particular online dictionary, *Baidu Hanyu* (`https://hanyu.baidu.com/`). Second, the first author inspected each compound with respect to whether the compound is a label, i.e., a conventionalized name for an object, event, or idea, or whether the adjective can simply modify the noun as it would in an adjective-noun phrase. This resulted in a factor `Function` with as levels `labeling` and `modification`. We assigned a compound to the `modification` category when a modification reading is possible. However, often, these compounds also have a conventionalized metaphorical reading. For instance, 大器 (*da4qi4*, 'big instrument') can mean 'outstanding talent'. Third, the first author rated each compound for whether it was semantically transparent, in the sense that the meaning of the compound had a clear relation to the meanings of its constituents, resulting in a factor `Transparency` with levels as `transparent` and `opaque`. Table 4 provides examples for each of the combinations of `Listedness`, `Function`, and `Transparency`. The assignment of values for `Function` and `Transparency` are unavoidably impressionistic. The purpose of these classifications is therefore a very simple one, namely to be able to probe, however imperfectly, whether the semantic vectors make sense.

Unsurprisingly, there are very few cases in Table 4 of compounds that we judged to be opaque, that are not listed in the *Baidu* dictionary, while allowing a modification reading rather than labeling. One such example is 重者 (*zhong4zhe3*) which can mean 'heavy person', but also has the

Table 4: Examples of compounds cross-tabulated by `Listedness`, `Function`, and `Transparency`. For the translations, we consulted not only the *Baidu Hanyu* but also the yabla dictionary at `https://chinese.yabla.com/chinese-english-pinyin-dictionary.php`.

`Listedness`	`Function`	`Transparency`	Count	Example	Translation
listed	labeling	transparent	409	淡水	dan4shui3, light water, ‘sweet water’
not-listed	labeling	transparent	34	干姜	gan1jiang1, dry ginger, ‘Rhizoma Zingiberis’
listed	labeling	opaque	331	长工	chang2gong1, long work, ‘bitter labor’
not-listed	labeling	opaque	10	香兰	xiang1lan2, fragrant orchid, ‘cymbidium’
listed	modification	tranparent	441	大奖	da4jiang3, big prize, ‘grand prize’
not-listed	modification	transparent	166	薄纸	bo2zhi3, thin paper, ‘thin paper’
listed	modification	opaque	47	大器	da4qi4, big instrument, ‘outstanding talent’
not-listed	modification	opaque	2	重者	zhong4zhe3, heavy person, ‘severe situation’

metaphorical meaning ‘severe situation’. The category with the largest number of compounds contains transparent compounds that are listed in the dictionary and that can have a modification reading along with a more specialized meaning, such as 大奖 (*da4jiang3*, big prize, ‘grand prize’). Here, we are dealing with conventionalized collocations that block (Aronoff 1976) alternative possible words such as * 大标 (*da4biao1*, ‘big prize’).

In order to obtain an exploratory validation of the semantic similarity measures, we conducted a series of t-tests contrasting `listed` and `not-listed` compounds, compounds with the `labeling` versus `modification` function, and `transparent` and `opaque` compounds, for each of the three correlations $r_{\text{A-AN}}$, $r_{\text{N-AN}}$, and $r_{\text{A-N}}$. Results are listed in Table 5. For all three measures, we find that the mean correlation is higher for the `modification` function than for the `labeling` function. Likewise, higher means characterize `transparent` as compared to `opaque` compounds. It is only for $r_{\text{N-AN}}$ that we find a difference in listedness, in which compounds that are listed in the dictionary have smaller correlations.

Table 5: T-tests comparing differences in semantic vector correlations for `Function`, `Listedness`, and `Transparency`.

correlation	`Function`	`Listedness`	`Transparency`
$r_{\text{A-AN}}$	modification $>$ labeling	listed $\approx$ not-listed	transparent $>$ opaque
	Est: 0.02; $p = 0.0007$	Est: = 0.006; $p = 0.437$	Est: 0.043; $p < 0.0001$
$r_{\text{N-AN}}$	modification $>$ labeling	listed $<$ not-listed	transparent $>$ opaque
	Est: 0.076; $p < 0.0001$	Est=-0.076; $p < 0.0001$	Est= 0.087; $p < 0.0001$
$r_{\text{A-N}}$	modification $>$ labeling	listed $\approx$ not-listed	transparent $>$ opaque
	Est: 0.154; $p = 0.0004$	Est: -0.006; $p = 0.331$	Est: 0.020; $p < 0.0001$

These results suggest that all three correlation measures capture aspects of semantic transparency. This raises the question of whether all three measures are equally predictive for the category-conditioned productivity of an adjective-noun compound. Since productivity is defined over morphological categories, i.e., sets of words sharing both form and meaning, one may expect that the correlation measure $r_{\text{A-AN}}$ that compares the vector of the adjective with the vectors of the compounds is the best predictor of category-conditioned productivity. To evaluate this prediction, as a first step, we fitted a mixed model to $r_{\text{A-AN}}$, with fixed-effect predictors `Function`, `Listedness`, and `Transparency`, and with `Adjective` as random-effect factor. As `Listedness` turned out not to be predictive, we removed it from the model specification. The resulting model is summarized in Table 6. The effects of `Function` and `Transparency` are as expected.

Of primary interest is the random effect for `Adjective`, which clarifies that the different adjectives differ in their transparency as quantified by $r_{\text{A-AN}}$. We therefore extracted the by-adjective adjustments a_A to the intercept and compared these to the category-conditioned productivity values $\mathcal{P}(A, N_A)$ of the adjectives. A correlation test revealed a well-supported positive correlation, $r[a_A, \mathcal{P}(A, N_A)] = 0.514, t(43) = 3.04, p = 0.0041$. In other words, the more transparent the meaning of the adjective is in the meaning of the compound, the greater its category-conditioned productivity is.

Table 6: Summary of a generalized additive mixed model fitted to $r_{\text{A-AN}}$, with fixed-effect predictors `Function` and `Transparency`, and with random intercepts for `Adjective`. The model was fitted with the `gam` function from the **mgcv** package for R.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept (Fuction= modification, Transparency= opaque)	0.3921	0.0141	27.843	< 0.0001
Function=label	-0.0179	0.0054	3.3367	0.0008
Transparency=transparent	0.0294	0.0058	5.022	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
random intercepts Adjective	49.02	53	17	< 0.0001

We carried out the same analyses for $r_{\text{N-AN}}$ and for $r_{\text{A-N}}$. The by-adjective adjustments b_A to the intercept did not correlate significantly with $\mathcal{P}(A, N_A)$ ($p = 0.08$). For $r_{\text{A-N}}$, a minor correlation was present with the by-adjective adjustments c_A, ($r[c_A, \mathcal{P}(A, N_A)] = 8.25, p = 0.05$). This suggests that possibly, the semantic similarity of the adjective to the noun also shapes its productivity. However, since the random intercepts a_A and c_A enter into a strong correlation ($r = 0.71, p < 0.0001$), we evaluated their relative importance by including them as predictors in a linear model predicting $\mathcal{P}(A, N_A)$. In this model, a_A retained significance, whereas c_A was no longer significant ($p > 0.9$). These results lead to the conclusion that the crucial semantic determinant of the productivity of an adjective-noun construction is the semantic relatedness of the adjective and the compound. Given that the morphological category for a given adjective consists of compounds, and that the form and meaning that are shared by the compounds in their morphological category are grounded in the adjective, this result makes perfect sense.

The correlation between by-adjective random intercept adjustment a_A and category-conditioned productivity $\mathcal{P}(A, N_A)$ is visualized in Figure 6 (see Table 5 in supplementary materials for complete data). 甜 (*tian2*, 'sweet') has the largest by-adjective random intercept adjustment and also the largest category-conditioned productivity. By contrast, 少 (*shao3*, 'few') is at the other extreme. In other words, since 甜 (*tian2*, 'sweet') is more semantically transparent than 少 (*shao3*, 'few'), 甜 (*tian2*, 'sweet') can be more productive in compounding.

In addition to the by-adjective random intercept adjustment a_A, the average of all pairwise correlations between pairs of compounds ($\bar{r}_{AN}$) also turned out to be a good predictor of the category-conditioned degree of productivity ($r[\bar{r}_{AN}, \mathcal{P}(A, N_A)] = 6.33, t(43) = 3.36, p = 0.0016$). That is to say, adjectives with more semantically similar compounds are more productive. However, $\bar{r}_{AN}$ and a_A are strongly correlated ($r[\bar{r}_{AN}, a_A] = 0.86, t(43) = 10.66, p < 0.001$). This raises the question of which measure is the superior predictor for $\mathcal{P}(A, N_A)$. Unfortunately, their correlation is so strong that both predictors are not significant (both $p > 0.1$, see Table 6 in supplementary materials) when both entered into a generalized additive mixed model. Whereas simultaneously a significant proportion of the variance ($F(2, 42) = 5.641, p = 0.0067$) is attested — the hallmark of collinearity (see, e.g., Chatterjee and Hadi 2012). We conclude that both predictors provide a useful window on the category-conditioned degree of productivity, namely, both the semantic similarity of the compounds within the morphological category, and as well the semantic relatedness of the adjective and the compounds, are important for an adjective-noun construction to be productive.[2] [3]

In order to provide an overview of the relative importance of the different factors predicting category-conditioned productivity, we conducted a random forest analysis (Breiman 2001; Torsten Hothorn and Zeileis 2006). As predictors we included token frequency N, type counts $V(A, N_A)$, the count of hapaxes $V(1, A, N_A)$, by-adjective random intercept adjustment a_A, the mean of pairwise correlations between compound's semantic vectors $\bar{r}_{\text{AN}}$, and one additional variable that we were interested in, polarity.

Recall that the adjectives included in the present study are all polar adjectives with either positive or negative meanings (Kennedy 1998). According to the Pollyanna principle, humans tend to use positive words more frequently than their negative counterparts in evaluations and judgements (Boucher and Osgood 1969; Matlin 2016). Presumably, positive adjectives in polar adjectives are more productive than negative ones in Mandarin adjective-noun compounds.

Figure 7 presents the variable importance of the six predictors according to a random forest analysis. The predictor with the highest variable importance is the number of tokens N. This is unsurprising as the number of tokens figures in the denominator of $\mathcal{P}(A, N_A)$. Unavoidable, higher values of $\mathcal{P}(A, N_A)$ go hand in hand with lower values of N ($r = -0.53, p < 0.0001$). The next most important predictors are the by-adjective random adjustments a_A and the number of types $V(A, N_A)$, which are closely followed by the mean pairwise correlations $\bar{r}_{\text{AN}}$. Consistent with the analysis using the linear model, the by-adjective random intercept adjustment a_A and the average of the pairwise correlations of compound semantic vectors $\bar{r}_{\text{AN}}$ are basically equally important for predicting $\mathcal{P}(A, N_A)$. These two predictors capture two related facets of semantic transparency: $\bar{r}_{\text{AN}}$ highlights the importance of semantic similarity within the set of compounds, whereas a_A

[2]When the log-transformed mean relative frequency (Hay 2001) is included as an additional covariate for predicting $\mathcal{P}(A, N_A)$, it is significant with a negative slope, as expected, since a morphological category with more hapaxes must have lower frequency ratios. Importantly, both a_A and $\bar{r}_{AN}$ remain predictive with positive slopes. For further details, see the supplementary materials.

[3]In order to validate the correlation between semantic transparency and category-conditioned degree of productivity, we conducted a follow-up study with 34 different kinds of word embeddings, obtained from 9 different corpora and 4 different context features. Across all combinations of corpus and method, we observed the same results as reported in the present study, with effects being the strongest for word-character-based vectors combined with corpora of either `Sougou News` or `Zhihu Question Platform`. See Table 8 in supplementary materials.

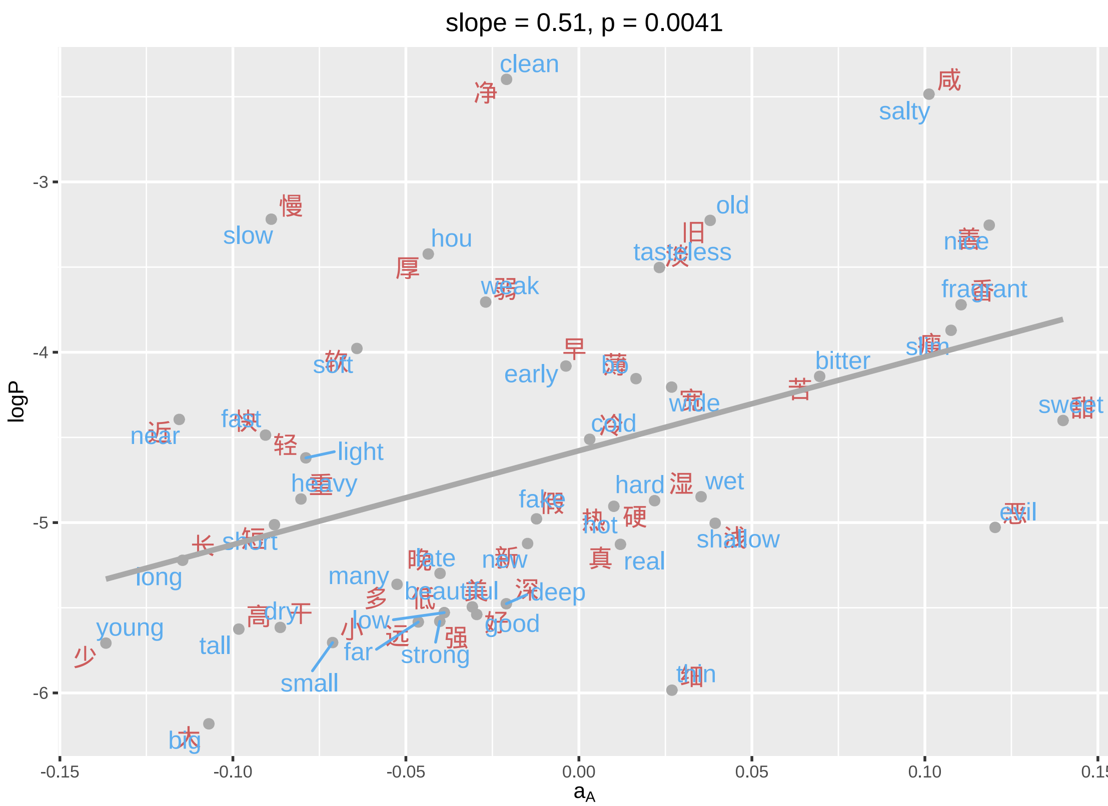

Figure 6: Scatterplot of category-conditioned productivity $\log \mathcal{P}(A, N_A)$ as a function of by-adjective random intercept adjustments a_A.

highlights the role of the semantic transparency of the adjective in relation to its compounds. Of course, it makes sense for adjectives with higher values of a_A to also have higher values of $\bar{r}_{\mathrm{AN}}$, and vice versa: the more the adjective contributes to compounds' meanings, the more similar these compound meanings will be to each other.

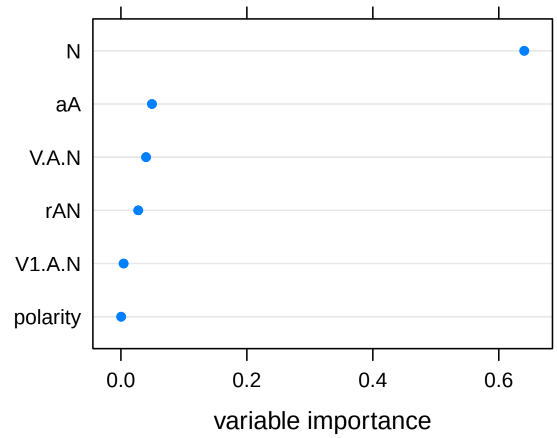

Figure 7: Variable importance according to a random forest analysis of predictors of category-conditioned productivity $\mathcal{P}(A, N_A)$. N: token frequency (N), V.A.N: type counts $V(A, N_A)$, V1.A.N: hapaxes $V(1, A, N_A)$, aA: by-adjective random adjustment (a_A), rAN: mean of correlations of semantic vectors for compounds itself ($\bar{r}_{\mathrm{AN}}$), and polarity.

The number of hapaxes and polarity are hardly predictive for $\mathcal{P}(A, N_A)$. The irrelevance of polarity as predictor of $\mathcal{P}(A, N_A)$ is further confirmed by a linear model ($t(44) = -0.038, p = 0.892$, see Table 7 in supplementary materials). However, there is some indication that positive adjectives have more hapaxes than negative adjectives ($t(44) = 0.648, p = 0.011$). This suggests that positive adjectives may contribute more to the overall growth rate of the vocabulary (the hapax-conditioned degree of productivity), in accordance with the Pollyanna principle, but that polarity is irrelevant when it comes to how semantic transparency within morphological categories. Once the focus is on the morphological category, i.e., on the set of compounds sharing a given adjective, then the polarity of that adjective is fixed. In this case, $\mathcal{P}(A, N_A)$ reflects the extent to which the category is semantically regular and transparent, but it remains blind to the general usefulness of the category itself for communication — the level at which the Pollyanna principle might come into play.

A further complication is that for some polar adjectives, it is not straightforward to evaluate the kind of polarity involved. For instance, it is unclear whether 咸 (*xian2*, 'salty') and 淡 (*dan4*, 'tasteless') are opposite on the same scale. Furthermore, for pairs such as 干 (*gan1, 'dry'*) vs. 湿 (*shi1*, 'wet'), it is difficult to decide which of the two is on the positive end of the scale, and which on the negative end. In summary, no firm conclusions can be drawn about whether the productivity of Mandarin adjective-noun compounds is co-determined by the Pollyanna principle.

The results obtained thus far are based on semantic vectors (word embeddings). To consolidate these results, we carried out two further complementary explorations.

First, we inspected the number of lemmas and the number of senses given for the adjectives in the Chinese Wordnet (Huang et al. 2010), and regressed the log-transformed counts of types V and tokens N on the WordNet counts of lemmas and senses, respectively. Significant positive correlations emerged for V and the count of senses ($\hat{\beta} = 0.94, t(43) = 3.86, p = 0.0001$) and the count of lemmas ($\hat{\beta} = 0.63, t(43) = 2.78, p = 0.008$), as well as for N and the count of senses ($\hat{\beta} = 1.73, t(43) = 4.6, p < 0.0001$) and the count of lemmas ($\hat{\beta} = 1.09, t(43) = 2.95, p = 0.005$). Unsurprisingly, adjectives with more meanings and senses are used more extensively in adjective-noun compounds: their greater semantic range offers more opportunities for word formation, and supports a greater extent of use. Importantly, a strong negative correlation ($\hat{\beta} = -1.07, t(43) = -4.25, p < 0.0001$) is present for the count of senses and log-transformed $\mathcal{P}(A, N_A)$. The more senses an adjective has, the less predictable the meaning of a novel adjective becomes. This decrease in semantic systematicity within the adjective's morphological category is detrimental for its capacity of giving rise to novel, unseen adjective-noun compounds.

Second, we extracted constituents' semantic transparency ratings from the Chinese Lexical Project (Tse et al. 2017). These ratings only cover 42% of the compounds that we examined in the present study. Nevertheless, we still find solid correlations between subjective semantic transparency ratings for the first constituent and the compound (sub$[A, AN]$) and the second constituent and the compound (sub$[N, AN]$) on the one hand, and the corresponding measures (emb$[A, AN]$, emb$[N, AN]$) based on word embeddings: r_s(sub$[A, AN]$, emb$[A, AN]$) $= 0.46$, $p = 0.001$, r_s(sub$[N, AN]$, emb$[N, AN]$) $= 0.45, p = 0.002$ on the other hand. This supports the reliability of semantic vectors for capturing semantic transparency. Importantly, when log $\mathcal{P}(A, N_A)$ is regressed on sub$[A, AN]$, a regression line with positive slope is obtained ($\hat{\beta} = 1.88, t(43) = 4.05, p = 0.0002$). Conversely, when log $\mathcal{P}(A, N_A)$ is regressed on sub$[N, AN]$, there is no solid evidence for a functional relation between the two measures ($p > 0.5$). Clearly, it is the semantic transparency between adjectives and compounds that shapes category-conditioned productivity.

5 General discussion

This study addressed the question of how Mandarin adjective-noun compounding typically subserving the creation of names for things and events in the world can be productive. As things and events in the world are in many ways "sui generis", with their own very specific properties, a word formation process that creates names for these very different things and events runs the risk of not having its own clear semantics. Whereas the English de-adjectival prefix *un-* (as in *unkind, unreliable*) simply specifies negation, compounds such as 大家 (*da4jia1*, big family, 'everyone'), 大写 (*da4xie3*, big write 'capital letter') and 大亨 (*da4heng1*, big prosperous, 'magnate'), although sharing the adjective 大 (*da4*, 'big'), have meanings that are by far not as compositional compared to *un-*.

We first clarified that adjective-noun compounds sharing the same adjective construct morphological categories, sets of words sharing form and meaning. For instance, the above compounds with 大 do not only share form (the adjective) but also the sense of being at the extreme end of some scale (e.g., group size, seriousness, or degrees).

Next, we clarified that there are substantial adjective-specific differences in the productivity of Mandarin adjective-noun compounds. We considered four measures of productivity for a given adjective A: the type count $V(A, N_A)$ (a measure of extent of use, also referred to as realized productivity), the count of hapaxes $V(1, A, N_A)$ (the hapax-conditioned degree of productivity), the category-conditioned degree of productivity $\mathcal{P}(A, N_A)$, and the number of unseen types $V(0, A, N_A)$. We verified that there are statistically substantial differences in the extent of use of the different adjective-noun compounds.

The extent of use $V(A, N_A)$ is positively correlated with the number of hapaxes $V(1, A, N_A)$. This is unsurprising, as the hapaxes tend to comprise roughly half of the types.[4] Extent of use is not correlated with the number of unseen types $V(0, A, N_A)$, which we estimated using the GIGP model. However, the hapax-conditioned degree of productivity, gauged with the count of hapaxes, is positively correlated with the number of unseen types. This finding clarifies that for all but the least productive adjectives, the number of attested types does not exhaust the number of possible types, and that hence we can rule out that the existing adjectives exhaust available onomasiologically sensible word formation possibilities. In other words, among the different adjective-noun compounds, we find truly productive morphological categories.

The extent of use $V(A, N_A)$ enters into a negative correlation with the category-conditioned degree of productivity $\mathcal{P}(A, N_A)$. By itself, this negative correlation is perhaps unsurprising. The type count $V(A, N_A)$ is a monotonically increasing function of token frequency N_A. Its first derivative, the growth rate $\mathcal{P}(A, N_A)$ necessarily decreases as the type count $V(A, N_A)$ and token frequency N_A increase. What is surprising, however, is that in earlier quantitative studies of English derivation (Baayen and Lieber 1991; Hay and Baayen 2002), no such negative correlation between $V(1, A, N_A)$ and $\mathcal{P}(A, N_A)$ was present. For the data presented in Baayen and Lieber (1991), $r = 0.09, p = 0.6819$. For the table of prefixes given in Hay and Baayen (2002), $r = 0.29, p = 0.1536$, and for their table of suffixes, $r = 0.1, p = 0.4702$. This raises the question of why Mandarin adjective-noun compounds and English derivational affixes show such diverging relations between $V(A, N_A)$ and $\mathcal{P}(A, N_A)$.

Possibly, the high diversity of the semantic functions of English derivational affixes is at issue. The suffixes *-less* and *-ation* realize very different semantics. To control such large differences in meaning, researchers have therefore focused on so-called rival affixes, affixes that are semantically fairly similar, such as *-ness* and *-ity*, and *un-* and *in-* (Aronoff 1982; Baayen et al. 2013). However, even for pairs that at first glance seem to realize the same semantics, many subtle differences in meaning and use exist (Riddle 1985). By contrast, for the present Mandarin compounds, we always have adjectival modification with polar adjectives. Thus, the adjective-noun compounds of Mandarin make it possible to study differences in productivity while controlling for semantics for much larger datasets than is possible for analyses based on rival affixes.[5]

[4] A strong positive correlation between $V(A, N_A)$ and $V(1, A, N_A)$ is also attested for English affixes. For the data presented in Baayen and Lieber (1991), Table 3 and Table 4 combined, $r = 0.94, p < 0.001$.

[5] To verify that indeed the adjective-noun compounds are more similar to each other than English derived words, we calculated all pairwise correlations of the compounds in our dataset, as well as all pairwise correlation of 898 English derived words (realizing 24 different affixes) that were studied in Baayen et al. (2019). Mean semantic similarity was much lower for English (0.016) than for Mandarin (0.239, $p < 0.0001$, Wilcoxon test).

Thanks to the high degree of semantic control in our study, the category-conditioned productivity emerges as correlated with semantic transparency. The more transparent the meaning of the adjective is with respect to the meaning of the compound, the greater its category-conditioned productivity is. Although correlation is not causation, this finding is consistent with the hypothesis that semantic transparency makes productivity possible. The positive correlation between log $\mathcal{P}(A, N_A)$ and the mean of all pairwise correlations of semantic vectors of compounds $\bar{r}_{\mathrm{AN}}$, a formalization of Aronoff (1976)'s concept of semantic coherence, likewise points to the importance of semantic transparency.

Considering all measures for productivity jointly, two groups of measures emerge. The first group comprises $V(A, N_A), V(1, A, N_A)$, and $\hat{V}(0, A, N_A)$. These measures all reflect profitability, the extent to which a morphological category is used or can be useful in the future. The second group has one member only, the category-conditioned degree of productivity $\mathcal{P}(A, N_A)$. This measure characterizes the internal systematicity of a morphological category. Baayen (2009) observed that an affix can have a high category-conditioned degree of productivity and yet have a low profitability. The present study adds to this observation that the internal semantic systematicity of morphological categories is correlated with $\mathcal{P}(A, N_A)$, and not with the measures of profitability. In other words, profitability measures provide insight into the onomasiological usefulness of a word formation process and the extent to which it is "fashionable" in the language community, whereas the category-conditioned productivity measure gauges the internal semantic systematicity of the morphological category, which likely is a qualitative, structural prerequisite for it to be productive. In other words, the agreement between semantic transparency and morphological productivity proposed by Baayen (1993) gains robust support. Since this is the very first study on quantitative investigation of the productivity of Mandarin adjective-noun compounds, it is worth exploring whether the findings could generalize to other kinds of compounds in Mandarin, and also to compounds and compound-like constructions in other languages in the further work.

Author note

Corresponding author: Shen Tian. Email: sweetilovefreedom@126.com. This research was funded by the China Scholarship Council (Grant No. 201906230064). The authors are indebted to Chuang Yu Ying and Karlina Denistia for their comments of an earlier version of this paper. We also thank our reviewers for their constructive feedback which helped us strengthen the paper.

References

Aronoff, M. 1976. *Word Formation in Generative Grammar*. MIT Press, Cambridge, Mass.

Aronoff, M. 1982. Potential words, actual words, productivity and frequency. In *Preprints of the Plenary Session Papers of the XIIIth International Congress of Linguists*, 141–148, Tokyo.

Aronoff, M. and Fudeman, K. 2011. *What is morphology?* John Wiley & Sons.

Baayen, R. H. and Lieber, R. 1991. Productivity and English derivation: a corpus-based study. *Linguistics* 29. 801–843.

Baayen, H. 1993. On frequency, transparency and productivity. In *Yearbook of morphology 1992*, 181–208. Springer.

Baayen, R. H. 1994. Productivity in language production. *Language and Cognitive Processes* 9. 447–469.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. 1996. The CELEX lexical database (cd-rom). University of Pennsylvania.

Baayen, R. H. 2001. *Word Frequency Distributions*. Kluwer Academic Publishers, Dordrecht.

Baayen, R. H. 2009. Corpus linguistics in morphology: morphological productivity. In Kytö, M. and Lüdeling, A. (eds.), *Corpus Linguistics. An international handbook*. 900–919. Mouton de Gruyter, Berlin.

Baayen, R. H., Janda, L. A., Nesset, T., Endresen, A., and Makarova, A. 2013. Making choices in Russian: Pros and cons of statistical methods for rival forms. *Russian Linguistics* 37. 253–291.

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. P. 2019. The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. [Special issue]. *Complexity* 2019. 1–39.

Bauer, L. 2001. *Morphological productivity*. Cambridge University Press, Cambridge.

Boleda, G. 2020. Distributional semantics and linguistic theory. *Annual Review of Linguistics* 6. 213–234.

Bolinger, D. L. 1948. On defining the morpheme. *Word* 4(1). 18–23.

Booij, G. E. 1977. *Dutch Morphology. A Study of Word Formation in Generative Grammar*. Foris, Dordrecht.

Booij, G.E. 2010. Construction morphology. *Language and linguistics compass* 4(7). 543–555.

Boucher, J. and Osgood, C. E. 1969. The pollyanna hypothesis. *Journal of verbal learning and verbal behavior* 8(1). 1–8.

Breiman, L. 2001. Random forests. *Machine Learning* 45. 5–32.

Ceccagno, A. and Basciano, B. 2007. Compound headedness in Chinese: An analysis of neologisms. *Morphology* 17(2). 207–231.

Chatterjee, S. and Hadi, A. 2012. *Regression analysis by example*. John Wiley & Sons, New York.

Corbin, D. 1987. *Morphologie derivationelle et structuration du lexique* [Derivational morphology and lexical structure]. Niemeyer, Tübingen.

Evert, S. and Baroni, M. 2006. The zipfR library: Words and other rare events in R. useR! 2006: The second R user conference, Vienna, June 2006.

Firth, J. R. 1957. *Studies in linguistic analysis*. Wiley-Blackwell.

Goldberg, A. E. 2016. Partial productivity of linguistic constructions: Dynamic categorization and statistical preemption. *Language and cognition* 8(3). 369–390.

Harris, Z. S. 1954. Distributional structure. *Word* 10(2-3). 146–162.

Hay, J. B. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39. 1041–1070.

Hay, J. B. and Baayen, R. H. 2002. Parsing and productivity. In Booij, G. and Van Marle, J. (eds.), *Yearbook of Morphology 2001*, 203–235. Kluwer Academic Publishers, Dordrecht.

Huang C. R., Hsieh, S. K., Hong, J. F., Chen, Y. Z., Su Y. L., Chen Y. X., and Huang, S. W. 2010. 中文词汇网络: 跨语言知识处理基础架构的设计理念与实践 [Chinese Wordnet: Design, Implementation, and Application of an Infrastructure for Cross-lingual Knowledge Processing]. 中国语文 [*Study of Chinese language*] 24.

Huang C. R., Hsieh, S. K., and Chen, K. J. 2017. *Mandarin Chinese words and parts of speech: A corpus-based study.* Taylor & Francis.

Kastovsky, D. 1986. Productivity in word formation. *Linguistics* 24. 585–600.

Kennedy, C. 1998. On the monotonicity of polar adjectives. John Benjamins Publishing, Amsterdam.

Kruisinga, E. 1932. *A Handbook of Present-Day English, Part II: English accidence and syntax*. Noordhoff, Utrecht.

Landauer, T. and Dumais, S. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review* 104(2). 211–240.

Lieber, R. 2010. *Introducing Morphology*. Cambridge University Press, Cambridge, UK.

Matlin, M. W. 2016. Pollyanna principle. In Pohl, R. (ed.), *Cognitive Illusions: Intriguing Phenomena in Thinking, Judgment and Memory* 315–335. Routledge, London.

McDonald, S. and Shillcock, R. 2001. Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech* 44. 295–323.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. *arXiv preprint arXiv:1310.4546*.

Perek, F. 2018. Recent change in the productivity and schematicity of the way-construction: A distributional semantic analysis. *Corpus Linguistics and Linguistic Theory* 14(1). 65–97.

Perek, F. and Hilpert, M. 2017. A distributional semantic approach to the periodization of change in the productivity of constructions. *International Journal of Corpus Linguistics* 22(4). 490–520.

Plag, I. 2003. *Word Formation in English*. Cambridge University Press, Cambridge, UK.

Plag, I. 2010. Compound stress assignment by analogy: the consituent family bias. *The Mind Research Repository (beta)* (1).

R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Riddle, E. 1985. A historical perspective on the productivity of the suffixes *-ness* and *-ity*. In Fisiak, J. (ed.), *Historical Semantics, Historical Word-Formation*, 435–461. Mouton, New York.

Sahlgren, M. 2001. Vector-based semantic analysis: Representing word meanings based on random labels. Paper presented at the Workshop on Semantic Knowledge Acquisition and Subcategorization at the XIII European Summer School in Logic, Language and Information, Helsinki, August 2001.

Schultink, H. 1961. Produktiviteit als morfologisch fenomeen [Productivity as a morphological phenomenon]. *Forum der Letteren* 2. 110–125.

Shaoul, C. and Westbury, C. 2010. Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods* 42(2). 393–413.

Sichel, H. S. 1986. Word frequency distributions and type-token characteristics. *Mathematical Scientist* 11. 45–72.

Song, Y., Shi, S., Li, J., and Zhang, H. 2018. Directional skip-gram: Explicitly distinguishing left and right context for word embeddings. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 175–180.

Torsten Hothorn, K. H. and Zeileis, A. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15. 651–674.

Tse, C.-S., Yap, M. J., Chan, Y.-L., Sze, W. P., Shaoul, C., and Lin, D. 2017. The Chinese lexicon project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words. *Behavior Research Methods* 49(4). 1503–1519.

Wood, S. N. 2017. *Generalized Additive Models: an introduction with R*. CRC press.

Xu, Z. 2018. The word status of Chinese adjective-noun combinations. *Linguistics* 56(1). 207–256.