

**能产性理论和汉语复合构词初探**  
**A STUDY ON THEORIES OF**  
**MORPHOLOGICAL PRODUCTIVITY AND**  
**CHINESE COMPOUND WORDS**

李萌

LI MENG

新加坡国立大学中文系  
NATIONAL UNIVERSITY OF SINGAPORE  
北京大学中国语言文学系  
PEKING UNIVERSITY

2015

能产性理论和汉语复合构词初探  
A STUDY ON THEORIES OF  
MORPHOLOGICAL PRODUCTIVITY AND  
CHINESE COMPOUND WORDS

李萌

LI MENG

( *B.A., Southeast University* )

新加坡国立大学中文系  
北京大学中国语言文学系  
双学位硕士（研究）学位论文

A THESIS SUBMITTED  
FOR THE DOUBLE DEGREE M.A. BY RESEARCH  
IN CHINESE LANGUAGE  
DEPARTMENT OF CHINESE STUDIES, NUS  
DEPARTMENT OF CHINESE LANGUAGE AND  
LITERATURE, PKU

2015

## Acknowledgement

My interest in morphological productivity originated from Professor Chen Baoya (陈保亚) 's course *Theoretical Linguistics*, then I gave a talk about it later in his another course, *Linguistic Methodology*, after reading Bauer and other scholars' works. Furthermore, Professor Wang Hongjun (王洪君), a very knowledgeable and amiable teacher, taught me a lot of things about lexical frequency and word-formations during the break in her courses. All of these enlightened me on this subject, and an idea about productivity of Chinese compound words hit me one day.

The idea can successfully develop into this thesis, which owes a great deal to my supervisor, A/P Ye Wenxi (叶文曦), who poured his time and energy into my thesis writing, and offered remarkable comments. Additionally, I benefit a lot from my supervisor at NUS, A/P PHUA Chiew Pheng (潘秋平), whose modesty and kind-hearted help are unforgettable.

I would like to express deep gratitude to Dr. Richard Xiao (肖忠华, Department of Linguistics and English Language at Lancaster University), who generously provides the wordlist used in this thesis, and I cannot complete this thesis without it.

My colleagues and friends, Jin Yanwei (金延伟, Department of Linguistics at SUNY, Buffalo), Lei Tangxun (雷塘洵, Department of Chinese Language and Literature at PKU), Yang Xi (杨希, Department of Chinese Studies at NUS), Ye Shumian (叶述冕, Department of Chinese Language and Literature at PKU) shared deep insights and provided constant encouragement.

In addition, Chow Chai Khim (周凯琴, NUS Library) and Xu Qingbai(徐清白, PKU Library) offered me professional reference service. Lin Tao (林涛, Department of Linguistics at University of Colorado, Boulder) also helped me to search and deliver relevant papers. Special thanks to them.

I would also express great gratitude to my teachers: Professor Dong Xiufang (董秀芳), who equipped me with a solid theoretical foundation in her morphology course; A/P Wang Feng (汪锋), who reminded me of logical fallacies in my proposal; Dr. Lin Youjing (林幼菁), who taught me how to observe and analyze natural language in fieldworks; A/P Li Juan (李娟), who generously offered me materials of Charles Hockett. Professor Bao Zhiming (包智明), A/P Peng Rui (彭睿), A/P Shi Yuzhi (石毓智), Dr. Wang Hui (王惠), Dr. Xu Zheng (徐政) helped me a lot when I studied at NUS.

Finally, I owe my deepest gratitude to mom and dad, for their wholehearted support.

# 目录

Acknowledgement.....	i
目录.....	ii
统计表一览.....	iii
插图一览.....	iv
Abstract.....	v
能产性理论和汉语复合构词初探.....	1
第一章 引言.....	1
1.1 能产性与“规则+词库”模式.....	1
1.1.1 复合.....	2
1.1.2 屈折.....	3
1.2 来自其他理论模型的挑战.....	4
1.2.1 联结主义模型.....	4
1.2.2 基于使用的语言理论：信息流与频率效应.....	6
1.3 能产性的研究路径.....	7
1.3.1 内省和定性研究.....	7
1.3.2 基于语料库的词汇统计学方法.....	9
1.3.3 类推推理、机器学习和发现程序.....	13
1.4 研究数据.....	15
第二章 汉语双音节复合词能产性的计算.....	18
2.1 研究范围界定.....	18
2.1.1 汉语中的词缀和词根.....	18
2.1.2 复合词和单纯词的界限：语素的识别、同一性和词缀的多义性.....	25
2.1.3 黏着词根复合和词复合.....	31
2.2 汉语双音节复合词能产性的计算方法.....	32
2.2.1 基本计数.....	32
2.2.2 无法落实的概念公式.....	33
2.2.3 基于孤词型频的计算方法.....	33
2.3 实例分析.....	40
2.3.1 黏着词根的能产性.....	40
2.3.2 定位与能产性的关系.....	46
2.3.3 汉语词缀的能产性分析.....	49
2.3.4 词库和句法之间：“1+1”和“2+1”格式能产性对比.....	53
第三章 相对频率和心理词库.....	58
3.1 存储和加工的效率平衡.....	58
3.2 相对频率.....	60
结论.....	68
附录一 词根语素能产性指数(示例).....	71
附录二 三音节词表(示例).....	113
参考文献.....	116

## 统计表一览

表 1-1 语料库语体频率信息.....	16
表 1-2 语料库词频分布.....	16
表 2-1 (类)词缀英汉对译.....	20
表 2-2 词缀多义性.....	28
表 2-3 词根的类型不对称分布.....	38
表 2-4 黏着词根的能产指数.....	41
表 2-5 孤词型频的区间分布.....	43
表 2-6 能产性指数的区间分布.....	44
表 2-7 能产性指数频率分布举例.....	45
表 2-8 定位词根语素的能产性指数.....	46
表 2-9 词根语素能产性指数的对称和不对称分布.....	47
表 2-10 汉语词缀的能产性指数.....	49
表 2-11 联绵词的能产性指数.....	51
表 2-12 “1+X”和“2+X”格式对比.....	53
表 3-1 二字组和三字组相对频率对比.....	62
表 3-2 二字组和三字组相对频率对比.....	63
表 3-3 相对频率与“2+品”结构.....	65
表 3-4 相对频率与“2+器”.....	66

## 插图一览

图 1-1 PDP 加工模型 (Pinker, 1999, p. 105) .....	5
图 1-2 频率的认知机制 (Fenk-Oczlon, 2001, p.434).....	6
图 1-3 Nokia 和 Samsung 变化趋势 .....	12
图 1-4 “渣男”和“暖男” .....	13
图 1-5 类推的有限状态自动机形式 .....	14
图 1-6 语料库词频“长尾”分布趋势 .....	17
图 2-1 基于孤词的趋势预测 (Baayen & Lieber, 1991, p. 812).....	34
图 2-2 整体能产性 (Baayen & Lieber, 1991, p. 819).....	35
图 2-3 孤词和新词的关系 .....	36
图 2-4 “快男”和“跑男” .....	52
图 2-5 “给力”的变化走势 .....	52
图 2-6 “1+家/2+家”和“1+者/2+者” .....	55
图 2-7 “1+手/2+手”和“1+者/2+者” .....	55
图 2-8 “1+品/2+品”和“1+机/2+机” .....	56
图 3-1 词库的复杂度与效率 .....	59
图 3-2 相对频率消长和结构自由/黏着变化 .....	65

## **Abstract**

Native speakers have morphological knowledge of their languages; and this is the reason why they can always understand or invent new words that they have never heard before. Thus, those actual words within a language only form a subset of the speaker's knowledge. In Generative Grammar, language productivity is thought as a combination of two modules--Lexicon and Rules, which proves problematic in some cases. And other theories have different ideas/hypotheses about how morphological knowledge works in speakers' mind. The crucial difference between these versions of theories lies in that whether language productivity is realized as discrete categories or continuous quantities. In recent years, Progress has been made from various perspectives, like corpus-based word frequency distributions, measurement of morphological productivity, frequency effects in language processing, inference and discovery procedures of morphological knowledge. Within these studies, most research has been done on the formation of derivational words; and compounding, as a way of forming words, receives little attention in literature. In this paper, a corpus-based approach is adopted to study the productivity of Chinese compound words by calculating the degree of productivity and distribution of relative frequency.

In chapter 1, we would briefly talk about the problem of language productivity and offer several current theoretical approaches to this problem. In chapter 2, after we define our study scope, we would discuss how to calculate productivity of Chinese compound words; and at the same time, we would use large-scale examples to analyze it. Since language productivity is related to parsing and combining elements, in chapter 3, relative frequency would be introduced to illustrate the combination of morphemes in Chinese in forming compounds and their variations. In summary, this study concentrates on describing rich empirical data both qualitatively and quantitatively; and also it lays a solid empirical foundation for future follow-up experimental studies or machine learning studies.

**KEY WORDS:** productivity, compounding, Chinese, frequency

# 能产性理论和汉语复合构词初探

## 第一章 引言

预期和实现在语言里相接触。

——维特根斯坦

### 1.1 能产性与“规则+词库”模式

能产性(productivity)是语言系统的重要属性(Hockett, 1958; Chomsky, 1965),表现为人们可以理解之前从未听到过的话语,并根据已知的语言形式创造出新的组合形式,表达新的意义。Berko (1958)的经典实验显示,0-4岁的学前儿童可以给从未听到过的人造名词 wug 添加复数标记-s。

美国结构主义语言学家 Hockett (1960)曾将能产性列为语言的十三个设计特征(design features)之一。生成语法理论兴起后,语言学研究开始转向追求理论的预测能力,是否具有能产性被视作区分语言形式间不同性质的重要标志。在生成语法理论中,词法规则的能产性,与创造那些没有列在词库中的新形式相关,而词库则仅仅收录那些无法由规则生成的任意(arbitrary)或特异(idiosyncratic)形式(Chomsky & Halle 1968, p.12)。这种思想可以追溯至 Bloomfield (1933, p.274):“ The lexicon is really an appendix of the grammar, a list of basic irregularities [词库确实是语法的一个附录,是基础不规则形式的一个列表]”。换言之,能产性规则与词库相互依存,确定了词库的收录标准和项目,同时也就确定了需要用规则分析和处理的另一部分语言形式。从学术发展上讲,生成语法理论继承了 Bloomfield 这种思想并付诸实践。

这种模式的语法理论预设了规则(regular)能产与不规则(irregular)不能产的截然对立,在实际研究中将语料放入两个对立的范畴中时暴露出一些缺陷和问题。

### 1.1.1 复合

有的学者认为英语复合词极度能产，性质接近句法，而有的则认为其性质处在句法和词法之间，还有的认为其更接近于词法，其生成方式在早期生成语法理论研究中不断引发争论。Lees (1960)认为复合是底层短语结构的表层实现，将复合视为规则移位的结果。Chomsky (1970)的词汇理论 (lexicalist theory)则支持词汇假说，认为复合是存在词汇规则 (lexical rules) 的证据。而在随后的 GB 理论中，复合又被认为至少一部分是  $\alpha$  移位的结果 (Baker, 1988)。英语复合词，尤其是 NN 结构复合词，性质难以确定表现在几个方面：

一是许多试图划分复合词和短语界限的标准并不清晰有效。NN 结构复合词是英语复合词中最大的子类 (Bauer, 1983)。书面上拼写为一个词或者使用“-”连写的 NN 组合，如 smallpox 和 small-time，通常被认为是词；而对于书面上拼写为两个独立的词的 NN 组合，如 small talk 和 small print，是否可以划清分为短语和复合词两类的界限，Bauer (1998)在对可能的六种标准进行讨论后，认为这些标准都不足以划出这样的界限。

二是英语复合词子类间的能产性存在较大差异。英语中的 VV 结构复合词，与 NN 结构复合词相比，则比较少见，构造也受到更多限制。Tylor (2012, p.36)指出 (to) sleep-walk 是一个标准的表达，而 (to) walk-dream 则不是。

三是 NN 结构复合词的意义结构复杂：a) 有的整体意义无法由成分预测得到，如英语中 airport 的意义已经不太能分解为 air + port (Tylor, 2012, p.37)，相反，母语者很多时候掌握复合词的意义和用法时并没有联系成分 port 的意义；b) NN 结构复合词内部语义关系不止一种，如 sugar house 可以指用糖做的房子，另外也可以指出售糖果的地点；c) NN 结构复合词的语义可能是高度依赖于语境解释的，如“apple juice seat”可能表示“前面桌子上有苹果汁的座位” (Downing, 1977)。这些不规则的特征更像是词的特性。

一方面，英语复合词中的一些结构如 NN 接近于无限能产，按照“规则+词库”理论，应当由规则处理；另一方面，其产生的词则表现地像已有的词项，具有词汇完整性，无法通过规则来预测意义，应当被列入词库，

被母语者一个一个学习，因而产生理论上的矛盾。

### 1.1.2 屈折

尽管英语中许多历史上的构词形态已经脱落，但在名词复数以及动词过去分词形式方面，仍然存在能产的标记方式和历史残留的不规则形式并存的格局，这为观察语言习得和心理词库的工作方式提供了观察的窗口。Bloomfield (1933)曾论述区分名词复数的规则与不规则，并建议将 *ox* 等变化不规则的形式列入词库中：

<b>regular</b>	<b>irregular</b>
<b>box + -es</b>	
<b>fox + -es</b>	<b>ox + -en</b>
<b>cox + -es</b>	

然而，Bauer(2001)指出，在计算机专家中间使用的 *Vax* 一词的复数形式是 *Vax-en*，而非 *Vax-es*。在历史上，英语动词过去分词不规则形式的变化也可以类推扩展到规则形式上，如 *caught, cost, flung, knelt, quit, slung, stuck, strung* (Jespersen 1942)。可见，不规则的形式并非绝对不能产，那么将不规则形式放入有限的词库，而将规则形式放入规则部分处理的做法是否有问题，就令人产生了一些疑问：

(1)对于一些构词模式而言，在何种意义上可以被称之为完全不能产？如果放入词库中的词完全不能产，是否意味着历时上不应当出现不规则形式的类推，而这与来自历时的实际证据相矛盾又如何解释？如果不存在绝对不能产的词，则词库存在的意义是什么，存储的内容又是什么？

(2)能产性是一个是非二元的性质问题还是一个程度渐变的数量问题？如果能产性是一个程度渐变的数量问题，如何度量能产性？能产性和频率是什么关系？

(3)能产性的现象是否可以有“规则+词库”之外的理论来模拟实现？

下文在 1.2 节中针对问题(3)作简要的文献回顾，介绍其他学者的备选思路和理论模型；在 1.3 节中介绍能产性研究常见的几种路径和方法。在第二章，针对问题(2)，在接受能产性是程度渐变的数量假设前提下，界定汉语双音节复合词的范围，结合能产性指数做具体实例分析，探讨汉语双音节复合词的能产性计算问题，并分析基于孤词的计算方法的理论的优缺点。一般来讲，能产性与结构的分解和透明度有关，在第三章，以结构的组合性为切入口，以汉语的字组为观察平面，分析相对频率(relative frequency)对能产性的心理加工的影响，并探讨词库的存储内容，来回应问题(1)和第二章能产性计算的遗留问题。

## 1.2 来自其他理论模型的挑战

### 1.2.1 联结主义模型

早期的生成语法理论建立在离散的符号逻辑和范畴规则基础之上，所作的预测是二值的，不允许连续状态的出现，一旦遭遇反例，线性执行、环环相扣的推导规则就需要大面积修改，使得生成语法理论虽然严密但是脆弱。而联结主义(connectionalism)计算则依靠神经网络(neural networks)的连续数学方法：矩阵(如联结权重赋值)，概率论和统计(如统计推断、学习)等。与以往离散的逻辑符号计算方法不同，通过连续的联结主义计算实现对渐变语言现象的模拟，先后出现了两种不同的、影响较大的思路(Smolensky, 1999):

其一是心理学家 Rumelhart 和 McClelland 及其领导的 PDP 研究小组提出的并行分布式计算的神经网络模型。Rumelhart 和 McClelland (1986) 选择 Bybee 和 Slobin (1982)报告的成人和儿童在英语过去分词规则、不规则形式方面的心理实验数据，进行了计算模拟，挑起了关于规则和频率在语言处理方面的争论(Pinker and Prince, 1988; Pinker, 1991; MacWhinney & Leinbach, 1991 等)。该研究强调对儿童在习得过程中的过度类推(overgeneralization)现象的计算模拟，其核心是从输入层(input layer)输入动词词干的音系序列，激活不同的单元来表征语音形式，通过联结权重的概率差异影响输出层(output layer)过去分词词缀的语音输出形式(见下图)。通

过经验材料的输入训练，计算机可以掌握规则的去分词形式，并且可以像儿童一样，在初始学习期间输出合理的过度类推的错误形式。

联结主义模型在利用语言成分的经验输入得到基于原型的表征后，就不再参照这些单独的语言成分(见图 1-1)。这种模型的缺陷表现在两个方面：首先是该模预测的语言结构模式会出错，其模式与网络节点记录的原型有关，而与单独的语言成分无关；其次是越来越多的证据显示单独语言成分的记忆会影响到语言的加工和输出形式。与之不同的另一种方法是基于典型(exemplar)表征的类推模型，利用大量保留在记忆中的语言成分个例来进行类推(Skousen & Chandler, 2011)，详见下文。

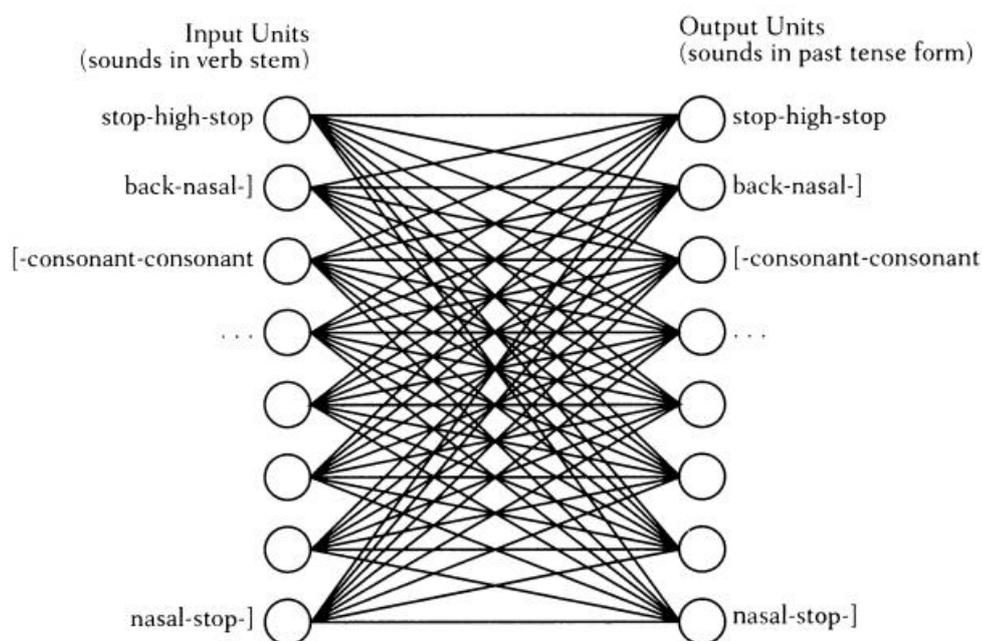


图 1-1 PDP 加工模型 (Pinker, 1999, p. 105)

第二个是随后 Legendre, Miyata 和原 PDP 小组成员 Paul Smolensky 提出的 Harmonic Grammar(Legendre, Miyata, & Smolensky, 1990), 以及演变为后来与生成学派中的 Alan Prince 等学者合作创立的优选论(Optimality theory)模型。这种模型没有使用特定的网络联结，而是直接采用通用的、可以偏离的限制条件，来推导具体的语言现象。优选论在具体实践中面临音系晦涩，理论复杂度和表述变异能力不足等问题。

联结主义心理学家与生成派语言学家的论争结果表现在两个方面：一

方面，一些生成派语言学家开始部分地接受一些能产性较低的模式是由类推产生的，但他们仍坚持高度能产的词法和句法规则，即混合模式的理论模型，包括类推和符号规则两种加工过程(Pinker, 1991; Jakendoff, 2002)；另一方面，这场论争间接导致了生成学派内部的分化，一部分学者在旧理论的基础上开始了新的探索，创立了优选论。这场学术争论大大推进了关于构词能产性的研究，使语言学家增进了对能产性问题的理解。语言学理论如何处理渐变的经验数据，在构词能产性的问题上打开了突破口，也对此后语言学理论的发展产生了深远影响。

### 1.2.2 基于使用的语言理论：信息流与频率效应

基于使用(usage-based)一词最早出现在 Langacker (1987)，指母语者对语言的特殊知识，而后被用来指涉一系列假说，其核心是语言使用是语言结构形成的来源，二者密不可分。这种思想可追溯至 Greenberg (1966)阐释使用频率在决定音系、形态和词库的标记性(markedness)方面的重要作用。该理论假设母语者的语言知识由经验塑造，语言成分的语例(token)会通过范畴化进入记忆系统，并进而影响认知表征(Bybee, 2011)。Fenk-Oczlon (2001)认为语言成分的频率并不直接作用于语言结构，而是首先影响认知过程(见图 1-2)：高频率的语言成分会增加人们对该成分的熟悉度(familiarity)，而降低在感知和输出该成分时的认知成本(cognitive costs)。认知成本、熟悉度这些概念都是相对性的，一个语言成分对某些人而言是不熟悉的，可能对另外一些人来讲则是熟悉的，或者在过一段时间后就会由不熟悉变得熟悉。

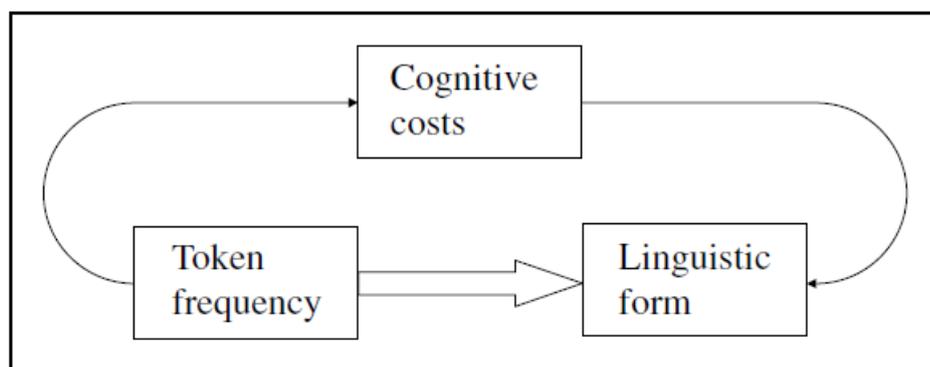


图 1-2 频率的认知机制 (Fenk-Oczlon, 2001, p.434)

使用频率影响的认知过程包括学习、记忆和感知(Ellis, 2012)。学习幂律(the power law of learning)描述了许多认知技能学习中练习和表现之间的联系,即练习越多,表现越好,尽管练习的作用会逐渐衰减 (Anderson, 1982; Newell, 1990)。

在语言表征、加工和演变中,目前已知的频率效应(frequency effects)表现在几个方面:

1)一个高频率成分由于频繁使用而有更稳定的记忆表征,通常更能抵抗词形变化的规则化而保持不变,如英语中动词过去分词的不规则形式,即高例频的保守效应(Conserving Effect of high token frequency)(Bybee, 2010);

2) 高频词比低频词更容易识别,有歧义的词汇更容易被理解为高频的词义(Jurafsky, 2003);

3) 结构的组合频率会影响语感的合法 (well-formedness) 判定 (Pierrehumbert, 2003);

4) 听者的概率推断会导致语言演变,并可以用来解释词汇的历时规则是如何维持运转的(Zuraw, 2003)。

### 1.3 能产性的研究路径

能产性的研究方法包括使用内省语感(intuition)、语料库(corpora)频率统计、心理学实验(experiment)设计和机器学习(machine learning),这些方法并不互斥,相反,因为相互补充而在实际研究中经常混合使用。

#### 1.3.1 内省和定性研究

Van Santen (1992) 强调能产性存在于既有词和不可能边界之间的空间中(转引自 Bauer, 2001)。如果能产性是有关新造词的产生,那么它一定要能满足能够判定新造词形式的合法性,并且与既有的词汇进行相互对照。Bauer (2001)在梳理了相关的既有词(existing words)、新造词(new words)、潜在词(potential words)和可能词(probable words)几个概念后,认为如果接

受这样的理论假设，那么就必须回答下面的几个问题：

首先，既有词是构词能产性的前提和基础。Aronoff (1976, 1983) 强调既有词(actual words)应当由词库定义。这里的问题在于既有词是如何存在的？如果既有词以大型工具书的形式存在，那么词典收录词汇都具有一定的时效性，如何保证其中的词汇仍在使用而不是已经被废弃？其次，一个词要成为一个公认的既有词，必然要为言语社团的次级群体所使用。Bauer (2001) 以 Oxford English Dictionary (OED) 为例说明，词典收录存在标准偏好，以致于在我们能够捕获的词和我们认为需要研究的既有词之间存在一定的偏离。在词典编纂和二语教学活动中，高频的组合类型是词典学家和二语教学老师们主要的兴趣所在。由于词频分布的齐夫律(Zipf's Law)的存在，能产性的构词组合在语料库中有相当大量的低频词汇出现，其结果是大量的不能产的构词组合和一部分更高频的能产性组合被收录入词典中。而能产性的构词组合形式的收入数量，要依据词典的编纂规模而定，因为词典的有限容量可以容纳有限的不能产的构词组合形式，却无法容纳相当大量以至接近于无限的能产组合(Baayen & Lieber, 1991)。

其次，在传统的语言学理论中，根据一个新造词是否最终为言语社团所接受，从历时角度可以区分偶造词(nonce)和新词(neologism)。所谓偶造词，就是在特定场合临时地创造出来的词；而新词则是成为言语社团使用语言一部分的词。这里的问题在于在一个词刚刚被创造出来的时候，语言学家无法区分该词是偶造词还是新词。如果接受潜在词(potential words)的概念，那么不论是偶造词还是新词，在语言结构上并不存在本质差异，都是由语言系统提供的造词方式，那么这两种类型的词都应该成为构词能产性的例证，而不是像一些语言学著作中给偶造词贴上“修辞”或“临时活用”的标签，排除在外。

第三，能产性始终与潜在性问题相关。潜在词的概念直接源自生成语法框架下的词法研究。潜在词假定了语言系统提供了造词方式，由于空间并未填满，而产生词汇空缺(lexical gap)的现象，因而有些词只是潜在地存在，所以 Aronoff (1976) 认为词法学的主要任务应当是预测说话者能输出的构词组合类型。按照 Aronoff 的词汇理论，填充词汇空缺的词通常会被词汇系统中已有的同义词阻断(blocking)。值得注意的是，这里的问题在于实

际的语料中存在不以填补词汇空缺为目的的造词的潜在性，即在词汇系统中存在特定词汇的情况下，还会引入同义的新词，如英语中的动词 **Orient** 和 **Orientate** (Bauer, 2001)。此外，语言系统中也存在出现不符合语言系统规则的新造词。

最后，由于母语者并未穷尽语言系统提供的词汇形式，因而并非所有的潜在词都有可能变为现实。可能词(**probable words**)与潜在词中不太可能实现的词相对立。自足的语言系统可以界定潜在词的边界，但能产性更关心哪些词形是更加可能的？一些语言学家试图区分不同的限制因素，确定各种限制因素的性质，即哪些是语言系统内部的，哪些是语言系统外部的。

依靠内省，我们能对构词模式的能产性作出较为简单的、离散范畴性的判定，如能产/不能产、能产/半能产/不能产，或者是具体的语言成分/构词模式间的能产程度比较，而在如何落实这些概念的实证上，则显得能力不足。总的来说，内省是检验关于能产性学说和论断的重要标准，但是单纯依靠内省，无法对能产性作出更进一步的刻画和检验方法。一些学者开始寻求更客观的方法来拓展构词能产性的研究，检验已有的学说。自 20 世纪 60 年代 **Brown** 语料库建立以来，计算机的性能和文本处理技术有了长足的进展，英语语料库建设方面的理论积累和技术铺垫，为这种拓展提供了可能。

### 1.3.2 基于语料库的词汇统计学方法

#### 1.3.2.1 词典是统计既有词的可靠材料吗？

**Baayen** 和 **Lieber**(1991, p. 803)指出，要想落实对 **Aronoff** 提出的既有词(**actual words**)进行实际测量，那么必须首先承认存在一份相对合理的词表。一个大型的平衡语料库可以提供一个具有代表性的语言样本切面。而比较词典和语料库这两种文本集合，**a)**语料库可以提供词频信息，**b)**并且包含词典没有列举的高度能产的新造词，相反，出于商业和使用价值的考虑，词典更加偏好收录那些频率较高、语义凝固的词，此外，**c)**词典具有滞后性，还可能收录了实际生活中已经不再被使用的词汇。然而，这种选择并不是绝对的(见下文)。

### 1.3.2.2 语料库词频表获取的工作程序

词和其他语言单位的频率信息是语料库语言学研究的重要基石 (Abney, 1996)。文本的分布统计的基本数据是例频(token frequency)和型频(type frequency)。

所谓型频是指不同词型的数量，例频是相同词型的不同数量。以具体文本为例，在下面这段话中，“苹果”一词共出现 3 次，其例频为 3，在整段话的型频统计中，不同的词例如果是相同的词型，如“苹果”，那么计作是 1 个词型，即型频为 1。

“**苹果**是异花授粉植物，大部分品种白花不能结成果实。**苹果**能够适应大多数的气候。在南北纬 35-50 度之间是**苹果**生长的最佳选择。”(摘自百度百科)

其工作程序主要包括：

(1) 文本预处理：根据研究目的决定是否清理文本中的标点符号、数字字符及其他字符；切分词例(token)，如 didn't 是否处理为 2 个词。

(2) 词型(type)的同一(type identification)，即确定不同的词例是否属于同一词型。如英语中大小写不同的单词是否需要归为一类，有的如 The 和 the 应视为同一个词的不同形式，但是 Goldsmith 是人名，而 goldsmith 是金匠，则不应归为一类。德语中的限定词 der, die, das 是与后接名词性、数、格保持一致的不同变化形式，是否应当归为一个词位(lexeme)? 同形异义的“花(~朵)”和“花(~钱)”是否需要分开计数？汉语文本中主要需要处理的词型同一问题是异形词(如“掺假”和“搀假”)和多音词(如“教 jiào 授”和“教 jiāo 授”)。根据国家语委先后于 2001 年、2004 年发布的《第一批异形词整理表》、《第二批异形词整理表》(草案)和自己的认识，我们对词表中出现的异形词进行了取舍，保留一项进行统计。由于在实践操作中无法对词表中的每个多音词在语料中逐个确认读音、统计词频，因而相关数据不纳入统计范围。

(3) 得到词频表并分析频率分布模式。词频表包含两个基本数据：词例的数量，即语料库样本大小；词型的数量，即词表样本大小。词频表可以重新组织为排序频率表(rank/frequency profiles)和频率谱(frequency

spectra)来研究词频分布。

### 1.3.2.3 来自历时的证据

Plag (1999)的研究特色之一在于选择以 Oxford English Dictionary(OED)作为材料来源,巧妙地利用 OED 词典标注的文献历史时间信息来使用词典数据。针对 Baayen 和 Lieber (1991)批评词典的说法,他为选择 OED 词典的立场提出回应的意见:首先,词典编纂过程中的确倾向于收录高频、语义凝固的词,因为许多能产的新词语义可以被预测,而没有被收录其中,这种偏差对其他词典来说是致命的,但 OED 作为体量巨大、详尽涵盖英语词汇发展的历史词典,则未必。尽管 OED 也存在这种偏差,但这并不意味着 OED 的材料对观察和描写能产性是无效的,这种偏差可能导致材料无法准确测量能产性的等级,但是在辨别构词能产性的有无这一点上,仍然是可靠的材料。其次,由于 OED 的词条义项都标注了最早出现的时间点,因而词典的滞后性问题完全可以通过切割时间段来避免。

此外,监测语料库(monitor corpora)随着时间变化不断收录和扩充不同性质的语料(Sinclair, 1991),因而可以通过新词语的变化趋势来提供构词能产变化的证据材料、检验相关的理论推断。伯明翰大学(the University of Birmingham)的 John Sinclair 主持的 BOE(the Bank of English)语料库和美国的 COCA(the Corpus of Contemporary American English)是著名的英语监测语料库。中国国内的张普 (2004)提出了类似的“动态流通语料库”的观点,来辅助发现动态语言知识更新和新词语提取。2004-2005 年中国教育部陆续与国内 5 所高校合作建立国家语言资源监测与研究中心,动态流通语料库以每年 12 亿字符,越 10 亿汉字符的数量增加,其监测结果从 2005 年开始以《中国语言生活状况报告》绿皮书的形式在商务印书馆出版(张普, 2009, pp. 346-7)。Baayen 和 Renouf (1996)利用泰晤士报(Times) 1989-1993 年的文本建立历时的监测语料库,通过对文本的词汇逐层过滤,来研究特定的英语词缀创造新词的时间进程。

第三,与建立监测语料库相似的是将互联网作为语料库的研究思路。互联网语料库存在一些问题(McEnery & Hardie, 2012):首先,与其他语料库相比,互联网语料库的内容没有按照语体分类,文本呈现的是混合分布;

其次，互联网文本中有很多书写的错误字，如“奢糜(误) -奢靡”，这些噪音会干扰对语言数据的分析。

近些年来，随着互联网产业的发展，搜索引擎的技术进步和数据科学(data science)的兴起，许多新词在搜索引擎网站上的搜索次数可以作为一个统计指数来间接反映其历时的变化趋势。下图分别是“谷歌趋势(Google Trends)”和“百度指数(Baidu Index)”中反映的热词变化。第一幅图中显示的是手机厂商诺基亚(Nokia)和三星(Samsung)在手机通讯行业中的兴衰交替。智能机兴起后，诺基亚无法及时调整产品战略，由盛转衰，而三星由于加入谷歌的安卓系统阵营，在智能机领域中遥遥领先，其反映的商业兴衰也可以由搜索指数间接反映出来(见图 1-3)。第二幅图显示了汉语中最近出现的新词“渣男”和“暖男”的热度变化趋势，其反映的走势可以作为考证词源和起讫日期的一个重要参考依据(见图 1-4)。

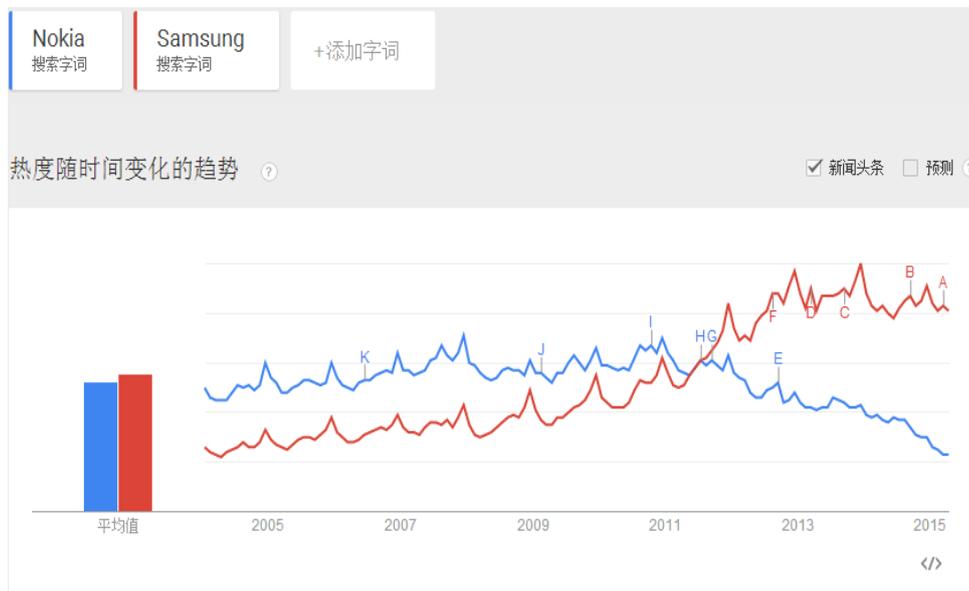


图 1-3 Nokia 和 Samsung 变化趋势 (<http://www.google.com/trends/?hl=zh-CN>, 检索日期 2015-4-21)



图 1-4 “渣男”和“暖男” (<http://index.baidu.com/>, 检索日期 2015-4-21)

尽管语料库提供了丰富的数据来观察词汇的能产变化，但是不同构词模式间能产性的差异是什么导致的？是语言系统内部的结构原因还是语言系统的外部因素？这种因果关联的探究依然是一个悬而未决而又非常重要的问题。语言理论的构建不能停留在语言数据的记录上，必然涉及到因果关系的论断。在观察语料基础上形成的假设，除了可以借助心理实验来验证之外，还可以通过建立推理模型来实现模拟。

### 1.3.3 类推推理、机器学习和发现程序

基于典型表征的模型进行类推的基础是留存在记忆中的大量语言实例，这种观点与早期生成语法中将词库冗余成分最小化(minimal redundancy)的做法(Kiparsky, 1982)是不同的。事实上，这种存储经济优先的做法有其特定的历史背景，早期的个人计算机文件存储容量相当有限，只有几十 KB，因而形式语言的生成系统将评估存储的经济性作为一个重要议题(Baayen, 2003)。然而人的大脑记忆系统的容量却不存在此类问题，换言之，在特定历史背景下出现的存储经济性优先的理论模型并没有多少心理学证据的支持。而且这种做法忽视了在给定的心智系统中存储的经济性(economy of storage)和心理加工的经济性(economy of processing)之间存在平衡取舍的问题，即过度强调存储的经济性，必然会尽可能增加更多的系统规则。与

记忆提取相比，规则的调取和施用需要占用更多的加工时间。而语言作为一种实时(real-time)的交际活动，人们不可能无限制地放缓语速导致交际困难和障碍，因而加工效率优化要优先于存储经济。

记忆大量词例是学习一些能产构词模式必不可少的基础。Aitchison (2012)指出，以英语为母语的儿童在早期创造新词时更倾向于使用复合或类转(conversion)而不是词缀添加，即使到 10 岁以后的青少年阶段，离接近于成人的词缀类推能力还有不小的差距。在一个实验中，青少年表达小 wug 的概念时使用的表达包括 pug-wug, wuggist 和 wiggle, 仅仅有小部分被试与成人的语感一致，使用 wuglet 的形式。

比较著名的模型包括概括语境模型(The Generalized Context Model) (Nosofsky, 1992), 基于记忆的学习模型(Memory Based Learning)(Daelemans & van den Bosch, 2005), 类推模型(Analogical Modeling)( Skousen, 1989)。

支持类推模型的一个例子是德语和荷兰语中的 NN 结构复合词。德语和荷兰语构造新的 NN 结构复合词，有时会在两个名词中间插入一个-s-的构形成分(formative)，无实义(有的文献称为中缀，interfixes)，如德语“Krankheit(病)-s-zeichen(标记)，病症”。在 CELEX 数据库中，有 69%的荷兰语 NN 复合词同英语 NN 结构复合词相似，没有-s-的构形成分，而剩余复合词中包含约 20%的-s-，11%的-en-及其他边缘中缀成分。由于各种规则都存在不少例外，在直接成分上使用类推策略要比归纳规则更加有效(Krott, Baayen, & Schreuder, 2001; Krott, Schreuder, Baayen, & Dressler, 2007)。

从更为普遍的推理机制上看，类推应当属于学习过程和发现程序中的一个环节，而对语言输出的稳定控制，在一定程度上应当依赖于从语言材料中获取的词法结构知识。而语言成分的形式类推构造，close:closes:closing → act:acts:acting, 可以表示为有限状态自动机形式(见图 1-5):

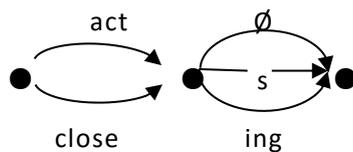


图 1-5 类推的有限状态自动机形式

John Goldsmith 的 *Linguistica* 项目顺着这个思路展开研究，致力于从文本中自动发现和切分词汇，自动分析词法结构和获取词法知识，在这个意义上，与 Zellig Harris 发展分析语言材料的自主方法的目标是一致的，即发现程序(discovery procedure)。尽管类推提供了构造新形式的可能，但是提供进一步评估类推构造形式的工作方法也是必要的部分(Goldsmith, 2009)。或者更宽泛地讲，如何从经验数据中发现可靠的归纳方式是学习理论(learning theory)的重要问题，为此，必须区分两类方法，一类方法是观察数据的特征并对其分类，或预估变量的数值，另一类方法则是通过使用经验数据来发现、挑选这种分类或预估方法的方法，这种发现归纳方法的核心就是模式识别(pattern recognition)。未知的环境概率分布(background probability distribution)会决定分类方法的可靠性，而将意外错误最小化的贝叶斯原则(Bayes rules)可以挑选最优的分类方案(Harman & Kulkarni, 2012)。

借助不同的研究方法，语言学家拓展了构词能产性的材料获取和观察方法，辨别分析了能产性的概念维度，以及在能产性输出的推理和建模方面的进行了有益的探索。尽管构词能产性方面的研究还远远不那么令人满意，但是语言类型学材料的增加、研究方法的多样化，都将对增进认识能产性有所贡献。

#### 1.4 研究数据

数据来源主要是基于汉语语料库的词表，由 Lancaster University 的肖忠华博士提供。该语料库的词表基本信息如下(见表 1-1, Xiao, Rayson, & McEney, 2009, p. 8):

表 1-1 语料库语体频率信息

文体	词例数	字数
口语	2,692,315	3,824,579
新闻	12,147,572	20,185,322
小说	11,973,365	16,424,649
纪实文学	11,900,160	17,954,729
总计	38,713,412	58,389,279

下表显示了该语料库词表中的顶级和底层的词频分布状况(见表 1-2):

表 1-2 语料库词频分布

顶级词频			底层词频		
排序	频率	词	排序区间	频率	举例
1	2321615	的	51699 - 52989	10	胶印 麦芒 扫尾
2	739321	了	52900 - 54379	9	衰朽 豆奶 领唱
3	691153	是	54380 - 55929	8	滑石 重水 旁及
4	598740	一	55930 - 57660	7	纲目 围屏 地标
5	525029	在	57661 - 59548	6	包场 矿层 工位
6	417646	我	59549 - 61653	5	语序 填料 倒票
7	408160	不	61654 - 64180	4	脱靶 底板 狙杀
8	335000	他	64181 - 67288	3	歹意 闲雅 玉兔
9	326169	和	67289 - 71792	2	抄件 酣睡 羽冠
10	294690	个	71793 - 84833	1	屈体 道长 等角

词表的词频顶部主要是语法虚词，并且词频急速衰减，第 1 位的“的”词频是第 2 位的“了”3 倍多。随着排序下降，这种词频差异逐渐减小，但是其区间逐渐扩展。词频排在前 10 位的词型(词例之和为 6657523)占据整个词表的词例总和(38713412)的 17%。

而在词表底部，词频关联的数量更多，且随着词频衰减而增加。词频为 3 的词例有 3108 个(64181 - 67288)，词频为 2 的词例有 4504 个(67289 - 71792)，词频为 1 的词例有 13041 个(71793 - 84833)。词频为 10 及其以下的词型共 33135 个，占词表词型总数(84833)的 39%，但其词例仅占词表的词例总数(38713412)的 0.3%。极少数高频词型分布有大量的词例，而大量低频词型只分布有极少数的词例数量，构成了低频词的“长尾”，这种扭曲的分布与经典统计学中的分布迥异(见图 1-6)：

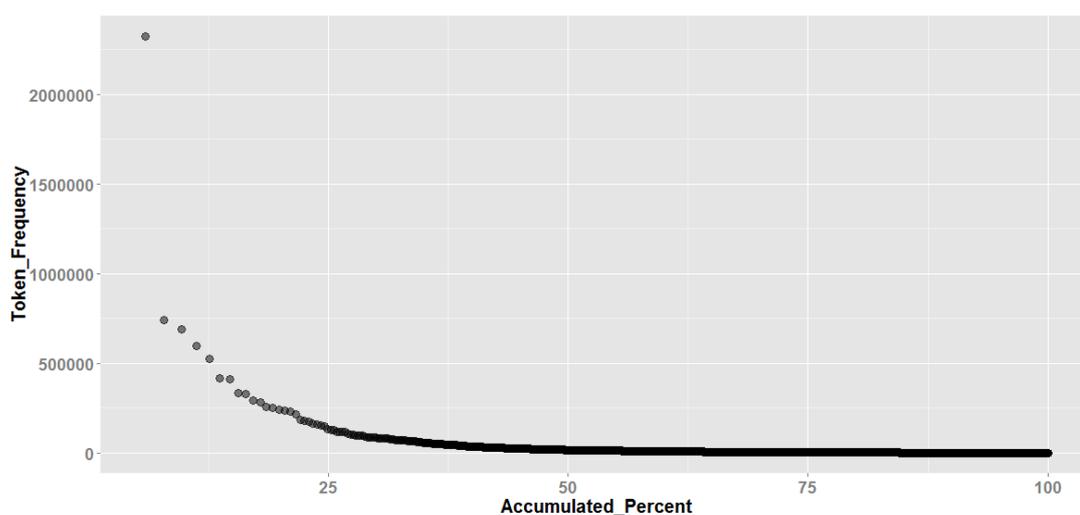


图 1-6 语料库词频“长尾”分布趋势

上图中的点代表词表中的词型，y 轴数值是不同词型的例频数值。不同词型的例频数值占整个词表词例总数的百分比，在 x 轴上从左向右，依次相加累积，就是上图中的 x 轴数值，累积百分比(accumulated percent)。这种频率分布结构并不限于该语料库的词频分布，而是广泛见于自然语料文本中。Zipf (1949, 1965)是最早对这种扭曲的词频分布结构进行系统研究的著作。

值得注意的是，例频为 1 的并非完全是新造词，其中的“道长”等词对于汉语母语者来说，并不陌生。例频为 1 的词型数量占词表词型总数的 15%，随着语料库容量的增长，会有更多的低频词进入语料库。在有 1 亿词容量的英语 BNC(British National Corpus)语料库中，这一比例高达 52%(Leech, Rayson, & Wilson, 2001, p. 9)。

此外，本文也使用一些自省和来自互联网语料库的语例。

## 第二章 汉语双音节复合词能产性的计算

### 2.1 研究范围界定

本小节界定汉语复合词的范围问题包括三个方面：

- 1) 复合词和派生词的界限；
- 2) 复合词和单纯词的界限；
- 3) 复合词和短语的界限。

由于汉语词缀的判定标准各家学者观点不一，导致一些词在判定是复合词，还是派生词上，存在模糊性。马庆株 (1995)认为“词缀的范围影响构词法的格局，是派生词与复合词之间划界的关键”。其次，不透明的复合词还会存在语素识别和同一的问题，本文将探讨语素的同一和定义。此外，二字组结构中还存在词和短语界限划分的问题。本节将围绕以上几个问题简要论述词表和计算对象的筛选标准。

#### 2.1.1 汉语中的词缀和词根

词缀的定义和标准通常包括 (1) 意义抽象、虚化, (2) (不是词根的)黏着语素, (3) 是词而非短语或句子的直接成分(Haspelmath & Sims, 2010; Carstairs-McCarthy, 2006, pp. 83-84) 三点。其中，第3条标准主要是与黏附形式(clitics)相区分，如 *She'll come tomorrow* 中的'll，尽管在音系上与 *She* 更近，但在语法上修饰的是短语 *come tomorrow*。

词根是构词的语义核心成分，在实际的结构分析过程中，作为自由语素的词根通常会被先分析出来，而剩下的、依附在词根上的非词根黏着语素，通常就是词缀，如英语词 *reconsideration*，*re-*和*-ation* 词缀依附于词根 *consider*。换言之，词根和词缀的区分方法行之有效的语言结构基础是形式不对称。词根和词缀的不对称(*root-affix asymmetries*)与许多音系现象有关，一般来讲，词缀在音段数量和组合模式的选择上都较为有限(Jakobson, 1965)，而词根不然；词根需要遵守的音系长度限制词缀可以违

背 (Inkelas, 2014)。此外, 一些特定的音系模式和特征都只出现在词根或词缀上, 如在许多语言中, 挤喉音(ejectives)只出现在词根上(Bybee, 2005)。而对于汉语来说, 这种形式结构上的不对称特征并不显著, 相反, 汉语的单音节语素占绝大多数, 语音语义间具有较为稳定的对应关系, 语素与语素间的边界清晰 (徐通锵, 1991,1994,1997; 王洪君, 1994), 因而确定汉语中的词缀范围并非易事, 各家著作也分歧较大。总的来说, 汉语的语素音位对应模式在词根和词缀不对称方面并不显著, 这是确定汉语词缀困难的一个重要的结构原因。

在讨论汉语词缀的标准之前, 这里厘清一下本文使用的“自由”和“黏着”术语的意义。Bloomfield (1926, p. 155)对“自由”和“黏着”的定义是“A form which may be an utterance is free. A form which is not free is bound.” Bloomfield 的区分标准是是否能单说为一个话段, 国内不少著作沿用这一说法, 如吕叔湘 (1962), 朱德熙 (1982)。在这个意义上, 自由/黏着同是否“单说”大致等同。但是后来一些外国语言学家在继承和使用这对术语时, 对“自由”和“黏着”的区分标准改为是否成词 (Bauer, 2004; Trask, 2007; Crystal, 2008), 因为成词也不一定能够成句, 而是否成词则更容易在具体条件和范围下被确定。陈保亚 (1999)在讨论提取词的方法时也对“单说论”混淆不同层面提出批评。本文将在是否成词意义上使用这对术语, 因而与语素分类中“成词/不成词”标准大致等同, 尽管汉语中的“词”并不容易确定, 这是要特地说明的。

学界提出的汉语词缀和黏着词根的区分标准主要包括以下 4 类:

(1)意义标准: 意义虚化(几乎所有著作均提到)

与词根语素相比, 词缀的意义倾向于概括、抽象和虚化, 各家根据意义标准所得到的词缀数量和范围相差很大。此外, 许多著作在列举词缀和“类词缀”的时候, 通常都无意受到外语词缀对应的语义概念的暗示。近代以来汉语同其他外语的密切接触, 特别是翻译西方著作, 使汉语中批量出现一些表达形式, 举例如下(表 2-1):

表 2-1 (类)词缀英汉对译

(类)词缀	英语词缀	例词
非-	a).non-	非关税(nontariff)、非物质(non-physical)
	b).in-	非正式(informal)
	c).un-	非常规(unconventional)
	d). ?	非卖品(X is not for sale)
单-	a). mono-	单核(mononuclear)、单语(monolingual)
	b). uni-	单极(unipolar)
	c). ?	单向街(one-way street)
-化	a).-ize/-ization	全球化(globalization)、数字化(digitalize/digitalization)
	b).-ify	美化(beautify)
	c).-en	恶化(worsen)
	d). ?	老龄化(aging)
-学	-ology	社会学(sociology)、生物学(biology)、地质学(geology)

在接触的过程中,汉语逐渐形成较为系统的对应模式。值得注意的是,这些语素成分的英语词缀来源有时并非是唯一的,而是一对多的形式。上表中的“老龄化”一词对应的是英语 **aging**,指逐渐变老的过程,以及“非卖品”一词,这些都不是直接翻译的结果,而更多的是汉语构词系统自身构造的结果。赵元任(1968)认为这些新兴“前缀”、“后缀”也是复合词中组合能力较强的语素,之所以在文中加上引号称之为“前缀”、“后缀”的原因,是由于来自外语翻译,并且这些词缀会改变原有词干的语法功能(功能标准见下文)。

此外, **-ology** 在英语构词法中是否应当划为词缀也是存在争议的。英语中存在一批从希腊语和拉丁语借入的语素,被广泛用于创造新的科学术语,如 **bio-**和**-ology**(Bauer, Lieber, & Plag, 2013, p. 441):

biodegradable	geology
biorhythm	narratology
biofuel	euology
bionomic	analogy

Bauer(1979, 1983)指出如果将这些成分归为词缀,那么会出现一个词是由两个词缀组成、没有词根结论,如 **biology**。因而,一种可行的分析方法就是将其分析为黏着词根,其构成的词是复合词, Bauer 称之为新古典复合词(neo-classical compounds)。由于这类复合词构成的词根黏着,在句子层面没有分布,因而也无法像其他复合词一样用词类标注结构,如 **NN, AN** 结构等。

可以相互对照的是,大妈(Dama)、土豪(Tuhao)这些汉语词汇借入英语中,被收入牛津英语词典(OED),不能简单地依据汉语词汇的语义结构来确定其是否为英语复合词一样。

因而语素成分是否应当归入词缀的范畴,不应当简单依照词语在源语(source)中的语义和语法类别来决定,而仍应当参照词缀的语言类型共性和目标语(target)的结构特征来决定。

(2)形式标准:定位 (朱德熙, 1982; 郭良夫, 1983; 马庆株, 1995)

朱德熙 (1982)指出:“词缀都是定位语素”,“真正的词缀只能粘附在词根成分上头,它跟词根成分只有位置上的关系,没有意义上的关系”。定位是词缀的重要特征,因而可以依据词缀粘附词根的相对位置,对词缀进行分类,尽管这些不同位置出现的概率并不相等。一般来讲,跨语言的证据显示后缀要比前缀更常见,较早指出这一点的是 Sapir (1921, p.70): “Of the three types of affixing—the use of prefixes, suffixes, and infixes—suffixing is much the commonest. Indeed, it is a fair guess that suffixes do more of the formative work of language than all other methods combined”,之后陆续为更多的类型学数据证实(Greenberg, 1963)。此外, Matthew (1991)指出,有时也会出现一个词缀成分的位置和类型不易确定

的情况，有时在一些例子中可以充任中缀，有时在另一些例子中充任前缀或后缀。拉丁语中的一个鼻音成分在词中的结构可以表示为 X V nasal (C)，该鼻音成分是中缀还是后缀取决于词末辅音成分是否出现。

马庆株(1995)区分了绝对定位和相对定位，“绝对定位是指位置无条件地固定，即不管是在构词平面上还是在句法平面上都是定位的”，“相对定位有两种情形：(1)构词平面的定位与句法平面的定位不一致……(2)相对定位还指在词缀义 A 上定位，在非词缀义 B 上不定位，AB 两个意义相关。”

词缀都是定位语素，反之，定位语素未必都是词缀，也可以是实义的词根语素：

前字(定位)	举例	后字(定位)	举例
咨	咨询、咨政	虑	思虑、考虑
匡	匡扶、匡助	龄	学龄、高龄
巩	巩固、巩膜	谊	友谊、情谊

事实上，许多著作中筛选词缀时并未严格执行定位标准，因而得到的结果也相差很大。如 Packard (2000) 列举的否定前缀“无-”、“非-”，这些语素在后字位置都存在分布，如“毫无”、“尚无”；“若非”、“并非”、“绝非”。严格遵守定位的限制标准会大大缩小词缀的范围，本文认定词缀将严格执行定位标准。

(3)数量标准/能产性标准：词缀可能是能产的(Dai, 1992)；词缀比黏着词根能产(Packard, 2000)；能产性量级：类词缀>词根>词缀(王洪君、富丽, 2005)

使用数量标准作为词缀鉴定标准的做法主要见于汉语构词法研究文献中。尽管许多学者使用了“能产性”的概念作为数量标准，但却没有像 Plag (1999)、Bauer (2001)进行多个维度的详细辨析，因而大多数学者使用的“能产性”概念大致等同于词表或词典中既有词的数量，只有王洪君、富丽(2005)区分了既有词频和“类推潜能”之间的区别。

Dai (1992, p. 146)在列举词缀的能产性标准之后指出：“Although the

attachment of *lao-* in Chinese is not productive, it is much more productive than the attachment of English *a(c)-*, which occurs perhaps only in acknowledge.”首先，他将包含词缀的既有词数量等同于能产性，因而得出汉语中的“老-”要比英语中的 *a(c)-* 更加能产的结论；其次，跨语言的能产性对比是否是可能的，作者并没有论证；第三，作者认为汉语词缀“老-”是不能产的，同时又比英语词缀“*a(c)-*”更加能产，十分令人费解。由于没有严格界定能产性的计算方法，Dai (1992)认为“预-”（预习、预报、预测），“-器”（瓷器、漆器、玉器），“-体”（物体、人体、气体）也可以划为词缀。

Packard (2000, pp. 72-73)根据《倒序现代汉语词典》中收录“-员”和“-者”的词频信息（以“员”结尾的词语型频为 76，以“者”结尾的词语型频为 26），认为“-者”更具有能产性，因为词典中“-者”结尾的词语型频较低暗示了仍有大量以“者”结尾的词语没有进入词典，并进一步认为这是论证“-者”是词缀，“-员”是黏着词根的一个证据。首先，词典收录词条更倾向于收录高频的词语，但是词语型频数值低并不能推导出仍有大量词型没有进入词典的结论，也有可能词典收录的是不能产的词语，这部分词语因为高频而仍保存在语言系统中并被收入词典，即词典收词的频率标准与构词能产与否没有直接关系。其次，Packard 将汉语词缀比黏着词根更能产作为论据使用却没有论证。在下文中可以看到有的黏着词根的能产性指数并不比词缀低，尽管能产性指数的比较和词缀的划分可能有循环论证之嫌。

词根词缀的划分是基于形式结构的标准，而能产性的度量是基于数量标准，已知的语言证据并没有显示二者之间存在简单的一一对应关系，换言之，词缀和词根的个例理论上都可以表现出高能产性或低能产性。赵元任(1968)在分析汉语的前后缀时候，也将复合词中组合能力强的语素，放在一起并列考察，以示数量无法区分词根和词缀。由于能产性计算的问题没有严格论证，单凭直觉和既有词频无法得到一个较为理想的对能产性的测量，因而大部分试图用能产性的数量标准来界定词缀的方法难以令人信服。

(4)功能标准：标识特定的语法功能(Dai, 1992; 陈光磊, 1994; Packard, 2000)

功能鉴定标准认为在词干上添加词缀，通常会改变原有词干的题元角色(thematic role)或者词类性质。这里的问题在于，不管是添加词缀还是添加黏着词根都会带来整体语义成分的改变，比如“-者”和“-员”都可以用来指称一个施事者，这样的标准并不足以将二者区分为黏着词根和词缀两类，也就是说 Packard 将二者区分为两类的这条理由并不充分。而识别词类性质的改变，首先面临的问题是黏着语素不能自由成词，无法通过观察在句法层面的句法分布来获得词类；其次，汉语的词缀对词干成分的词类性质和音系特征并没有特别明显的限制和选择。

以上讨论了几种可能的词缀鉴定标准。在没有详尽讨论构词能产性的基础上引入含混的数量标准，并不能解决区分词缀和黏着词根的问题，原因在于词根与词缀的区分在很大程度上是基于结构的划类，而不是数量。由于使用严格的定位标准，因而词缀范围被大大缩小。汉语中的后缀大多带有语音形式上的弱化。本文中讨论的词缀会局限到“阿-”、“老-”、“-子”、“-儿”、“-头”、“-巴”、“-乎”、“-们”、“-么”几个。

类词缀(semi-affix)的思想和说法较早见于 Marchand (1969)。Marchand 使用的术语不是类词缀，而是类后缀(semi-suffix)，指英语构词法中介于完整的词和后缀中间状态的成分，如-like(manlike, godlike, daylike)，-worthy(unpraiseworthy)，由于只分布在复合词的后位，这些构词因而介于复合词和派生词的边界。通常支持设立“类词缀”的学者所持理由通常包括意义和数量标准两个方面。意义上没有词缀虚化程度高(吕叔湘, 1979等)，然而如上文所述，意义的虚化程度难以把握，有时还受到外语语义概念的暗示，很难作为判定汉语结构的标准。而数量标准指由于在能产性计算问题上没有严格论证，因而理由也不充分。朱德熙 (1982)指出“-性”、“-式”、“自-”等是词根语素，理由是它们是不定位语素，但马庆株 (1995)主张区分义项，这些类词缀义可以是定位的，这里涉及到语素的同一性、词缀的多义性和西方词法学研究中的传统处理方法，见下节 2.1.2。但在这里要指出的是，在西方语言学文献中，一个词缀成分可以是多义的，但不

存在一个同形的构词成分根据不同的意义，被归属为词根和(类)词缀的情况。

### 2.1.2 复合词和单纯词的界限：语素的识别、同一性和词缀的多义性

由于汉语中真正的词缀语素很少，大部分都是词根语素，因而区别复合词和单纯词的界限与识别、提取语素问题密切相关。

Bloomfield (1933, p.161)将语素定义为内部不包含部分音义关联的简单形式 [a linguistic form which bears no partical phonetic-semantic resemblance to any other form, is a simple form or morpheme], 与后来学者做法不同的是，Bloomfield 将英语中的名词复数形式如 mice, geese 和动词过去分词形式 ran 视作一个语素，额外的语法意义视作对线性音段成分的替换操作，但是不被处理为单个的语素。因而尽管 Bloomfield (1933, p.162) 宣称一个完整的语素序列足以描写一个语言所有的语音形式 [Since every complex form is made up entirely of morphemes, a complete list of morphemes would account for all the phonetic forms of a language], 这种语素形式可以穷尽描写的是语音音段形式，而不是意义。Bloomfield 对语素的定义简明清晰，但是识别和同一的方法则没有特别明确地说明。在他之后的 Harris 和 Hockett 则将抽象的意义单位也视作语素变体，即可以在音位层面实现的语子(morph)，20 世纪下半叶美国的生成音系学研究也对这种原子化的方法起了推波助澜的作用。

Hockett (1958)后来将语素定义为最小的有意义的成分，而 Bolinger (1948)认为使用意义标准来识别语素会存在两方面的疑问：首先，“有意义”是对谁而言的(Whose meaning)? 其次，“有意义”是什么性质的意义(What kind of meaning)?第一个问题会面临由于历时词源信息的不对称导致语素义识别的社会变异问题，是以普通人语感为准，还是历史语言学家的分析为准；而第二个问题，则牵涉到现实语料中形式和意义间不是一一对应的复杂关系。

首先，语素的识别和同一依赖母语者的语言知识。如英语中的 disease 一词，从历时角度看，是由 dis-和-ease 两个成分组成，但是这两个成分在当代英语中的意义很难和 disease 的词义联系在一起，尽管使用词源信息

会有助于分析结构，但是却没有证据显示语言社团中的大部分母语者有这样的知识(Bolinger, 1948)。汉语的复合词由于结构逐渐凝固而变得成分语义和结构关系不可识别，如“麻利”、“冬烘”，寻求历时的词源信息通常会有助于识别结构(赵元任，1968；吕叔湘，1979)，但是由于大部分母语者都不具备语言的历史知识，因而无法作为共时层面鉴定语素的标准。事实上，这种信息的不对称可能会导致共时层面上“错误”的类推和重新分析，一些语言学家称之为“民俗词源”(folk etymology)，如不少英语母语者认为 godhead 和 maidenhead 的第二个成分是 head，但事实上只是历史上-hood 后缀的方言变体(Fertig, 2013)。而在汉语中，也存在一些边缘性的词汇，介于单纯词和复合词之间：

萧条	仓促	鸪鹑	螳螂
萧索	仓惶	鸺鹠	蟑螂
			蜣螂

一些联绵词的部分形式也可以同一，如“萧”(冷峻义)，“仓”(急迫义)，但是“萧条”另一半的单字“条”则意义不明。而“鸪鹑”、“螳螂”这些词的字形偏旁相同，从侧面揭示语义结构已经不再透明，接近于单纯词的语义，但是形式的同一仍然具有心理现实性。在 2009 年央视春晚小品《不差钱》中描绘了一个场景，赵本山第一次听到 dāo láng 这个词的时候，误以为是一种昆虫。

小沈阳：我重唱一个，毕老师，我再重唱一个行吗？

赵本山：唱啥啊？

小沈阳：我唱个刀郎的吧。

赵本山：拉倒吧。你唱个屎壳郎吧。

毕福剑：老哥，刀郎他不是一个昆虫的名字，他是一个人名叫刀郎，歌手的名字。

在这里，由于个人大脑中 dāo láng 没有存储为专有名词，因而作为一

一个新词，理解过程与包含“X 螂”(láng)的部分音形相同的已知词汇联系起来，在特定语境下产生误解的意义，从侧面证明了这种形式同一在识别语素成分过程中具有心理现实性。

其次，语言中形式和意义之间的复杂对应关系，表现在音段成分在词中的意义依赖于词义而定，同形成分之间无法建立起语义同一的关系。以英语为例，如 re-，在 research 中的意义是“加强”，在 recall 中的意义是“回来”，在 rewrite 中的意义是“再”，而在 religion 中没有意义，这些 re- 意义各不相同，仅仅是同形异义关系(Bolinger, 1948)。Marchand (1969)则详细分析了英语中后缀 -er 的多义性：

语义类型	举例
a. 施事	baker, commander, driver
b. 工具	boiler, eraser
c. 受事	slipper
d. 场所	counter
e. 带有地理属性的人	Londoner, Icelander
f. 与数字相关的人/事物	five, sixer, tenner

许多 -er 后缀的语义所指可以是多重的，swimmer 可以指人、鸟类和游泳器官，上表中将 -er 添加至数字后面的，可以指任何与该数字相关的事物或人。更多的例子可以见下表，在不同词中分布的一些词缀、词根没有统一的意义，语素义依赖于整体的词义(表 2-2, Lieber, 2009)：

表 2-2 词缀多义性

	in-	ex-	con-	re-	trans-	de-
-port	import	export	comport	report	transport	deport
-mit	-	-	commit	remit	transmit	demit
-ceive	-	-	conceive	receive	-	deceive
-duce	induce	-	conduce	reduce	transduce	deduce
-cede	-	exceed	concede	recede	-	-
-fer	infer	-	confer	refer	transfer	defer
-scribe	inscribe	-	conscribe	-	transcribe	describe
-gress	ingress	-	congress	regress	transgress	-
-sist	insist	-	consist	resist	-	desist

从共时角度来看，我们有两种可能的处理办法：一种处理办法是继续追求意义相同，而区别为一组同音异义的语素，即  $re_1, re_2, \dots$  或  $er_1, er_2, \dots$ ；另外一种是以形式的同一作为区分标准。而从实际操作层面讲，只有形式同一是可操作的，原因有两点，第一是意义的同一是模糊的和难以界定的，第二是这些词缀在音段序列上会有相同的音系变化表现，在构词中对词根有相同的形式限制，这种共同的形式特征，使词法学家更倾向于视为同一的成分，因而在词法学研究实践中，语言学家更倾向于处理为同一个语素成分。相应地，语素的定义标准也需要修改，先后有一些语言学家提出了自己的思路。

Bolinger(1948, p.21)认为语素是能够进入新组合(new combinations)的最小成分，将进入新组合的潜能性作为标准有两个优点：第一，可以用更加客观的分布标准代替难以把握的意义标准，即可以在新环境中复现；第二，尽管实际出现的新组合可能并不多，但正是这种进入新组合的潜能性使母语者始终将该成分视为具有意义的独立单位，这种潜能性不仅对于语素适用，也适用于词。黏着和自由是相对的，即使是分布在不同环境中的词，在这些广泛但是有限的环境中，某种意义上讲也是黏着的，而正是这种潜能性使其避免成为短语层面的黏着。汉语中存在一批双音节复合词，分布基本局限于四字格的成语结构，是一种短语层面的黏着，以单列词条

的形式收录于《现代汉语词典》中，如“国是”（共商~），“就里”（不明~），“辞色”（不假~），“灼见”（真知~），“矢口”（~否认），这些词缺乏进入新组合的潜能性。

Bolinger 看到了这种潜能性对应识别和保持语言成分的心理现实性的意义，但是缺乏这种潜能性只是意味着语言的搭配组合更倾向于被当作一个单位整体记忆和处理，并不意味着没有在形式上识别的可能性。因为人类拥有非常核心的模式识别的能力(Kurzweil, 2012)，依靠模式识别机制，我们可以识别不同表情的人脸是否为同一个人、不同画风的网络表情符号(😊 😊)、相似的语言结构和不同字体的文字符号(A ✍️ A)，一些模式识别的理论认为这些模式以形式特征的层级存储在大脑中。

Aronoff (1976, p.15)认为语素最基本的特点不在于它的意义，而是我们能够识别它，一个语素是能够连接更大的语言单位的语音序列形式[*what is essential about a morpheme: not that it mean, but rather merely that we be able to recognize it. A morpheme is a phonetic string which can be connected to a linguistic entity outside that string.*] 这种观点立场更接近与 Harris(1951)的处理方式。

此外，还有的语言学家走的更远，Hockett (1961, 1987)在后期对语素的基本单位假设进行了不遗余力的批评。Anderson (1992)在对语素的问题做了更详尽的讨论后，认为构词法的基本单位应当是词。

对汉语来说，一些传统方法通过细化义项来划分更多的语素和词缀，这种做法通常存在几个问题：

一是从根本上缺乏对不同分类方法评估选择的评价程序(evaluation procedure)，导致出现各种方案，却无法选择符合研究目标的最优方案；

二是这种原子化的分类方法试图把所有意义落实到单个语素成分上，却无法回避语义组合性(composition)假说无法完全说明结构意义的问题(Langacker, 1987)。在语义结构不透明的词汇上表现为语素义识别的困难和语素义无法说明整体词义，这是一个问题的两面。从工作程序上来讲，语素义来源于对更大语言单位的分解，上文已经论述了这种识别依赖于母语者的语言知识。这里必须区分作为历史语言学研究策略的语言知识和作为母语者共时推理的、具有心理现实性的语言知识。历史语言学家研究这些

凝固结构的历时来源，分析其内部结构，达到识别的目的，这可以是一种研究策略，然而却不足以构成共时的、具有心理现实性的理论的推理依据。因为这种做法往往忽视了一个共时的言语社团中的母语者大多数是不知道其历时来源的，相反，基于大脑的模式识别机制，母语者有时会产生一些偏离这种历时知识的“错误”分析(见上文的民俗词源举例)。因而有些历时的语言知识在一定程度上只能是语言学家的个人知识，假设母语使用者在输出这种结构不透明的语言单位时，使用了这些历时的语言知识进行组合，是不合理的。而忽视这种信息不对称，简单地拿历时的语言事实来代替共时的分析，会使语言学家落入信息不对称的陷阱(“历时信息的不对称悖论”)：离历时的语言事实越接近，离在真实推理基础上解释语言演变的目标越远。

汉语中语素义组合无法推导出整体词义，不仅表现在无法解释凝固词义多出的意义，如“洗钱”、“背书”、“码农”(程序员不等于写代码的农民)；而且还表现在一些并列结构复合词的意义需要在语境确定，单纯的语素义组合无法解释词义变少的原因，如“好歹”、“多少”(例句来自 CCL 语料库)：

1a). 你现在也不拍电影，但好歹也算有个名，人家都认识你，你就在我们这边写一个专栏吧。

1b). 您万一有个好歹，我怎么向中央交待？

1c). 人家好心好意，真不知 好歹！

1d). 但是，她总觉得，他好歹应该有一点歉意。

2a). 一曲《咱们工人有力量》激励了多少中华儿女。

2b). 那次长谈的结果，孙中山多少有些遗憾，宋蔼龄也为父亲的执拗有些不快。

2c). 每种货物应该纳多少税都明白的载于条约，那就可以省除争执。

1a 和 1d 句中的“好歹”都修饰存在的概念，但是是否有偏指好的意义，需要语境信息、社会知识和价值判定。比如“他好歹也有个博士学位”，这句话包含一个预设前提“读博是好的”，不认可读博的人是会说这样的话的。而 1d 句中则较为中性，好歹只是字面义引申，指“不管什么情

况”。1b 中的意义只出现在特定的结构，通常是“(万一)出/有个……”结构。1c 句的表达也通常限定在一些特定的场景中，通常是说话人指其他人不理解自己的好意，因而一般不会偏指歹意，在句法上的表达也相当固定，通常是“不知好歹”类似的表达。

2a 句中“多少”通常出现在一些感情强烈、正面评价的语境，是足量的肯定。2b 句中有些/有点儿后接的消极词汇更多，如“多少有些遗憾/盲目/问题”，是语用上往小量的方向说，只肯定其存在，因而偏指少。2c 句是中性的字面义引申，指数量义。

这些词的意义和用法依赖于语境，只激活特定的成分义，简单地从语素义组合的角度无法说明为什么有的成分义在语境中丢失了。汉语中还有一些并列结构的复合词已经固定地只表达一部分成分的意义，如“窗户”、“国家”等等，它们的意义演变可能和语境有关。Horn (2008)曾讨论过蕴涵意义和词汇语义的关系，认为一些词汇意义范围的扩大和缩小与蕴含量级的推理相关，这些语言现象需要词汇语用学(lexical pragmatics)的解释和说明。

总的来说，语言系统中形式和意义之间不是简单的一一对应关系，按照意义标准来界定语素，并不能有效地说明一些语言现象，反而带来分类的分歧以及形式处理的不便。此外，一个具有进入新组合的潜能性的语言成分，更容易被视为具有心理现实性的独立单位，但语言成分意义的凝固、脱落和从形式上识别、重新分析在一定程度上相互独立。因而本文采取的语素定义标准更接近于 Bolinger (1948)和 Aronoff (1976)的形式标准，即构词中的形式替换单位，同一个语素并不一定具有相同的意义，在汉语中通常以单字的形式出现，书写系统在这方面会起到一些辅助提示作用。

### 2.1.3 黏着词根复合和词复合

复合指一个语言单位由一些成分组成，而依据不同的语言，这些成分既可能是词根成分，也可以是词。英语中的“新古典复合词”就是黏着词根复合类型，而汉语中则有更多，然而成分性质的差异并不影响其复合词的地位。而汉语中词复合构词和短语之间则存在模糊的界限。界定汉语中的词和短语界限问题，是一个比较困难的问题，尤其是汉语复合词和短语

的界限。前人先后提出了“同形替代法”(陆志韦, 1937)、“扩展法”(王力, 1944; 陆志韦, 1964)、意义推断法等鉴别方法。

一般来讲, 结构松散、偏向句法性质的, 结构组合更多、能产性更高, 而结构凝固、偏向词汇性质的, 结构组合受限、能产性降低。梁源(1999)指出了在二字组结构中前后字替换的型频数量会影响二字结构的凝固程度。本文的主要研究目标不是制定区别二者的结构标准, 相反, 下文要探讨和展示的是使用能产性计算方法捕捉这种介于词法和句法之间的连续性质, 用定量的数值分布区间来表示, 因而本文没有对词表中的二字组做这方面的细致区分。通过上文中对研究范围的界定, 我们得到一个二字组的词表, 共 42106 个。

## 2.2 汉语双音节复合词能产性的计算方法

本节首先回顾文献中的能产性计量方法, 并结合汉语实际情况, 探讨复合词能产性计算的方法。接下来, 展示汉语复合词能产性计算的实例, 并结合语料分析一些理论问题。最后, 对这种计算方法的优缺点进行评析。

尽管不同学者在对能产性的度量等级上仍存在分歧(Bauer, 1992; Jackendoff, 1997), 但是如果接受能产性是一个区间的连续状态的假设, 那么就需要考虑到能产性的度量问题。Baayen 和 Lieber (1991)指出, 对能产性的度量应当满足 3 个标准: 1)这种测量应当反映语言学家对能产性的直觉; 2)可以“在统计上确定地表达一个成分进入新组合的预备性”(Bolinger 1948); 3)考虑到语义或者形式特异的词会降低能产性指数。遵循这一原则, 我们将逐个考察几种可能的计量方法。

### 2.2.1 基本计数

第一种方法就是计数带有特定词缀的词的数量, 即某词缀的词的数量越多, 该词缀就越能产。在一些语言学著作中, “XX 能产性高”的模糊用法通常是单纯计数意义上的, 即其出现的词的型频或者例频多。

这种简单的计算方法产生的结果通常会与直觉语感相悖, 即数量大可能是不能产的, 而数量小的反而能产。换言之, 其计算的数值结果无法相

互比较。如汉语中的“-子”通常认为是不能产的或接近于不能产，而“-家”(表示带有某种特征的人)在被认为是具有“新生类推潜能”(王洪君、富丽，2005)，王洪君指出在语料库中“-子”有760余例(型频)，而“-家”仅有68例(型频)。此外，这种计算方法预设了特定的词典、词表可以等同于既有词的集合，而这两者之间存在一定差异。

### 2.2.2 无法落实的概念公式

该方法由Aronoff(1976)给出， $I = V/S$  ( $I$ =能产性指数， $V$ =型频数量， $S$ =造词规则理论上可以产生的型频数量)。这里存在的问题包括：首先，如上文所指，比较不同构词模式既有的型频数量会导致反直觉的结果。其次，更为棘手的是，如何计算一个词缀理论上可以生成的潜在型频数量？Baayen和Lieber(1991, p.804)指出，随着语料库样本增大，会有越来越多的既有词进入样本中，对于不能产的构词方式来讲，这个指数会逐渐趋于稳定；而对于能产的构词方式来讲，由于理论上能产的 $S$ 可以接近无限大，所以这个指数仍然会趋近于0。所以，这个指数与其说是能产性指数，不如说是不能产指数。此外，Aronoff(1983)还注意到取词频平均值的方法，即将一个词缀的例频之和除以型频之和。但这种方法忽视了词频分布的高度扭曲的特征(齐夫律 Zipf' Law)，即词的内部词频分布高度不均衡，标准差很大。

### 2.2.3 基于孤词型频的计算方法

该计算方法由Baayen和Lieber(1991)提出， $P = n_1 / N$  (就某个特定词缀而言， $n_1$  = 孤词(hapaxes)型频之和； $N$  = 包含该词缀的所有的词的例频之和。)所谓孤词，就是在文本中只出现一次、例频为1的词。就是以词缀-er为例，假设在一个特定容量的语料库中，包含-er的所有词的例频之和为57683，包含-er的孤词型频为40，即出现了40个例频为1的包含-er词缀的词，则-er词缀的能产性指数为 $40/57683=0.0007$ 。

其基本原理可追溯至Good(1953)估测动物种群的统计方法。许多学科的研究都存在估测种类范围的需求，生物学家关心如何估测动植物种群的种类数量，即物种多样性和灭绝速度，而语言学家则可能关心估测文本和

作家的词汇量(lexical richness)。特别是生物多样性测量的经验数据采集缓慢、困难，以及环境保护政策制定等现实社会需求，推动了相关统计方法的发展。

具体到构词能产性的计算方面，其工作的基本原理如下：随着语料库样本量的增加，在增加的词中，如果出现大量新词(第一次进入语料库的词，频率为 1，会增加型频的累积)，那么整个语料库收纳的词的类型频数值增大，曲线陡峭，因而暗示了随着继续增大，仍会有更多的新词进入；如果几乎没有新词进入，那么整个语料库收纳的词的类型频数值趋于没有变化，曲线平缓，暗示随着继续增大，不会有很多新词进入。语料库扩大前后的这种数值差(增加的新词型频)，与增加曲线的斜率相关，因而在某个特定的数值点上，这种斜率可以反映曲线未来的走向，即这种数值可以间接地反映所谓能产性的潜力，见图 2-1。

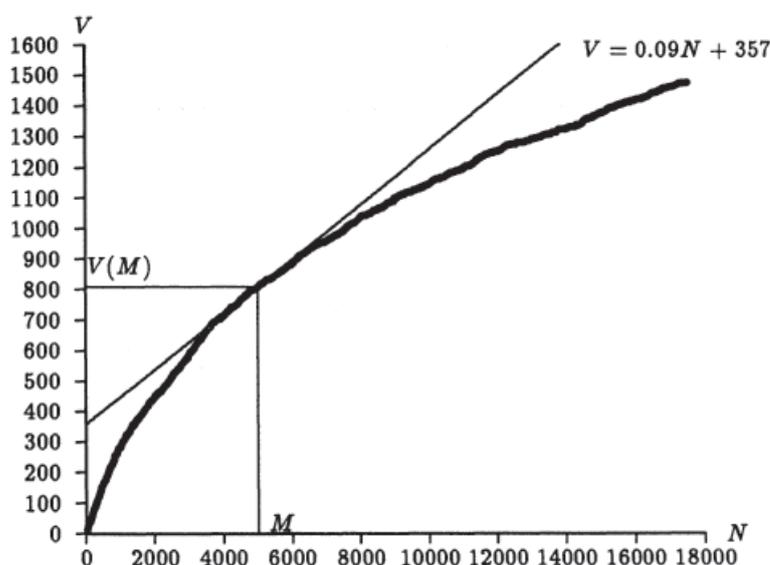


Figure 2.  $\mathcal{P}(M)$  as the slope of the tangent to the growth curve of  $V$  (the curve shown here is that of the third singular present tense flectional -s as observed for a random sample of 17,481 tokens from the Cobuild corpus; the slope of the tangent in the point  $(M=5000, V(M)=807)$ ,  $0.09$ , equals  $\frac{n_1}{N} = \mathcal{P}$  for  $N=M$ )

图 2-1 基于孤词的趋势预测 (Baayen & Lieber, 1991, p. 812)

这种计算方法通过捕捉遇到新的词型的概率来反映能产性指数，而与既有词型的数量无关，因而无法涵盖既有词型在能产性概念中的位置，因而 Baayen 和 Lieber (1991)又提出了整体能产性(global productivity)的概念，

包含特定构词模式的既有词型数量  $V$  和  $P = n1 / N$  两个维度的信息。整体能产性高的构词模式会有较高的  $V$  和  $P$  值，而整体能产性低的构词模式则  $V$  和  $P$  值都较低。这种做法的优点在于更全面地涵盖新词型和旧词型数量反映在能产性直觉中的地位，而缺点是当不同的构词模式在两个维度上数值各有高低时，则无法直观地比较出哪个模式能产性更高，如下图 2-2 中的  $-er$  和  $-ness$ 。

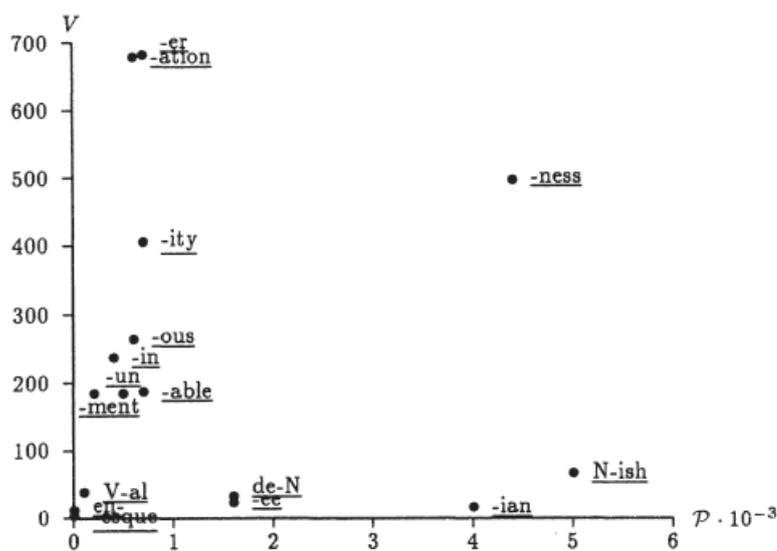


Figure 3. Global productivity  $P^*$  for various affixes (the degree of productivity  $\mathcal{P}$  is found on the horizontal ordinate; the vertical ordinate represents the extent of use  $V$ )

图 2-2 整体能产性 (Baayen & Lieber, 1991, p. 819)

### 2.2.3.1 孤词与新造词的关系

在继续讨论一些具体的语言成分的指数数值之前，我们需要厘清关于孤词的概念以及孤词是如何反映能产性潜能的。

对于一个假定的同质系统来讲，词汇的新旧可以在时间轴的序列先后上体现。对于母语者个体而言，词汇的新旧是一个相对的概念，同一个词，对于张三是新词，但对于李四则可能是旧词。对于同一个人来说，特别是从学习的角度看，第一次遇到的词是新词，但是过后可能会变为旧词，即新旧依赖于词汇知识的范围，是主观经验性的。而孤词的范围界定，是以客观的文本频率分布为基础，在 1.3 节展示的词频分布表中，频率为 1 的词中许多并不是新词，因而二者的范围有时并不一致(见下图 2-3 示意)。

从词汇监测的角度看，计算的斜率走势作为一种客观的指数，反映了

假定的经验学习的动态过程，即随着语言环境的扩大和词汇增长，结合“新”出现的词项进行的统计推断：特定的构词模式仍出现新词，那么一定还会继续出现这种模式的构词。或者可以不严格地类比为田野调查的研究过程，设想一个语言学家调查一个陌生语言，随着调查的深入，词表上的词越来越多，高频词碰到的次数多，而低频词则较少。调查语言的过程就是不断探索和扩展边界的过程，在任何一个特定范围的语料边界上，会存在一些构词模式相同、只碰到一次的“新”词(尽管对于母语者来说未必是新词)，由于语言系统提供了组合更多相似形式的空间，那么此时的语言学家可能会推断认为这个“新”词的构词模式是能产的，即在下来的调查过程中，碰到相同构词模式的“新”词的概率很大。当然，这只是从既有词型推断是否能继续观察到新形式的一种可能。

从整个语言系统的角度看，由于孤词中一部分并非主观感知的新词，而是系统中既有的旧词，因而从历时的角度看，其潜能性也有可能是反映特定构词模式的词例在一定语料容量范围内尚未被穷尽、逐渐衰退的趋势。这二者可以并存为一个语言学习者在学习一些书面文言词汇的场景，对于学习者而言是“新”词，是仍未完全穷尽的，但是对一个语言系统而言，它是逐渐退到词汇系统外围的词例。

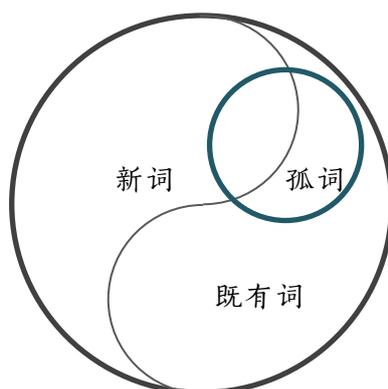


图 2-3 孤词和新词的关系

(大圆代表一个客观文本中的词汇总和，小圆代表频率为 1 的孤词总和，二者的区分标准是客观的文本频率分布。而对于母语者个体而言，新词和既有词的区别则具有一定的主观性，边界是模糊和流动的。)

当语料库样本量足够大时，孤词中仍可能包含那些对于言语社团来讲，

是新造的词。

孤词并不等于新造词，重要的是基于孤词的能产性指数计算可以暗示已有样本中的词型是否已经穷尽了可能的形式，当能产性指数大的时候，意味着还有许多词型尚未被捕捉到，当能产性指数小的时候，指几乎所有的词型已经体现在语料库中，当语料库样本量再增加的时候，几乎不会再增长。因而，基于孤词例频计算得到的数值所反映的能产性，是就这种潜能性而言的。

需要注意的是，这种测量方法只能间接反映能产性，其数值也只是基于特定的语料库样本特征基础上的数值，而不是固定反映能产性的指数。在具体的应用上，Plag (1999)指出，英语中的副词词缀-ly事实上已接近屈折成分，能产性极高，但是据Baayen和Renouf (1996)的计算结果，-ly计算的能产性指数极低，接近于不能产，因而产生了理论预测和实际语料不符的矛盾。

Baayen(1993)提出过另外一个相近的计算方法，即基于孤词的能产性程度计算(Hapax-conditioned Degree of Productivity)， $P^* = n_1/h_1$ 。其中， $n_1$ 是一个特定词缀的所有孤词型频，而 $h_1$ 则是整个语料库不同词型所有的孤词型频之和，由于同一语料库中所有的孤词型频之和相同，所以就是将特定词缀的孤词型频数值大小进行比较。由于该计算方法和上文中 $P = n_1 / N$ 的计算方法都是利用孤词的出现和分布来预测特定构词能产性的大小，因而具有原理上的一致性。在实际的语料检验中，一些词缀的能产性比单纯词的孤词型频数值还要低，与直觉相悖。在下文中计算 $P = n_1 / N$ 时也会将 $n_1$ 列出做比较。

此外，计算能产性的另一个思路是在词表频率谱的基础上估测当文本库趋近于无穷大时可能的词型数量 (Baayen, 1992)，其估测的理论基础是依据词汇分布的概率统计推断，但并非基于真实的文本，其结论仍有待检验。所以下文中将主要围绕使用Baayen和Lieber(1991)提出的公式，利用孤词型频来计算能产性，并结合汉语实例作讨论。

### 2.2.3.2 能产性计算中的位置要素

Sproat & Shih (1996, 2002)是最早利用语料库数据计算汉语复合词能产

性的研究工作，他们的计算方式是以黏着词根 X 为单位，将词根 X 分布在前字、后字位置的词语例频都列入统计，即不区分位置的能产性指数。复合词能产性指数计算是否需要区分复合词分布位置？我们认为有必要区分不同位置的计算，基于以下两点理由：

第一，词缀是定位语素，而词根则可能是定位或者不定位的，不定位的词根在复合词中的分布也可能是不对称的，其型频和例频会影响能产性指数，举例如下(表 2-3)：

表 2-3 词根的型频不对称分布

(1)前少后多

词根	型频	举例
际	前 1	际遇
	后 22	国际 交际 人际 洲际 边际 天际 星际 省际 校际 党际
间 (jiān)	前 3	间距 间架 间脑
	后 41	时间 期间 房间 中间 空间 此间 民间 瞬间 人间 世间
府	前 4	府上 府第 府邸 府绸
	后 19	政府 首府 学府 王府 官府 城府 幕府 县府 侯府 冥府
系 (xì)	前 6	系统 系列 系数 系念 系谱 系辞
	后 31	关系 联系 体系 世系 根系 直系 父系 支系 谱系 云系
术	前 4	术语 术科 术后 术式
	后 28	技术 艺术 学术 手术 武术 战术 美术 魔术 医术 算术
员	前 4	员工 员额 员外 员司
	后 95	人员 委员 成员 官员 队员 党员 演员 球员 动员 议员
况	前 1	况且
	后 18	情况 状况 病况 实况 盛况 境况 战况 路况 近况 概况
究	前 1	究竟

	后	10	研究 讲究 追究 终究 考究 探究 深究 学究 查究 推究
境	前	7	境内 境界 境外 境地 境遇 境况 境域
	后	35	环境 边境 困境 处境 入境 出境 心境 梦境 意境 家境
件	前	1	件数
	后	41	文件 案件 软件 邮件 硬件 证件 信件 稿件 配件 部件

## (2)前多后少

词根		型频	举例
全	前	81	全国 全面 全部 全球 全省 全市 全体 全身 全都 全年
	后	13	完全 安全 健全 齐全 成全 保全 俱全 周全 大全 两全
重 (chóng)	前	37	重新 重复 重申 重建 重组 重返 重演 重现 重温 重逢
	后	1	双重
投	前	43	投资 投入 投降 投票 投产 投诉 投身 投向 投标 投机
	后	4	空投 相投 远投 交投
增	前	40	增长 增加 增强 增进 增多 增产 增收 增幅 增添 增值
	后	12	新增 递增 净增 猛增 倍增 激增 剧增 大增 骤增 陡增
总	前	59	总统 总理 总结 总额 总体 总裁 总部 总数 总量 总会
	后	5	汇总 副总 一总 加总 归总
调 (tiáo)	前	39	调整 调控 调节 调解 调味 调料 调和 调配 调试 调理
	后	7	协调 空调 烹调 失调 下调 上调 微调
清	前	89	清楚 清醒 清理 清晰 清洗 清新 清单 清静 清白 清秀
	后	21	认清 弄清 澄清 凄清 冷清 查清 摸清 划清 肃清 分清
新	前	92	新闻 新鲜 新型 新年 新建 新增 新兴 新春 新人 新娘
	后	18	重新 创新 最新 更新 全新 崭新 清新 革新 刷新 维新

巨	前	36	巨大 巨额 巨人 巨头 巨型 巨资 巨响 巨石 巨变 巨幅
	后	1	艰巨

---

第二，从形式上来看，有相同的一个词根语素且分布位置相同(前字或后字位置)的一组复合词，即直接成分族(constituent family)，提供了语言习得、理解和输出的类推基础。直接成分族在类推中扮演了重要作用，是影响英语复合词重音位置的一个重要因素(Plag, 2006)。基于典型的模型仿真研究(Skousen, 1989)也使用直接成分族成员的型频数值作为挑选特征的依据。此外，在上文提到的赵本山小品，侧面反映了“X 螂”的直接成分族在新词理解中的作用。而共有一个相同词根但位置不同的复合词之间，一般来讲，无法通过形式类推连接起来，如都包含“车”词根的复合词“跑车”和“车牌”。陈保亚 (2009)根据理解和生成过程的条件不同，区分不规则语素组、解释性规则语素组和生成性规则语素组，这种结构上的相似性是区分辨析的前提条件。

直接成分族是观察词根能产性的基本单位，直接成分组成员的数量即该词根的型频，词根语素的形式同一必须保证。因而本文中能产性的计算与 Sproat & Shih (1996)相比，要分开前后字的分布位置计算，与董秀芳 (2004)列举的词法模式标准也不尽相同。具体来说，以“石”为例，“石”在前字分布的所有词例频之和为 6586，孤词型频为 6，所以“石”在前字位置的能产性指数为  $6/6586=0.000911$ ；“石”在后字分布的所有例频之和为 3961，孤词型频为 9，所以“石”在后字位置的能产性指数为  $9/3961=0.002272$ 。

## 2.3 实例分析

### 2.3.1 黏着词根的能产性

尽管基于不同语料库规模计算得到的能产性指数会不同，而且计算方法和对象的筛选方法都有一些差异，因而无法进行跨语料库的对比。Sproat 和 Shih (1996, p.54) 展示了 15 个能产黏着词根的数据，下表中展示了相同

的词根根据本文所用词表计算得到的能产性指数。

表 2-4 黏着词根的能产指数

词根	位置	型频	例频	孤词型频	能产指数	举例
石	前	66	6586	6	0.000911	石碑 石刻 石窟
	后	67	3961	9	0.002272	宝石 玉石 碎石
盒	前	4	80	2	0.025000	盒带 盒盖 盒底
	后	14	451	2	0.004435	饭盒 烟盒 纸盒
蚁	前	5	86	0	0.000000	蚁穴 蚁巢 蚁后
	后	4	108	1	0.009259	白蚁 工蚁 蝼蚁
蛙	前	3	133	0	0.000000	蛙泳 蛙鸣 蛙人
	后	3	196	0	0.000000	青蛙 牛蛙 树蛙
龟	前	5	92	1	0.010870	龟裂 龟缩 龟甲
	后	6	335	2	0.005970	乌龟 海龟 金龟
饺	前	0	0	0	0.000000	-
	后	2	45	0	0.000000	水饺 蒸饺
蝇	前	0	0	0	0.000000	-
	后	5	484	0	0.000000	苍蝇 蚊蝇 果蝇
棉	前	29	2266	4	0.001765	棉花 棉被 棉袄
	后	17	361	5	0.013850	粮棉 植棉 皮棉
菇	前	2	10	0	0.000000	菇类 菇农
	后	5	308	2	0.006494	香菇 冬菇 草菇
脑	前	18	3648	1	0.000274	脑海 脑筋 脑门
	后	16	9433	2	0.000212	电脑 首脑 头脑
木	前	66	3378	6	0.001776	木材 木板 木匠

	后	48	2665	3	0.001126	红木 伐木 苗木
驼	前	7	305	0	0.000000	驼背 驼铃 驼峰
	后	0	0	0	0.000000	-
肠	前	8	386	2	0.005181	肠癌 肠炎 肠衣
	后	17	616	1	0.001623	小肠 结肠 盲肠
蜂	前	13	375	1	0.002667	蜂巢 蜂群 蜂王
	后	9	348	2	0.005747	蜜蜂 黄蜂 雄蜂
肚	前	4	263	1	0.003802	肚皮 肚脐 肚量
	后	3	4	2	0.500000	泻肚 兜肚 下肚

如上表所示，黏着词根显示出不同程度的能产。“木”、“石”词根构词要比其他词根多出很多，但是相比之下，其能产性指数并不高，原因是既有词型多表示系统提供的构造形式已经被占用了一部分，因而出现更多词型的潜能性反而会降低。能产性指数显示“蛙”、“饺”、“蝇”、“驼”在前后字位置都不能产，由于在语料库范围内不再出现相关词根的孤词词型，因而推断在更大的范围内，可能不再出现。但是“蛙”、“蝇”是否会出现更多的生物种类和命名，应该是具有构词潜能性的。按照 Sproat 和 Shih (1996)不区分词长和构词类型的统计方法，这几个词根语素是能产的。但是，Sproat 和 Shih (1996)的计算结果显示“蟑”的能产性指数与“肚”接近，语感直觉却是不能产，他们给出的理由是数据点过少而不显著。而如果按照本文区分词长、构词类型、分布位置的统计方法，会有大量的词根语素在特定位置上的分布点稀疏，因而产生计算结果是否可信的问题，如“肚”在后字位置上的指数是 0.5。

### 2.3.1.1 数据稀疏和词根能产性指数的区间分布

在整个双音节复合词表中，同一词根的孤词型频浮动区间在 0 到 22 之间：在前字位置的孤词型频数量在 0 到 15 之间，在后字位置的孤词型频数量在 0 到 22 之间。各词根在前后字位置按照孤词型频数量划分的区间

分布如下：

表 2-5 孤词型频的区间分布

孤词型 频	词根数量	
	前	后
0	2866	3004
1	964	829
2	374	339
3	170	183
4	98	72
5	32	65
6	32	40
7	20	20
8	5	10
9	10	10
10	5	4
11	6	3
12	1	1
13	3	-
14	-	3
15	1	2
16	-	-
17	-	-
18	-	-
19	-	-
20	-	1
21	-	-
22	-	1
合计	4587	

与之相关的是，由于词根之间孤词型频的差异不大，而差异主要体现

在例频之和，范围从 0 到十多万，所以在横向比较不同词根的能产性指数大小时，能产性指数的数值大小和例频之和的数值大小呈负相关关系：分母越大，则数值越低；分母越小，则数值越大。下表中选取词根在前后字位置的例频之和，按照区间分段展示，可以见到每个区间段内能产性指数的大小随着在递增：

表 2-6 能产性指数的区间分布

例频	前		后	
	词根数量	能产性指数(最大)	词根数量	能产性指数(最大)
$N \geq 100000$	10	0.000095	9	0.000067
$10000 \leq N < 100000$	357	0.001055	364	0.001076
$1000 \leq N < 10000$	1017	0.004966	946	0.007463
$100 \leq N < 1000$	1341	0.033333	1260	0.044643
$10 \leq N < 100$	801	0.400000	701	0.230769
$1 \leq N < 10$	409	1	303	1
$N = 0$	652	0	1004	0

按照 Baayen 和 Lieber (1991)的预测，基于孤词的能产性指数计算应当能够反映不同词根的能产性程度，并可以在一个给定的语料库范围内相互比较。上表反映出来的趋势在一定程度上反映了既有词型多的词根能产性程度会降低。但是，一些分布稀疏的数据却会出现无法准确反映走势的问题，集中地表现在只有一个点的数据，“蠹-”、“-蔓”的能产性指数则为 1。

表 2-7 能产性指数频率分布举例

(1)前字位置

词根	位置	型频	例频	孤词型频	能产指数	举例
大	前	269	137291	13	0.000095	大会 大量 大师
土	前	69	12322	13	0.001055	土地 土壤 土匪
余	前	38	1611	8	0.004966	余额 余粮 余热
吐	前	11	120	4	0.033333	吐露 吐字 吐丝
卯	前	5	10	4	0.400000	卯年 卯月 卯日
蠹	前	1	1	1	1	蠹虫
虑	前	0	0	0	0	-

(2)后字位置

词根	位置	型频	例频	孤词型频	能产指数	举例
人	后	229	133532	9	0.000067	女人 男人 老人
线	后	107	13017	14	0.001076	路线 战线 航线
虫	后	41	1072	8	0.007463	臭虫 害虫 蛀虫
梨	后	10	112	5	0.044643	雪梨 酥梨 黄梨
焊	后	5	13	3	0.230769	电焊 点焊 脱焊
蔓(wàn)	后	1	1	1	1	瓜蔓
会(kuài)	后	0	0	0	0	-

上表显示的稀疏分布导致的高能产性是无法令人信服的,如那些只有 1

个词型的词根分布，由于例频也为 1，而得到 1 的能产性指数。换言之，这种计算依赖于密集连续分布的词型，由稀疏数据得到的计算结果与语感直觉相悖，因而是无效的。

目前关于构词能产性研究的文献大部分都是计算少数能产的词缀，而没有关注更多的稀疏数据，这些数据能产性如何度量的研究相对不足。Sproat 和 Shih (1996)计算的汉语黏着词根共 20 个，属于抽取个例的计算，没有对话料库中的词根数值进行大量对比，当数据点数过少时，会出现不能产的黏着词根“蟑”指数和“肚”接近的情况。由于这种计算方法在一些特定的词汇分布环境下不再适用，因而 Sproat 和 Shih (1996)反驳 Dai (1992)，认为汉语中有的黏着语素是能产的，不应将黏着词根复合和词复合两种机制分开，这里的论证方式是有问题的，尽管含有黏着语素的复合词能产可以被历时的证据证实，如近年来黏着语素“裸-”相关的新词“裸官”、“裸婚”、“裸考”。

### 2.3.2 定位与能产性的关系

Bolinger (1948) 认为自由和黏着是相对的，自由也是在有限环境下的“黏着”。马庆株 (1995) 认为需要区分“相对定位”和“绝对定位”，需要考虑句法和词法不同的平面。“定位/不定位”，在一定程度上也可以转化为数量特征，即在限定的语料中，在特定环境和位置上的频率是否为 0，以及是否在不同环境和位置上的频率都不为 0。

首先，汉语定位的词根语素既可以能产，也可以不能产。

表 2-8 定位词根语素的能产性指数

(1)前字定位

词根	位置	型频	例频	孤词型频	能产性指数	举例
各	前	46	47992	6	0.000125	各方 各界 各类
	后	0	0	0	0.000000	-
咨	前	2	2097	0	0.000000	咨询 咨文
	后	0	0	0	0.000000	-

(2) 后字定位

词根	位置	型频	例频	孤词型频	能产性指数	举例
龄	前	0	0	0	0.000000	-
	后	22	3837	3	0.000782	船龄 楼龄 树龄
虑	前	0	0	0	0.000000	-
	后	9	8041	0	0.000000	考虑 忧虑 焦虑

其次，汉语词根语素既可以定位能产，也可以不定位能产：

表 2-9 词根语素能产性指数的对称和不对称分布

(1) 前后字位置均能产

词根	位置	型频	例频	孤词型频	能产性指数	举例
国	前	90	112684	9	0.000080	国民 国会 国防
	后	64	80825	3	0.000037	贵国 异国 岛国
文	前	75	61563	8	0.000130	文书 文本 文稿
	后	83	9385	7	0.000746	论文 诗文 散文
棒	前	5	173	2	0.011561	棒球 棒槌 棒针
	后	13	222	2	0.009009	棍棒 木棒 冰棒

(2) 前字能产/后字不能产

词根	位置	型频	例频	孤词型频	能产性指数	举例
重(chóng)	前	37	11641	2	0.000172	重新 重复 重申
	后	1	558	0	0.000000	双重
联	前	49	23298	10	0.000429	联网 联手 联欢
	后	17	1453	0	0.000000	关联 并联 串联

巨	前	36	9097	5	0.000550	巨大 巨额 巨人
	后	1	700	0	0.000000	艰巨

(3) 前字不能产/后字能产

词根	位置	型频	例频	孤词型频	能产性指数	举例
社	前	17	45599	0	0.000000	社会 社区 社长
	后	18	2482	1	0.000403	报社 诗社 剧社
部	前	19	46267	0	0.000000	部门 部队 部分
	后	64	48746	7	0.000144	外部 局部 尾部
员	前	4	5214	0	0.000000	员工 员额 员司
	后	95	47386	22	0.000464	委员 成员 海员

根据数据统计，词根在前、后字位置能产性指数是否为 0 可以分为 4 类：a.在前、后字位置能产性指数均为 0 的词根占词根语素总数的 46.2%；b.只在前字位置能产性指数为 0 的词根占词根语素总数的 16.3%；c.只在后字位置能产性指数为 0 的词根占词根语素总数的 19.3%；d.在前、后字位置能产性指数都不为 0 的词根占词根语素总数的 18.2%。

尽管直接成分族的类推行为通常导致词根语素在单个位置的分布增长，但以上的分布比例说明词根语素的能产并不是简单地局限在单个位置的分布上，而是存在多种类型，既存在单个位置的能产，也存在多个位置的能产和不能产。因为同一个位置的分布增长，潜在地可能是两个位置上的某一个单字类推导致的，或者说一个新词的产生，也可能会激发前后位置上的特定一个或两个词根语素的类推，形成链状结构。如与新词“虎妈”(Tiger Mom, 指用苛刻方法教育子女的母亲)相对的“虎爸”(指用苛刻方法教育子女的父亲)、“狼爸”(指用苛刻方法教育子女的父亲)、“猫爸”(指用宽容民主方法教育子女的父亲)等词汇，“狼爸”是基于“X 爸”的平行关系，与“狼孩”(指脱离人类社会、被狼抚养长大的人类儿童)的语义关系不平行。



同一直接成分族的词汇间语义关系不平行的一个重要原因，在于类推的词根位置和方向有时是不确定的，因而当新词占据特定的形式位置时，就可能同其他词汇在另一个词根位置方向上关系不平行，如“狼爸”和“狼孩”，这是语言形式和意义无法整齐对应的一个表现，也是语言系统形式填充具有动态性的一个表现(下文还会继续探讨语言形式增长的动态方式与频率的关系)。因而同一直接成分族的成员之间不一定具有词源上的类推关系，而只是语言系统形式构造既有的排列表现。

### 2.3.3 汉语词缀的能产性分析

表 2-10 汉语词缀的能产性指数

词缀	例频	型频	孤词型频	能产性指数	举例
阿-	2736	16	1	0.000365	阿爸 阿弟 阿姐
老-	29271	35	3	0.000102	老师 老板 老公
-子	136835	463	26	0.000190	日子 脑子 桌子
-头	13805	61	3	0.000217	石头 念头 骨头
-们	258603	10	0	0.000000	我们 他们 你们
-巴	2599	12	0	0.000000	嘴巴 尾巴 泥巴
-么	155823	7	0	0.000000	什么 怎么 多么
-乎	17562	17	0	0.000000	似乎 几乎 在乎

如上节所讨论，从语料库增容的角度看，上表的能产性指数高于 0 的反映了未来继续增加词型的潜能性，而等于 0 则反映了特定的构词模式可能已经封闭、不再有新的词型进入的可能性。从能产性指数上来看，无法通过数量特征来区分词根和词缀，二者都有能产性指数为 0 的例子，同时都有在单个位置上能产的例子。

而从语言系统的角度看，“阿-”是汉语方言中称呼亲属关系和人名的常见格式，因而是能产的。“老-”在汉语中仍然在产生一些新词，主要用于称呼人，标识一些属性特征，如“老外”（改革开放以后）、“老美”（指美国人）、“老赖”（指欠了别人钱，虽然有偿还债务能力，却拖欠不还的人）。值得注意的是，汉语中似乎尚未出现使用“老英”、“老法”、“老德”来指称英国人、法国人和德国人的例子。

需要指出的是上表中没有计算词缀“-儿”的能产性，主要原因是一些儿化的词在书写形式上不一定会体现出来，依赖于书写者的书写习惯，因而数据可能会产生大量偏差，所以没有列入计算项目。Nishimoto (2003) 对汉语词缀提出新的能产性计算方法；Arcodia 和 Bascino (2012) 利用台湾中研院近代汉语和现代汉语语料库，对“-儿”、“-头”、“-化”的历时变化作了对比，然而必须指出的是，这几位学者忽视了书写习惯对数据采集造成的偏差，因而基于书面文本频率计算的现代汉语“-儿”词缀的指数是不可靠的。

“-子”和“-头”的能产性指数显示仍存有新的词型还未被捕捉到，随着语料库进一步增大，可能会有新的形式进入，尽管这些新的形式不一定是“新造词”。因而从历时的角度来看，此时这种指数表现的构词模式“能产”只能被解释为在当前语料库容量下词型尚未被穷尽，因为一些构词模式不能产并不意味着从语言系统中完全消失，而是会逐渐衰退到更边缘的位置。一些构词模式衰退的速度可能要比单独的新词在社团中扩散要慢得多，因为它们仍可以通过大脑记忆和社会教育的方式进行母语者间的代际传递。换言之，在这些词缀的能产性计算上，基于孤词的计算方法是通过孤词来估测语料库容的词型增长潜能，但是无法区别是新兴产生还是衰退。

### 2.3.3.1 能产性指数和历时视角下的衰退

与之相关的一个问题是，这种衰退的、没有被穷尽的趋势可以达到多少？通常认为语言符号具有任意性，语言符号的创造相互独立的。由于单纯词的创造不是基于结构模式的构造，因而其产生提供了评估能产性的基准线，介于能产构词模式和不能产模式之间(Baayen & Lieber, 1991)。汉语系统中双音节单纯词的类型和来源包括古汉语遗留的联绵词、从其他语言中翻译借入的借词，以及一些逐渐凝固、词源信息丢失的词汇。由于借词涉及语言接触和语言系统的外部因素，因而通常无法准确预测其变化趋势，但联绵词提供了观察在现代汉语中已经不能产的构词模式衰退的趋势：

表 2-11 联绵词的能产性指数

例频	型频	孤词型频	能产性指数	举例
16534	131	9	0.000544	仿佛 蹉跎 踟蹰 徘徊 缥缈 犹豫 趑趄 狰狞

联绵词的能产性指数要高于上表中词缀的能产性指数，反映其在语料库中尚未穷尽的潜能更高，但语感告诉我们这并不意味着联绵词更能产，相反，那些数值更低的构词模式反映的潜能性可能才是有关“新词”的进入，如“阿-”、“老-”。

总的来说，基于孤词的能产性计算反映的是随着语料库增加尚未穷尽的趋势，由于这种统计并未涉及到母语者复杂的心理类推和学习机制，进入的“新”词可能只是历时上在衰退的构词模式的产物，对于学习个体的“新”并不等同于能产，即计算数值和语感直觉之间仍存在差距。

对于汉语来说，这从侧面印证了汉语书面语系统的稳定和持续性。词汇系统内部不同性质的词汇在变化速度方面的差异较大。语言和现实世界保持着复杂的镜像关系，一些新词汇的扩散与社会文化生活的流行趋势有时存在同步性。《快乐男声》是 2007 年起湖南卫视举办的歌手选秀赛的节目，参赛选手被称为“快男”；《奔跑吧，兄弟》是 2014 年浙江卫视开播的户外真人秀节目，又名《跑男》。“快男”和“跑男”两个词的兴衰走势与其代表的电视节目的播放有密不可分的联系，其中“快男”反映的峰值

是 2013 快乐男声的选秀比赛开播：



图 2-4 “快男”和“跑男” (<http://index.baidu.com/>, 检索日期 2015-5-8)

一些网络词汇的兴起和消退像病毒一样扩散和消退，如“给力”：

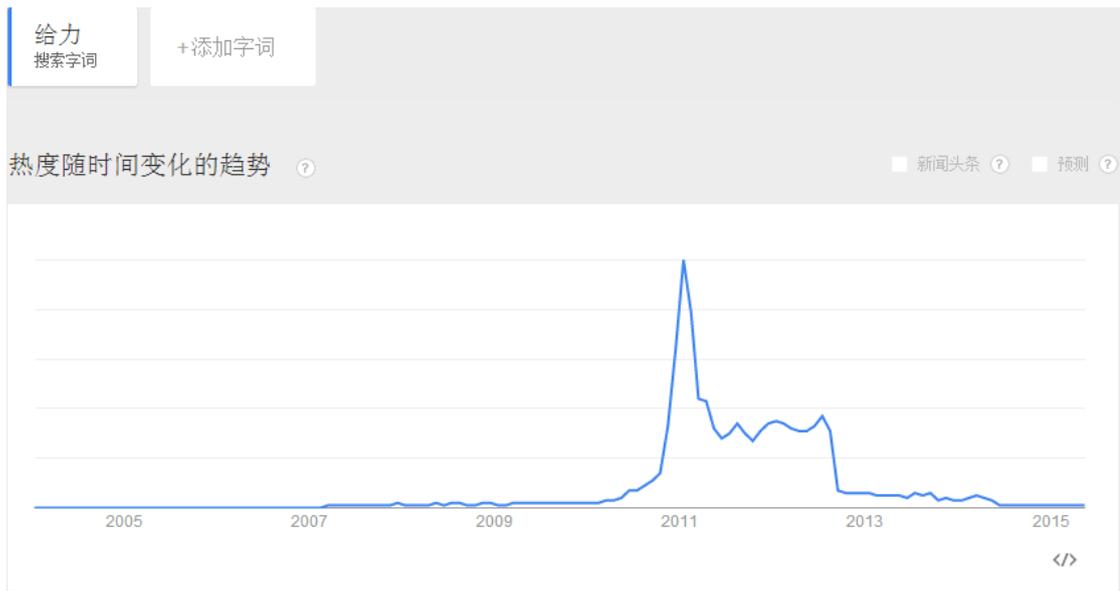


图 2-5 “给力”的变化走势 (<http://www.google.com/trends/?hl=zh-CN>, 检索日期 2015-5-08)

相比之下，一些书面语词汇的变化更像一种胶着状态的稳定和衰退。19 世纪的历史语言学家将语言演变比附为有机生命体的生长与死亡，这种隐喻用法在今天的历史语言学文献中仍可以见到。然而，从连绵词衰退趋势来看，这些古代汉语的遗留成分退出词汇系统核心的速度极为缓慢，并

且借助记忆和社会教育，可以通过书面语形式在代际间传递，它们的“生命”和变化与有机体的性质存在很大差异。如何用定量的方法记录和分析研究词汇的产生、传播和消亡，揭示词汇在动态系统中的全貌，仍存在许多未知。从理论上讲，这些不再能产的构词模式的词型可以在更大规模的语料库中被穷尽，但是基于孤词的计算指数无法在特定的库容下将这种衰退的构词模式与其他能产的构词模式区分开来。

### 2.3.4 词库和句法之间：“1+1”和“2+1”格式能产性对比

Bolinger (1948)指出话语片段(a + b) + (c + d)的新组合比单独的 a + b 或者 c + d 组合更自由，即小单元的语言环境比大单元的要更容易预测。换言之，从词汇到句子，结构组合的自由度和复杂度同时增加。相比较而言，词法结构的组合更加紧凑，能产性较低，而句法结构的组合更加松散，能产性更高。我们对汉语词根 X 的“1+X”和“2+X”两种格式分开统计，计算能产性指数 p。

表 2-12 “1+X”和“2+X”格式对比

语素	例频	型频	孤词型频	能产性指数	格式	举例
-者	70762	60	14	0.000198	1+X	记者 死者 弱者 读者
	19906	293	115	0.005777	2+X	消费者 劳动者 推动者
-人	133532	229	9	0.000067	1+X	古人 伟人 诗人 线人
	37602	232	38	0.001011	2+X	投保人 稻草人 领路人
-家	109965	95	3	0.000027	1+X	作家 画家 行家 赢家
	12369	91	10	0.000808	2+X	科学家 艺术家 空想家
-手	35577	139	6	0.000169	1+X	选手 对手 凶手 推手
	747	23	6	0.008032	2+X	狙击手 二传手 甩花手
-机	22503	88	10	0.000444	1+X	飞机 手机 坠机 停机
	11001	203	39	0.003545	2+X	计算机 交换机 打浆机
-战	15597	59	2	0.000128	1+X	内战 激战 参战 海战

	1188	50	7	0.005892	2+X	阵地战	价格战	贸易战	
-品	39842	63	1	0.000025	1+X	商品	食品	祭品	毒品
	3729	88	9	0.002414	2+X	艺术品	纺织品	纪念品	

上表显示，“2+X”的指数要比“1+X”的指数高出很多，即词根“2+X”格式能产性要比“1+X”格式能产性高，计算结果印证了这样的语感直觉，基于孤词的能产性计算通过数值反映了从词库到句法的渐变区间。这种能产性程度的差异不是通过既有词型频的差异来体现的，因为对不同的词根而言，“1+X”和“2+X”型频数量都有可能是更多的一类，如“-手”、“-战”“2+X”的型频就比“1+X”低，因而单纯依据既有词的型频和例频无法说明为什么“2+X”格式结合松散，能产性更高，而只有通过基于孤词的频率计算，才能将这种潜能性表示出来。

如果我们将词的例频划分为A组( $r > 100$ )，B组( $10 < r \leq 100$ )，C组( $2 \leq r \leq 10$ )，D组( $r = 1$ )四个区间，按照词的例频数值和不同格式进行分组对比，每一组的项目是某一特定格式在例频区间内的型频数量所占型频总和的比例。我们可以观察到“1+X”和“2+X”两种格式的构词模式的词频分布趋势：

以一个特定的构词格式看其在A，B，C，D四个区间的频率分布，“1+X”格式分布比例最高的一般在A或B组的高频区间，D组的比例要比A组低；而“2+X”格式分布比例最高的一般要后移至B，C，D区间。比较在同一个区间内的词根相同、音节数量不同的构词格式，在A组项，“1+X”格式的比例数值要比“2+X”高；而在D组项，“2+X”格式的比例数值要比“1+X”高。用一个比较形象的说法就是，如果将图表横置，“1+X”的分布图更像一个从波峰衰退的走势，而“2+X”则像是“1+X”的波峰后移的分布，因而在D组区间上的分布更多，即如果有大量的例频为1的词出现，占据型频相当一部分比例，通常意味着结构比较能产。总的来说，不同结构的能产性不同也反映在型频的数量区间分布上。

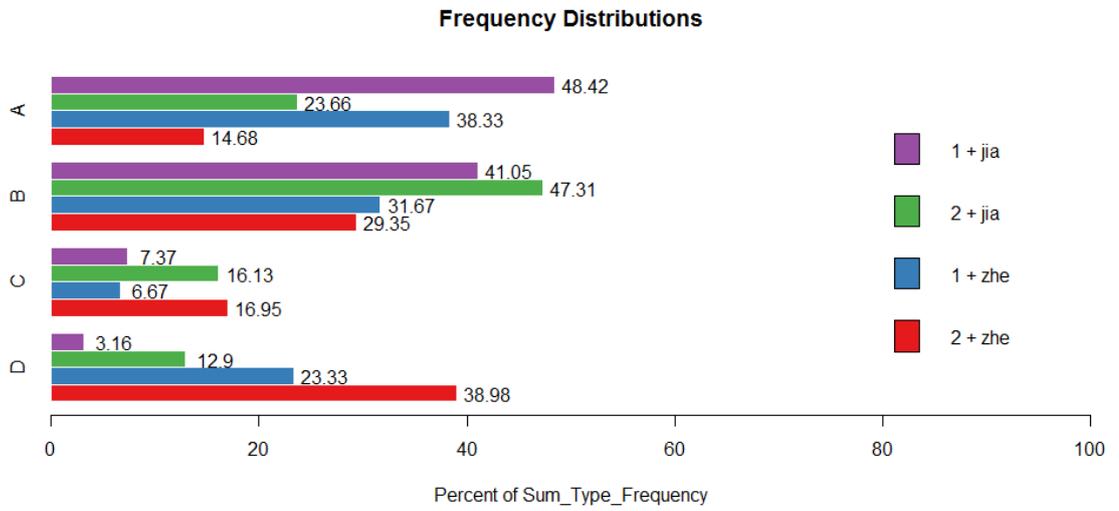


图 2-6 “1+家/2+家” 和 “1+者/2+者”

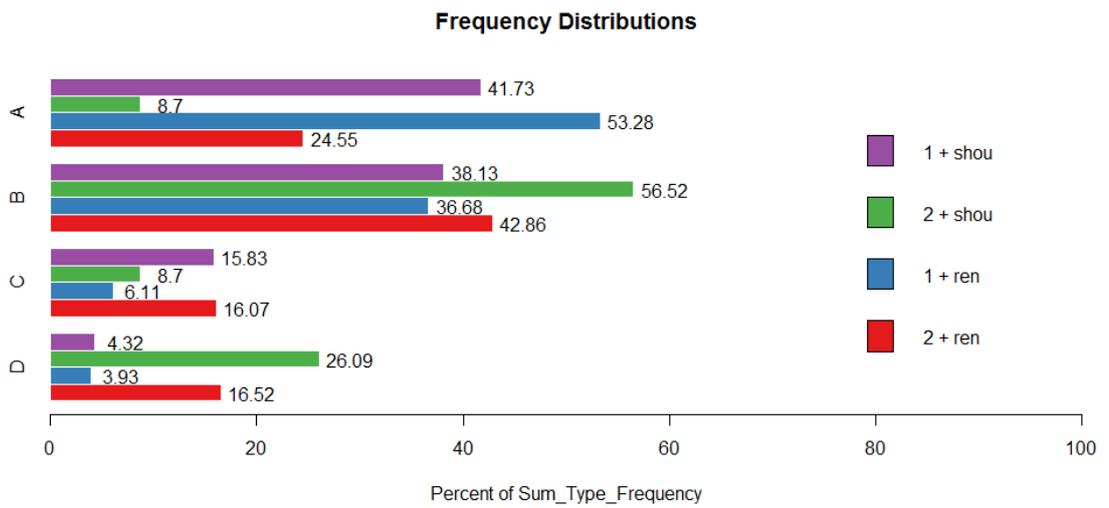


图 2-7 “1+手/2+手” 和 “1+者/2+者”

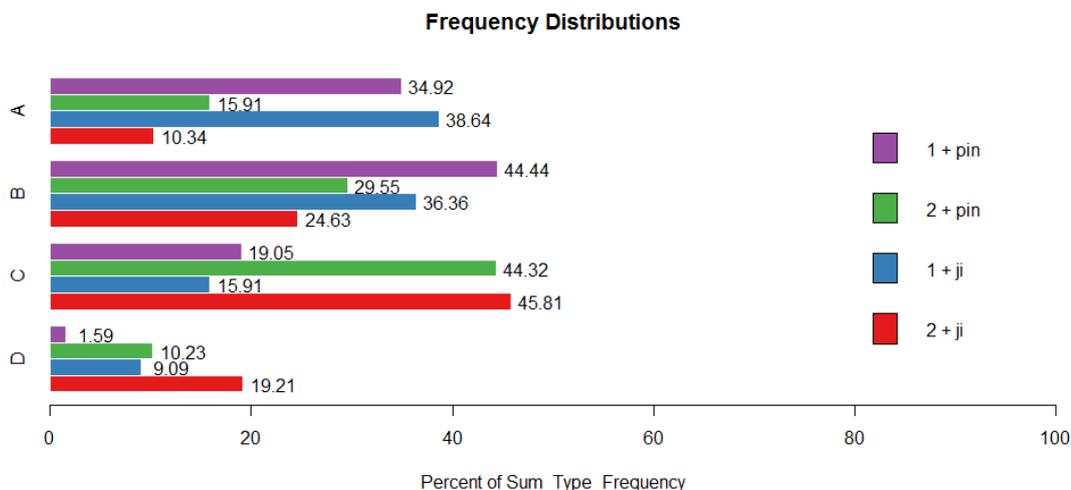


图 2-8 “1+品/2+品”和“1+机/2+机”

为什么基于孤词的能产性计算方法能够反映这种随着结构复杂度的增加，而增加能产性的渐变现象？一种可能的解释是 2+1 格式备选的双音节词汇比单音节词汇更多，因而可以组成更多的新词，但问题在于这种概率组合上的可能性与 1+1 格式无法简单地对比，得出哪个更能产，因为有的 2+1 既有组合也更多。第三章会引入相对频率分析组合性和频率的关系，分析新组合出现和频率分布间的关系。

### 小结

总的来说，基于孤词例频计算构词能产性在应用于汉语复合构词时，在特定的语料库容量下利用例频分布信息来推断新词进入语料库新词的潜能性，同时也可以大致反映出随着结构复杂度增加而能产性增加的趋势。在实际计算中也暴露出了一些问题：

a. 由于孤词本身并不涉及复杂的学习推理过程，因而“新词”的进入是基于学习者的个体视角和词汇信息判定，和从语言系统层面区别的词汇新旧存在交叉和出入，这种推理只是一种间接的统计反映；

b. 基于孤词例频的计算依赖于数据密集连续分布拟合的曲线走势。以往的研究集中在少数能产的词缀，而没有在给定语料库范围内对不同成分大规模对比。具体在汉语双音节复合词能产性的计算上，由于孤词数量增

长缓慢，能产性指数数值的大小在很大程度上取决于例频之和的数值大小，有时计算得到的数值会与语感直觉相背离。词汇统计学提出型例关系 (type-token relation) 来测量文本的词汇丰富程度，反映文本中的信息流动。这种方法也同样源于统计生物种群数量和丰富程度。汉语词汇的分布特征和在语料文本中的型例关系迄今为止了解地还不够，限于材料限制，上文中没有结合孤词的稀疏分布进行探讨。

本文中的能产性计算主要基于例频。梁源(1999)在研究二字短语凝固度中发现了两种有效的相关因素：(1)前后字替换频率，在定中类、状中类和述宾类短语中，呈阶突变性关联；(2)前后扩展的歧义度。其中前后字替换频率的观察是基于型频，而与二字组整体的例频计数无关。汉语双音节复合词的型频与例频在词汇心理加工过程中分别起什么样的作用？型频有效的原因是否在于这些词汇在词库中以个体形式离散地分布？这些仍是开放的问题，在第三章会部分地涉及心理词库的组织问题。

### 第三章 相对频率和心理词库

能产性与结构的分解和组合相关，结构的组合透明度会影响其能产程度。本章首先在 3.1 节从理论上讨论了复合词心理加工处理的可能方式，接下来在 3.2 节引入相对频率讨论在字组层面观察结构成分的自由和黏着性质的变化，及其对分析能产性的影响和作用。

#### 3.1 存储和加工的效率平衡

通常来讲，单纯词不具有内部结构，即意义和形式间是任意的关系，如“猫 māo”，指行动敏捷、擅长捕鼠的哺乳动物，由于没有内部结构，其意义无法从次级的语音形式推断得到，因而必须记忆。而复合词由于包含内部结构，则需要考虑两方面的特点，一方面是在言语理解和加工过程中，复合词可以被分解，其直接成分可以被识别，否则母语者无法理解在日常生活中遇到的新复合词。另一方面是结构不透明、语义具有特异性的复合词必须作为整体存储在记忆中，在输出时作为整个词项单位可以被检索和提取。因而复合词的语言加工方式和词库表征方式在理论上仍存在许多疑点和问题。

从理论上讲，复合词在心理词库中的表征形式有三种可能性(Libben, 2006):

- a). 存储效率最大化(Maximization of Storage Efficiency)/冗余形式最小化;
- b). 计算效率最大化(Maximization of Computational Efficiency);
- c). 混合概率最大化(Maximization of Computational & Storage Opportunity)。

所谓存储效率最大化，即尽可能去除重复的直接成分语素，复合词尽量被分解为单个成分存储，而不是作为整体词项被存储。这种将记忆负担最小化的做法，付出的代价是在言语理解过程中复合词都必须分解为成分序列并解释，相对地，由于词汇不作为整体单位从记忆中提取，所以在言

语输出过程中也必须使用规则，将语素序列组织起来，而规则的增加必然带来语言加工速度和效率的降低。

计算效率最大化则与之相反，则尽量将各种不同大小的成分存储，既有复合词，同时也包含直接成分语素。当遇到新词时，倾向于不分解为直接成分的形式，在言语理解过程中尽可能将词库中的词汇形式进行匹配，而在语言输出中也尽可能提取词库中现成的独立形式。

这种追求极致效率的操作思想大多带有上个世纪模仿计算机程序的时代烙印，着重评价计算的简洁性而忽视了大脑的工作机制。语言经验证据显示大脑的词汇系统并不是这两种极端方式中的一种，而更像是混合的工作机制。一个使用频率高的复合词被整体存储和表征，并不一定会阻断大脑对该词结构的分解机制。对于大脑这一机制是如何演化而来，目前研究的还不够充分。但是从功能的角度来看，无论是哪种方式提供语言输出，在面对外部的语言环境出现的各种词汇时，大脑都必须在很短的反映时间内完成即时信息处理(real-time processing)的任务。

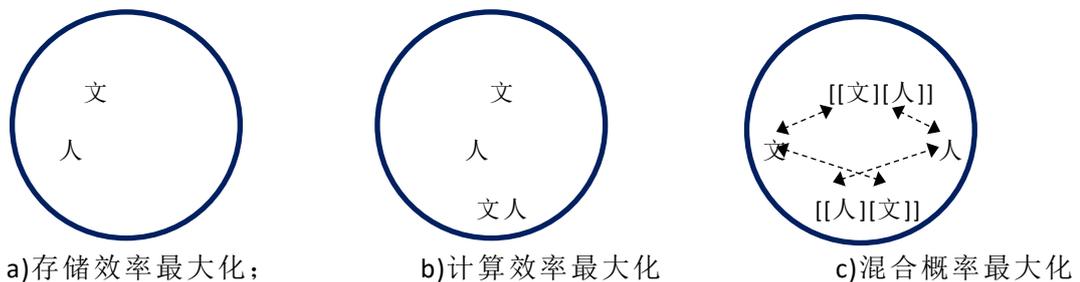


图 3-1 词库的复杂度与效率

事实上，理论的简洁性并不是评价理论最重要的指标，而成分冗余对维护语言系统的稳定和允许词汇层面的变异都是必要的。存有大量冗余的语言系统允许以词为单位，离散地发生变化，可以合理的解释一些词发生变异和另外一些词不变同时并存的现象，而使用组合规则推导的结果只能是要变一起变。如果复合词在词库中都以直接成分语素的形式存在，不存在冗余，那么复合词的输出应当以语素序列的形式组织，规则的施用应当保证语素在组合前后的同一性，换言之，语素 A 在组合输出的语言单位  $AX_1$ ,  $AX_2$ ,  $AX_3$  中应当保持形式和意义的同一。然而，在现代汉语中的确存在与上述预测有出入的情况。如“强 qiǎng(勉强，强迫义)”在“勉强”

中读 qiǎng，而许多人则将“强迫症”一词中的“强”读作 qiáng，而表达勉强、强迫义的“强”是同一的。这里的问题不在于一些语言使用者读音是否规范，而在于语言使用者的大脑没有遵循这种规则推导的一致性，从侧面揭示了至少一些词在大脑中应当以词的形式表征，而不是语素。类似的还有“卡 qiǎ(夹住，阻挡义)”在“卡脖子”中读 qiǎ，而在“卡位(篮球运动中，进攻人运用脚步动作把防守者挡住自己身后)”、“卡指标”中则读 kǎ。对于这些多音字读音出现的分布差异，或许还有更进一步的解释，但是这种读音的差异是以离散的词为单位分布，缺乏使用规则组合的一致性，因而整词表征形式更符合这些语言现象。

混合的工作模式将 a)或 b)任何一种作为默认处理机制，都会带来一些特定的问题。如果将 a)模式作为默认方式，则意味着所有的复合词在理解过程中都必须被分解为直接成分。语言使用者理解新词，第一次分解成分有助于理解新词义，而每次都分解则效率太低，语言系统中的高频常用词也类似。如果将 b)模式作为默认方式，则意味着所有的复合词在理解过程中都优先被整体存储而不是分解，这样就无法理解日常生活中临时语境下构造的新词。因而，大脑的工作方式必须能灵活应对外部环境的各种语言输入信息。相应地，一些学者先后提出了内部具有竞争关系的双通道模型(Frauenfelder & Schreuder, 1992; Baayen & Schreuder, 1999)。然而，大脑是如何选择不同的处理方式呢？或者说什么时候分解机制会触发？

### 3.2 相对频率

最近的派生词心理加工的实验显示，词基及其派生词的相对频率是影响词汇加工的重要因素，语言证据包括法语(Cole, Segui, & Taft, 1997)、英语(Hay, 2001, 2003)、意大利语(Burani & Thornton, 2003)、塔加洛语(Zuraw, 2009)、俄语(Antic, 2012)。

以英语为例，派生词 approach 和 dislike 整词的例频就是派生词频，而词基频率指的是没有添加词缀的词频，由于 -proach 是黏着词根，不能独立成词，因而 approach 的词基频率为 0，而 dislike 的词基部分 like 可以独立成词，所以其词基频率就是 like 的例频。而在汉语中，相对频率也可以

体现为一个自由成词的成分 A 和一个与之相关的 AX 复杂结构形式之间的频率关系。

Hay (2001, 2003) 认为词或词缀单独的绝对例频并不重要,重要的是词基和派生词的相对频率会影响到词汇加工机制。当词基比派生词例频更高时,如 *uncommon*(*common* 比 *uncommon* 词频高得多),该派生词倾向于分解处理;而当派生词比词基例频更高时,如 *impatient*(*impatient* 比 *patient* 词频更高),该派生词更倾向于整词处理。词汇的加工模式与词缀的能产性相关,倾向于分解处理的词,即词基比派生词例频更高,其词缀能产性更高,而倾向于整词处理的词,即派生词比词基例频更高,其词缀能产性较低。梁源(1999)发现二字组中前后字替换与结构凝固程度相关,与相对频率的思想是一致的。Yang (2013)利用中国国家语委语料库抽取 AN 复合结构,即形容词与名词的复合结构,来研究汉语复合的能产性并检验相对频率的假说。他在统计复合词右成分 N 的频率时,将 AN 和 NN 的右中心成分频率加在一起,作为词基频率(base frequency)。由于没有有效区分成分 N 在 AN 结构中和 AN 结构外的自由分布,因而他的统计方式是有问题的。相应地,他提出的相对频率在汉语复合加工中无效的说法也难以令人信服。

对相对频率做结构观察和研究可以追溯到 Thorndike (1943)提出的派生比率(derivation ratios),即按每个词型分别计算,将派生词例频除以词基例频,如 *proudly* 出现 101 次,*proud* 出现 396 次,则得到派生比率为 0.25。尽管这种派生比率计算与能产性并没有直接关联,但是却提供了观察词基组合复杂派生形式的频率分布的角度(Baayen & Lieber, 1991)。Harwood 和 Wright (1956)较早注意到通常派生词的频率要比词基的频率要低,因而派生词中的低频词分布要比词基频率分布更多的事实,他们认为这种分布显示了语言使用者使用记忆的词基形式和构词规则知识构造更多新词的创造机制,即构词规则的生成性。

在 2.3.4 小节,我们曾留下一个问题,为什么基于孤词例频的计算方法能区分这些结构间不同的能产性,反映出词法到句法的渐变?值得注意的是,两种结构的频率区间对比同样反映了与词基例频和派生词频率对比趋势相类似的分布,尽管严格来讲,“2+X”结构中的双音节词与“1+X”结构的双音节词不是相同的词。由于“2+X”结构的例频通常要比对应的双

音节例频低，因而低频的孤词也更多，因而能产性指数也相对更高。以“-者”为例，双音节词比相对应的“2+者”例频低的占少数：

表 3-1 二字组和三字组相对频率对比

2 音节词	例频	“2+者”例频	比较
消费	4742	2453	>
工作	59861	1669	>
投资	18578	1420	>
与会	1135	855	>
经营	9564	789	>
志愿	454	760	<
劳动	4647	724	>
爱好	866	504	>
受害	178	492	<
管理	23080	475	>
领导	19048	417	>
侵略	1625	377	>
组织	22574	312	>
获得	10221	296	>
旅游	6747	292	>
操作	1798	1	>
同情	1342	1	>
代销	46	1	>
开户	75	1	>
来稿	124	1	>
审问	130	1	>
承担	2909	1	>
持枪	145	1	>
贡献	5362	1	>
仿制	56	1	>

我们将各个“2+X”结构中的双音节成分和整体的例频相比较，分为双音节成分例频大于等于“2+X”例频和双音节成分例频小于“2+X”例频两种，并统计落在这两种类型的型频数量。一般来讲，双音节成分例频更高的情况占多数。

表 3-2 二字组和三字组相对频率对比

语素	$2 \geq 2+1$	$2 < 2+1$	举例(2<2+1)
-者	264	29	幸存者 目击者 无产者 倡言者 卖官者
-人	180	44	残疾人 当事人 候选人 明眼人 意中人
-员	306	26	公务员 售货员 宇航员 炊事员 守门员
-师	31	3	驯兽师 舞美师 农艺师
-家	85	5	作曲家 剧作家 汉学家 词作家 林学家
-手	15	8	多面手 刽子手 二把手 二传手 甩花手
-机	136	67	计算机 拖拉机 收音机 推土机 验钞机
-器	124	36	助听器 变压器 传感器 吸尘器 路由器
-品	77	11	战利品 随葬品 代用品 舶来品 护肤品
-馆	38	8	图书馆 博物馆 殡仪馆 国宾馆 水族馆
-室	79	13	办公室 教研室 阅览室 会客室 更衣室
-场	58	9	打麦场 垦殖场 储灰场 养蜂场 养鹿场
-率	133	23	成活率 上座率 出油率 产蛋率 通胀率
-度	31	8	摄氏度 感光度 能见度 冗余度 粘稠度
-值	18	2	附加值 缺省值

-型	73	11	外向型 粗放型 集约型 流线型 选装型
-化	185	12	多元化 白热化 低龄化 集约化 荒漠化
-性	426	36	一次性 前瞻性 能动性 可塑性 劣根性

构词成分可以自由成词，其构造组合越活跃，频率越高，在语言加工中倾向于使用组合方式构造，因而能产性更高。而那些频率较低，以至于黏着分布的成分，由于分布有限，反而容易从线性序列上被预测，因而更倾向于整体处理，因而能产性较低。相对频率不仅体现在派生词构造中，同时也体现在复合词构词中。如果说在第二章，使用基于孤词的计算方法存在有些数据稀疏导致计算数据与语感直觉违背，那么相对频率提供了另外的观察角度。

从更为普遍的角度来看，通过计算语言结构中的部分能否自由成词、在句子其他环境分布的频率和包含该成分的特定词语频率之间的关系，相对频率反映了复合结构中部分和整体的关系，与语言加工中成分组合和整词记忆之间动态平衡相关，从而提供了观察多种因素影响词汇加工的复杂机制的角度。尽管一般的语言学导论教材都会强调语言符号可以通过组合方式来构造更大的语言成分，具有组合性的特点，然而这并不是一个动态运作的语言系统的全貌。不论在词法，还是句法层面，语言成分在线性序列上组合数量都在发生变化。对一个系统而言，既存在组合相对自由、数量巨大的情况，同时也存在组合受限、数量较小的情况，两个极端一个是无限自由能产，一个是只能以黏着形式分布，无法自由运用而出现在与其他语言成分组合的结构中。换言之，这种语感直觉上的“自由/黏着”范畴性质可以体现在语言结构的成分和整体相对频率的数量对比上。

汉语在不同的字组层面都存在这种成分分解与整体加工之间的动态消长，构成了从结构透明的自由组合到成分黏着的固定搭配连续统：

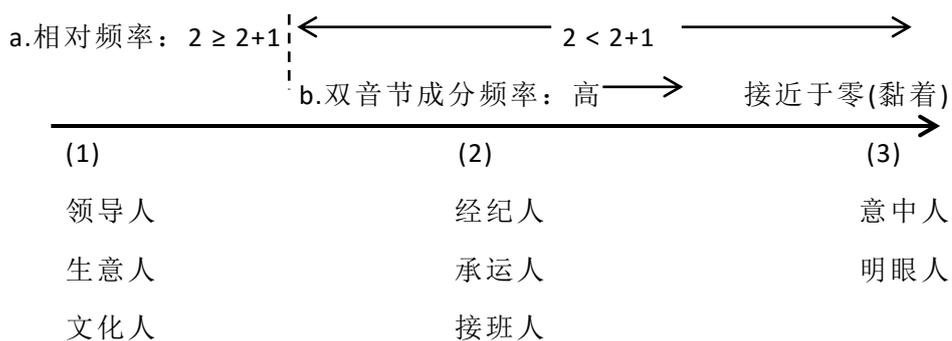


图 3-2 相对频率消长和结构自由/黏着变化

其中(1)和(3)类成分在有的语法体系中被划分性质不同的语法单位，前者是短语，后者是语素组(郭锐, 1996)。类似地，在二字组层面，依据语素的自由和黏着性质，“鸭蛋”被认为是词，而“鸡蛋”被认为是短语，这种划分通常与母语者的语感相违背。王洪君(2001)认为从区分自由语素和黏着语素、语素和词的角度无法区分不同的组合模式，因而以单字为出发点可以更好地区分汉语语法单位组合的不同模式。完全依照自由/黏着的性质划分单位层级，这种单位分类并没有从抓住语言单位组合变化的实质，在这一点上，以单字为出发点的确能够提供观察汉语不同组合模式的视角(王洪君, 2001)。而相对频率则提供了进一步描述其数量特征和性质变化的可能。因而，自由组合和黏着搭配结构之间的相互转化，需要在字组的层面观察相对频率的变化。

其次，有时结构直接成分划分的模糊性也和相对频率关联。例如，以“-品”结尾的三字组结构内部成分的相对频率可以显示出在划分直接成分结构“2+1”和“1+2”时的模糊界限。

表 3-3 相对频率与“2+品”结构

(1) [[复制][品]]			(2) ?[[木制][品]] ? [[木][制][品]]			(3) [[肉][制品]]		
	2 字组	2+品		2 字组	2+品		2 字组	2+品
复制品	353	70	木制品	69	12	肉制品	0	26
仿制品	56	20	竹制品	11	6	豆制品	0	63
精制品	68	1	纸制品	10	5	奶制品	0	47

首先，三字组中前两字组、后两字组和整体的相对频率显示其分解程度。从第(1)组到第(3)组，前两字组的例频在递减。第(3)组的“肉制”、“豆制”、“奶制”相对不具有独立成词的能力，上表中显示的例频为 0，在实际的语料中也存在“豆制食品”等搭配，但是一般没有进行扩展的能力，因而不存在“2+1”的划分可能性。与第(3)组相比，第(2)组的中间状态体现在，前两字组更具有独立成词的能力，前两字组与三字组的相对频率反映了结构的组合松散，“木制”、“竹制”和“纸制”可以加“的”扩展为“木制的桌子”、“竹制的凉席”、“纸制的荷花”。正是由于第(2)组前两字组的独立能力，使其结构划分存在语感上的两种可能性，这种特征正是被相对频率捕捉和发现的，尽管仍倾向于“1+2”的结构划分，而第(3)组的相对频率则显示其更倾向于被整词处理。

其次，三字组中前两字组和后两字组“制品”的频率对比关系，“制品”一词的例频为 731，与之相比，这三组的前两字组的例频都比较小，因而总的来讲，“1+2”更占据划分优势，在词表中第(1)组类型的有 4 个，第(2)和(3)组都是“1+2”结构的例子，有 11 个。

第三，语义因素在决定结构划分中也起作用。第(2)和(3)组的三字组第一个字都有比较齐整的材料属性，与第(1)组的差异较大，这些字组的语义是否齐整会影响直接成分结构的分析和新结构的类推。

另外一个例子是“2+器”结构。“青铜”的频率高于“铜器”，而“弦乐”则低于“乐器”。在不影响语义的前提下，内部两字组的频率高低会影响直接成分分析的划分。

表 3-4 相对频率与“2+器”

(1) [[青铜][器]]				(2) [[弦][乐器]]				(3) [[细][石器]]			
	前 2	后 2	2+器		前 2	后 2	2+器		前 2	后 2	2+器
青铜器	280	66	210	弦乐器	14	304	6	细石器	0	309	14
				管乐器	24		1	新石器	0		201
								旧石器	0		138

## 小结

在本章重点讨论了与能产性相关的结构组合透明度的问题，在字组层面观察相对频率，通过成分组合的数量特征和性质变化，可以显示出从黏着固定搭配到自由组合的能产性变化的连续统，这种变化既可以是基于心理感知的，也可以是历时上的组合性质变化。

## 结论

构词能产性现象反映了说话者具有创造新词的语言知识，因而是词法学研究的重要问题。生成学派继承了 Bloomfield 区分规则和不规则的思想，将是否具有能产性作为区分词法规则构词和词库词的重要标准。联结主义学派挑起的争论，大大促进了构词能产性和心理词库的研究。许多被揭示的语言现象都与基于语言使用的频率信息相关，因而频率成为揭示语言使用和语言结构形式互动的重要指标。而语料库技术的逐渐成熟和日益丰富的语料库建设成果，给那些不满于停留在内省定性研究方法的语言学家提供了有力的工具，来研究和观察词频信息和构词能产性的关系。在最近的二十余年，构词能产性研究主要围绕词频分布的统计、词频信息和词汇的心理加工、以及如何从文本中提取词汇知识的发现程序三方面展开。

上文在前人研究基础上，主要从两方面探讨了汉语复合词的能产性问题：

在第二章，利用基于语料库的词频表，本文大规模计算了汉语词根语素的能产性指数，并以汉语材料检验了基于孤词词频的能产性指数计算方法的准确性。计算结果显示，汉语双音节复合词中的一部分黏着词根仍具有能产性。同时，不同的词根语素在复合词中的分布位置上，能产性也不同，有的呈前后对称，而更多的是不对称的情况。第三，针对一些学者通过能产性来区分词根和词缀的做法，文中也给出了词缀能产性的指数，计算结果无法显示词根和词缀在能产性的数量标准上的区别，因而就目前已有的能产性测量方法来讲，无法证明区分词根、词缀的数量标准的有效性。此外，基于孤词词频的能产性指数计算方法能够显示出一些词根从二字组到三字组结构随着结构复杂度的增加而提高组合的自由度和能产性。

第三章则从结构成分和整体间的相对频率入手，来探讨结构凝固度和能产性之间的关系。通过对语料的分析，我们认为相对频率既可以成为语言心理感知和加工的信息，同时也可以显示出汉语单位组合的历时变化，从而为理论的整合解释提供可能性。此外，相对频率不仅通过影响语言结

构的分解和组合而影响能产性，同时会影响结构的直接成分分析，因而在范畴规则之外提供了可观察的数量特征。

本文在研究过程中还存在一些不足，主要体现在几个方面：

一是对汉语词汇系统在语料文本中的分布仍缺乏比较详尽的分析，型-例比例指数可以提供观察文本中新信息流动的速率，然而受材料所限，这方面研究并不充分。相关的问题是区分复合词分布位置来计算词根能产性的结果，由于汉语实际的文本分布中孤词随着文本扩充而增长的速度并不显著，因而产生大量的数据稀疏，导致计算结果与语感直觉相悖，因而如何计算低能产性成为一个问题。

二是虽然相对频率提供了语言感知和加工的心理基础，然而通用语料库的频率信息和个人词汇知识间仍存在较大的差距。事实上，关于个人词汇知识的构成和变异，相关研究仍比较薄弱。这会削弱将二者等同并论证频率信息和心理加工关系的论断。

三是本文由于主要基于词汇统计的分布描写和观察，尤其是基于个例观察而得到的关于相对频率的一些论断仍需要更多材料和更多的方法来继续检验其中的相关性。

本文遗留了一些将来有待于进一步研究的重要问题：

第一，对于稀疏的词汇分布，如何度量其能产性？

第二，文中在位置分布和相对频率部分都提到了语言系统新形式的构造和填充存在多种路径的可能性，这既是语言演化不确定性的表现，同时又可能导致直接成分族之间语义关系不平行，因而类推的推理和形式构造如何以动态的形式来实现？

第三，本文没有涉及语义对复合词能产性的影响，直接成分族的语义成分所处的概念层级的差异，是否会对能产性和类推产生影响？



### 附录一 词根语素能产性指数(示例)

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
1	爱	ài	30	8916	0	0	爱情 爱国 爱好	35	9607	1	0.000104	恋爱 可爱 热爱
2	碍	ài	2	113	0	0	碍事 碍眼	4	2792	0	0	障碍 阻碍 妨碍
3	安	ān	45	23917	0	0	安全 安排 安慰	17	7688	0	0	公安 治安 不安
4	按	àn	14	8152	0	0	按照 按期 按时	0	0	0	0	
5	案	àn	11	4457	0	0	案件 案例 案情	64	14797	3	0.000203	方案 草案 答案
6	把	bǎ	18	4459	4	0.000897	把握 把头 把手	8	442	0	0	火把 刀把 车把
7	摆	bǎi	18	3236	3	0.000927	摆脱 摆放 摆手	5	199	1	0.005025	摇摆 下摆 钟摆
8	败	bài	19	943	1	0.001060	败坏 败类 败北	19	8528	1	0.000117	失败 腐败 击败
9	班	bān	15	1438	1	0.000695	班长 班级 班机	35	5168	0	0	上班 下班 航班
10	般	bān	1	57	0	0	般配	5	10981	1	0.000091	一般 这般 百般
11	版	bǎn	10	1370	0	0	版权 版本 版图	33	5523	5	0.000905	出版 盗版 翻版
12	办	bàn	15	16053	0	0	办法 办理 办事	48	12314	1	0.000081	举办 主办 创办
13	半	bàn	29	3785	1	0.000264	半岛 半夜 半点	11	4144	2	0.000483	一半 多半 大半
14	伴	bàn	14	1281	1	0.000781	伴随 伴侣 伴奏	12	3480	0	0	伙伴 同伴 老伴
15	邦	bāng	2	166	0	0	邦交 邦联	7	2109	1	0.000474	联邦 邻邦 兴邦
16	膀	bǎng	1	3	0	0	膀臂	4	1774	1	0.000564	肩膀 翅膀 臂膀
17	包	bāo	41	17795	2	0.000112	包括 包装 包围	37	4289	7	0.001632	承包 面包 书包
18	胞	bāo	6	94	1	0.010638	胞胎 胞弟 胞衣	8	4632	3	0.000648	同胞 细胞 台胞
19	保	bǎo	51	47013	2	0.000043	保护 保持 保证	21	3980	3	0.000754	确保 担保 投保

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
20	报	bào	64	29558	2	0.000068	报道 报告 报纸	80	23873	5	0.000209	本报 汇报 日报
21	爆	bào	11	3686	1	0.000271	爆发 爆炸 爆竹	5	423	1	0.002364	火爆 引爆 起爆
22	背	bèi	27	5497	4	0.000728	背景 背影 背后	18	1828	3	0.001641	违背 后背 驼背
23	备	bèi	24	1269	2	0.001576	备战 备案 备用	29	25688	1	0.000039	准备 设备 具备
24	比	bǐ	25	38124	3	0.000079	比赛 比较 比照	16	5141	3	0.000584	相比 无比 对比
25	彼	bǐ	3	2415	0	0	彼此 彼岸 彼时	0	0	0	0	
26	币	bì	3	66	1	0.015152	币值 币制 币面	24	4562	2	0.000438	货币 港币 外币
27	必	bì	13	27220	0	0	必须 必要 必然	6	5828	0	0	不必 何必 未必
28	避	bì	12	3790	1	0.000264	避免 避开 避孕	8	1927	2	0.001038	回避 逃避 躲避
29	边	biān	28	5584	2	0.000358	边界 边境 边缘	59	30597	6	0.000196	一边 身边 旁边
30	编	biān	37	6268	1	0.000160	编辑 编制 编写	30	2249	5	0.002223	主编 改编 新编
31	变	biàn	42	19778	2	0.000101	变化 变成 变革	39	13155	0	0	改变 转变 演变
32	遍	biàn	6	1104	1	0.000906	遍布 遍地 遍及	7	4060	1	0.000246	普遍 传遍 走遍
33	标	biāo	32	15764	3	0.000190	标准 标题 标志	34	16162	0	0	目标 指标 招标
34	表	biǎo	46	47140	4	0.000085	表示 表现 表明	39	28447	8	0.000281	代表 发表 外表
35	别	bié	17	14187	0	0	别人 别墅 别国	37	34472	2	0.000058	特别 分别 区别
36	宾	bīn	8	1774	1	0.000564	宾馆 宾主 宾客	12	1292	0	0	嘉宾 来宾 贵宾
37	兵	bīng	34	3432	1	0.000291	兵力 兵团 兵器	67	10417	4	0.000384	士兵 官兵 当兵
38	病	bìng	49	7105	1	0.000141	病人 病毒 病情	54	6752	5	0.000741	疾病 毛病 生病
39	补	bǔ	40	4822	4	0.000830	补充 补偿 补贴	19	2205	5	0.002268	弥补 互补 填补
40	不	bù	150	195312	3	0.000015	不能 不同 不断	12	7728	0	0	毫不 绝不 从不
41	布	bù	40	3499	11	0.003144	布局 布置 布满	54	15047	6	0.000399	宣布 公布 发布
42	怖	bù	0	0	0	0		1	1404	0	0	恐怖
43	部	bù	19	46267	0	0	部门 部队 部分	64	48746	7	0.000144	干部 全部 西部

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
44	才	cái	19	2893	2	0.000691	才能 才华 才子	27	15301	2	0.000131	人才 天才 奴才
45	材	cái	2	4163	0	0	材料 材质	33	5496	7	0.001274	身材 题材 教材
46	财	cái	26	12789	1	0.000078	财政 财富 财产	17	1338	0	0	发财 钱财 理财
47	彩	cǎi	35	2756	4	0.001451	彩色 彩电 彩票	25	4255	3	0.000705	色彩 精彩 光彩
48	菜	cài	38	1908	5	0.002621	菜肴 菜单 菜刀	66	5731	4	0.000698	蔬菜 饭菜 白菜
49	参	cān	33	31548	0	0	参加 参与 参观	1	52	0	0	内参
50	操	cāo	13	3551	2	0.000563	操作 操场 操心	12	1076	0	0	体操 情操 早操
51	策	cè	6	2157	0	0	策略 策划 策应	13	18269	2	0.000109	政策 决策 对策
52	册	cè	2	67	0	0	册封 册页	20	2269	4	0.001763	注册 画册 手册
53	测	cè	14	2379	1	0.000420	测试 测量 测定	19	5040	0	0	预测 监测 检测
54	曾	céng	1	5500	0	0	曾经	2	982	0	0	不曾 未曾
55	层	céng	8	3283	2	0.000609	层次 层面 层级	30	10723	1	0.000093	基层 阶层 高层
56	差	chā	9	4873	1	0.000205	差距 差异 差别	21	1377	0	0	反差 偏差 逆差
57	查	chá	45	4704	4	0.000850	查处 查获 查询	28	17659	4	0.000227	调查 检查 审查
58	察	chá	3	722	0	0	察觉 察看 察访	18	13558	1	0.000074	考察 警察 观察
59	差	chà	2	1171	0	0	差点 差劲	1	380	0	0	相差
60	差	chāi	5	197	1	0.005076	差事 差使 差遣	12	864	0	0	出差 当差 苦差
61	产	chǎn	26	37456	1	0.000027	产品 产生 产业	61	32059	5	0.000156	生产 资产 财产
62	场	cháng	1	28	0	0	场院	3	54	0	0	打场 扬场 赶场
63	常	cháng	29	7554	1	0.000132	常委 常务 常见	20	35608	1	0.000028	非常 经常 正常
64	长	cháng	78	14318	9	0.000629	长期 长远 长久	28	4276	0	0	延长 漫长 全长
65	场	chǎng	21	6704	2	0.000298	场所 场面 场合	103	54197	9	0.000166	市场 机场 立场
66	厂	chǎng	18	2984	4	0.001340	厂长 厂家 厂商	29	3840	6	0.001563	工厂 电厂 总厂
67	超	chāo	38	11516	1	0.000087	超过 超凡 超越	4	328	0	0	高超 赶超 入超

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
68	车	chē	63	8792	3	0.000341	车辆 车道 车站	113	23141	8	0.000346	汽车 火车 轿车
69	彻	chè	3	3357	0	0	彻底 彻夜 彻骨	3	3852	0	0	贯彻 透彻 响彻
70	沉	chén	39	8167	3	0.000367	沉默 沉重 沉浸	9	1408	0	0	深沉 低沉 阴沉
71	晨	chén	11	337	2	0.005935	晨练 晨曦 晨报	5	3713	0	0	早晨 凌晨 清晨
72	臣	chén	4	181	0	0	臣民 臣服 臣子	12	1810	3	0.001657	大臣 功臣 忠臣
73	称	chèn	4	242	1	0.004132	称职 称心 称身	3	387	0	0	对称 相称 匀称
74	称	chēng	16	6087	1	0.000164	称为 称号 称赞	34	6247	1	0.000160	名称 声称 简称
75	成	chéng	72	80472	6	0.000075	成为 成功 成立	52	57136	3	0.000053	形成 完成 造成
76	呈	chéng	9	1516	0	0	呈现 呈报 呈交	3	150	0	0	辞呈 纷呈 签呈
77	承	chéng	25	13267	0	0	承认 承担 承诺	9	1901	0	0	继承 传承 奉承
78	城	chéng	29	22098	2	0.000091	城市 城镇 城乡	46	5073	5	0.000986	县城 古城 进城
79	程	chéng	5	8601	1	0.000116	程度 程序 程控	50	33797	2	0.000059	工程 过程 进程
80	持	chí	10	6113	0	0	持续 持久 持有	17	38187	0	0	支持 坚持 保持
81	冲	chōng	32	8076	4	0.000495	冲突 冲击 冲动	6	366	0	0	俯冲 缓冲 要冲
82	充	chōng	20	14779	0	0	充分 充满 充实	5	2196	0	0	补充 冒充 扩充
83	重	chóng	37	11641	2	0.000172	重新 重复 重申	1	558	0	0	双重
84	冲	chòng	3	66	0	0	冲劲 冲压 冲床	0	0	0	0	
85	出	chū	115	88089	1	0.000011	出来 出现 出去	100	76060	15	0.000197	提出 指出 作出
86	初	chū	53	7827	4	0.000511	初步 初期 初中	9	4195	2	0.000477	当初 最初 年初
87	除	chú	20	1943	2	0.001029	除非 除去 除夕	37	7581	3	0.000396	消除 解除 排除
88	厨	chú	3	1778	0	0	厨房 厨师 厨具	4	90	0	0	下厨 名厨 帮厨
89	础	chǔ	0	0	0	0		1	13955	0	0	基础
90	处	chǔ	20	16684	1	0.000060	处理 处于 处罚	14	4469	2	0.000448	查处 相处 地处
91	触	chù	13	1257	0	0	触动 触及 触摸	4	3661	0	0	接触 抵触 感触

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
92	传	chuán	52	22150	1	0.000045	传统 传来 传播	25	7270	1	0.000138	宣传 流传 遗传
93	窗	chuāng	20	4396	2	0.000455	窗口 窗外 窗户	15	1046	3	0.002868	车窗 门窗 橱窗
94	创	chuāng	4	221	0	0	创伤 创口 创面	1	174	0	0	重创
95	创	chuàng	23	21196	0	0	创造 创新 创作	10	1371	1	0.000729	开创 争创 首创
96	纯	chún	21	2697	1	0.000371	纯粹 纯洁 纯真	3	1458	0	0	单纯 清纯 提纯
97	此	cǐ	31	19451	3	0.000154	此外 此次 此时	14	29570	0	0	因此 如此 从此
98	刺	cì	20	3806	1	0.000263	刺激 刺杀 刺刀	14	1094	1	0.000914	讽刺 冲刺 遇刺
99	聪	cōng	4	2524	0	0	聪明 聪慧 聪颖	2	27	1	0.037037	失聪 耳聪
100	促	cù	7	11697	1	0.000085	促进 促使 促销	10	1803	0	0	督促 急促 敦促
101	脆	cuì	5	768	1	0.001302	脆弱 脆性 脆骨	6	1735	1	0.000576	干脆 清脆 酥脆
102	村	cūn	29	6433	1	0.000155	村民 村里 村庄	20	12653	3	0.000237	农村 乡村 全村
103	存	cún	27	12351	0	0	存在 存款 存心	36	7526	6	0.000797	生存 保存 库存
104	磋	cuō	1	1256	0	0	磋商	1	131	0	0	切磋
105	措	cuò	4	8150	0	0	措施 措辞 措词	3	1238	0	0	举措 筹措 失措
106	错	cuò	21	5681	3	0.000528	错误 错觉 错事	13	5300	0	0	小错 没错 认错
107	达	dá	7	13128	0	0	达到 达成 达标	18	11554	0	0	表达 发达 到达
108	答	dā	1	2776	0	0	答应	0	0	0	0	
109	答	dá	11	3273	0	0	答案 答道 答复	11	7549	1	0.000132	回答 解答 应答
110	打	dǎ	112	28476	7	0.000246	打开 打击 打算	30	3183	0	0	单打 双打 挨打
111	代	dài	32	22149	1	0.000045	代表 代价 代理	36	35775	1	0.000028	年代 时代 现代
112	贷	dài	4	4543	2	0.000440	贷款 贷方 贷户	9	985	1	0.001015	信贷 借贷 还贷
113	袋	dài	3	115	0	0	袋装 袋鼠 袋料	20	4071	2	0.000491	脑袋 口袋 麻袋
114	待	dài	16	2011	1	0.000497	待遇 待人 待客	21	11257	1	0.000089	对待 等待 接待
115	单	dān	69	18390	7	0.000381	单位 单纯 单独	27	9139	3	0.000328	简单 名单 订单

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
116	担	dān	10	9475	1	0.000106	担心 担任 担保	3	5647	0	0	承担 负担 分担
117	胆	dǎn	19	1235	2	0.001619	胆量 胆小 胆怯	12	1433	1	0.000698	大胆 斗胆 壮胆
118	但	dàn	2	23333	0	0	但是 但愿	3	3447	0	0	不但 非但 岂但
119	担	dàn	0	0	0	0		4	320	1	0.003125	重担 扁担 石担
120	当	dāng	46	61985	7	0.000113	当时 当然 当地	20	14914	1	0.000067	相当 应当 正当
121	党	dǎng	40	9734	3	0.000308	党员 党委 党内	27	3200	3	0.000938	全党 政党 入党
122	当	dàng	7	7706	0	0	当年 当天 当晚	14	4491	0	0	适当 恰当 不当
123	导	dǎo	26	8532	3	0.000352	导致 导演 导弹	26	30616	2	0.000065	领导 指导 引导
124	蹈	dǎo	1	1	1	1.000000	蹈袭	2	1285	1	0.000778	舞蹈 重蹈
125	到	dào	19	12855	0	0	到底 到处 到达	49	112348	3	0.000027	看到 得到 感到
126	德	dé	5	465	0	0	德育 德行 德政	20	5033	1	0.000199	道德 美德 品德
127	登	dēng	20	4211	0	0	登记 登陆 登山	5	871	0	0	刊登 攀登 荣登
128	的	dī	1	19	0	0	的哥	2	117	0	0	面的 打的
129	低	dī	49	7263	2	0.000275	低头 低落 低下	8	4611	0	0	降低 最低 高低
130	敌	dí	25	5409	2	0.000370	敌人 敌对 敌军	31	1308	2	0.001529	歼敌 情敌 无敌
131	底	dǐ	29	2278	2	0.000878	底下 底层 底蕴	48	11391	4	0.000351	到底 彻底 年底
132	抵	dǐ	19	4731	4	0.000845	抵达 抵抗 抵制	5	489	0	0	大抵 飞抵 相抵
133	第	dì	10	68334	0	0	第一 第二 第三	6	110	1	0.009091	门第 次第 宅第
134	地	dì	103	84417	9	0.000107	地区 地方 地位	120	49909	4	0.000080	土地 当地 各地
135	的	dì	0	0	0	0		4	6471	1	0.000155	目的 标的 寻的
136	弟	dì	5	1115	1	0.000897	弟兄 弟子 弟妹	13	4847	0	0	兄弟 子弟 小弟
137	帝	dì	6	852	0	0	帝国 帝王 帝制	8	4207	1	0.000238	皇帝 上帝 大帝
138	典	diǎn	10	3678	0	0	典型 典礼 典范	22	3188	3	0.000941	经典 古典 庆典
139	点	diǎn	32	4721	2	0.000424	点头 点燃 点评	104	41003	5	0.000122	重点 特点 有点

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
140	电	diàn	98	43320	9	0.000208	电话 电视 电影	52	5564	5	0.000899	邮电 机电 彩电
141	店	diàn	11	1149	1	0.000870	店铺 店主 店员	23	6094	3	0.000492	饭店 商店 酒店
142	调	diào	30	11231	2	0.000178	调查 调动 调研	43	11266	3	0.000266	强调 单调 情调
143	定	dìng	62	7006	5	0.000714	定期 定位 定义	68	96072	3	0.000031	一定 决定 规定
144	订	dìng	15	1848	0	0	订单 订货 订婚	16	3660	1	0.000273	签订 制订 修订
145	冬	dōng	26	3692	3	0.000813	冬天 冬季 冬眠	17	856	2	0.002336	过冬 严冬 隆冬
146	懂	dǒng	4	2594	0	0	懂得 懂事 懂行	3	260	0	0	懵懂 易懂 难懂
147	斗	dǒu	7	209	1	0.004785	斗笠 斗胆 斗篷	14	395	1	0.002532	烟斗 泰斗 漏斗
148	斗	dòu	12	5760	1	0.000174	斗争 斗志 斗嘴	18	5944	0	0	战斗 奋斗 搏斗
149	督	dū	9	851	0	0	督促 督察 督办	5	6811	0	0	监督 总督 都督
150	独	dú	33	9949	0	0	独立 独特 独自	5	2361	0	0	孤独 单独 唯独
151	毒	dú	39	3680	7	0.001902	毒品 毒瘾 毒气	40	5657	6	0.001061	病毒 吸毒 禁毒
152	渡	dù	6	603	0	0	渡过 渡口 渡船	14	2018	2	0.000991	过渡 偷渡 引渡
153	断	duàn	34	2246	2	0.000890	断定 断层 断然	36	19368	2	0.000103	不断 判断 打断
154	段	duàn	2	159	0	0	段落 段位	35	13295	2	0.000150	阶段 手段 身段
155	锻	duàn	7	1645	1	0.000608	锻炼 锻造 锻压	0	0	0	0	
156	对	duì	66	42885	2	0.000047	对于 对方 对象	19	23895	1	0.000042	面对 反对 绝对
157	队	duì	15	11125	1	0.000090	队伍 队员 队长	47	28593	1	0.000035	部队 军队 球队
158	顿	dùn	8	1742	2	0.001148	顿时 顿悟 顿觉	7	1993	1	0.000502	整顿 停顿 安顿
159	夺	duó	12	3761	1	0.000266	夺得 夺取 夺冠	13	2389	2	0.000837	争夺 剥夺 掠夺
160	朵	duǒ	0	0	0	0		3	2021	0	0	耳朵 花朵 云朵
161	额	é	6	999	1	0.001001	额头 额外 额度	30	7872	1	0.000127	总额 金额 份额
162	儿	ér	9	6433	0	0	儿童 儿女 儿时	18	9406	2	0.000213	女儿 婴儿 孤儿
163	耳	ěr	31	3692	2	0.000542	耳朵 耳边 耳光	16	731	0	0	刺耳 悦耳 木耳

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
164	发	fā	121	155931	4	0.000026	发展 发现 发生	66	30040	3	0.000100	开发 出发 爆发
165	乏	fá	2	416	0	0	乏味 乏力	10	4072	1	0.000246	缺乏 不乏 匮乏
166	伐	fá	3	84	1	0.011905	伐木 伐树 伐区	7	2250	0	0	步伐 采伐 砍伐
167	法	fǎ	51	24651	2	0.000081	法律 法规 法院	112	64901	10	0.000154	办法 方法 无法
168	发	fà	12	436	1	0.002294	发丝 发廊 发髻	28	4458	3	0.000673	头发 长发 白发
169	繁	fán	21	5483	1	0.000182	繁荣 繁华 繁殖	3	1086	0	0	频繁 纷繁 浩繁
170	烦	fán	12	1476	2	0.001355	烦恼 烦躁 烦心	4	2728	0	0	麻烦 厌烦 心烦
171	反	fǎn	62	33216	6	0.000181	反对 反映 反正	8	4679	0	0	相反 违反 造反
172	范	fàn	6	5717	0	0	范围 范畴 范例	12	6637	3	0.000452	规范 示范 模范
173	泛	fàn	7	690	1	0.001449	泛滥 泛起 泛舟	5	4532	0	0	广泛 宽泛 空泛
174	犯	fàn	14	5417	1	0.000185	犯罪 犯人 犯规	31	3691	5	0.001355	侵犯 罪犯 战犯
175	饭	fàn	21	4460	1	0.000224	饭店 饭馆 饭菜	35	7007	0	0	吃饭 晚饭 做饭
176	方	fāng	42	62683	4	0.000064	方面 方式 方法	98	66680	7	0.000105	地方 双方 对方
177	房	fáng	28	7595	2	0.000263	房间 房屋 房门	79	10525	3	0.000285	住房 厨房 病房
178	防	fáng	49	10286	3	0.000292	防止 防治 防御	31	5530	4	0.000723	国防 预防 边防
179	访	fǎng	7	9620	0	0	访问 访华 访谈	26	9751	2	0.000205	采访 来访 拜访
180	放	fàng	60	20773	1	0.000048	放在 放弃 放心	36	16944	5	0.000295	开放 解放 发放
181	飞	fēi	62	11036	8	0.000725	飞机 飞行 飞扬	6	1077	0	0	起飞 腾飞 放飞
182	非	fēi	13	20108	1	0.000050	非常 非法 非得	13	5215	1	0.000192	并非 除非 是非
183	费	fèi	13	3381	1	0.000296	费用 费力 费劲	44	14708	0	0	消费 收费 经费
184	氛	fēn	1	960	0	0	氛围	1	2664	0	0	气氛
185	份	fèn	2	620	0	0	份额 份饭	11	4944	0	0	身份 股份 省份
186	奋	fèn	8	3490	1	0.000287	奋斗 奋力 奋战	7	3570	0	0	兴奋 勤奋 振奋
187	丰	fēng	24	6865	1	0.000146	丰富 丰收 丰满	0	0	0	0	

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
188	风	fēng	96	17381	7	0.000403	风险 风格 风景	106	9928	5	0.000504	作风 党风 整风
189	否	fǒu	4	6089	0	0	否则 否定 否认	4	8422	0	0	是否 能否 与否
190	肤	fū	4	384	1	0.002604	肤色 肤浅 肤泛	2	1730	0	0	皮肤 肌肤
191	福	fú	15	1937	2	0.001033	福利 福气 福祉	19	5366	1	0.000186	幸福 祝福 造福
192	扶	fú	16	3713	1	0.000269	扶贫 扶持 扶植	4	394	1	0.002538	搀扶 帮扶 包扶
193	府	fǔ	4	89	0	0	府上 府第 府邸	19	32425	3	0.000093	政府 首府 学府
194	腐	fǔ	10	3542	0	0	腐败 腐朽 腐蚀	4	706	0	0	豆腐 陈腐 迂腐
195	妇	fù	9	11804	0	0	妇女 妇人 妇幼	22	4724	1	0.000212	夫妇 媳妇 寡妇
196	父	fù	11	17116	0	0	父亲 父母 父子	25	1914	2	0.001045	师父 祖父 岳父
197	负	fù	22	10802	3	0.000278	负责 负担 负面	14	3142	1	0.000318	欺负 担负 肩负
198	复	fù	60	8148	2	0.000245	复杂 复苏 复兴	16	11565	1	0.000086	恢复 反复 重复
199	富	fù	23	5780	0	0	富有 富裕 富强	9	8566	0	0	丰富 财富 致富
200	附	fù	25	4575	2	0.000437	附近 附属 附加	6	296	0	0	依附 吸附 攀附
201	付	fù	18	3077	3	0.000975	付出 付款 付诸	17	4758	0	0	支付 对付 应付
202	改	gǎi	41	44544	0	0	改革 改变 改善	17	3873	3	0.000775	修改 整改 劳改
203	概	gài	13	3860	0	0	概念 概括 概率	5	4298	1	0.000233	大概 气概 一概
204	感	gǎn	28	41084	0	0	感到 感觉 感情	34	7876	1	0.000127	情感 敏感 反感
205	赶	gǎn	22	8053	0	0	赶紧 赶到 赶快	2	461	0	0	追赶 驱赶
206	岗	gǎng	7	1882	0	0	岗位 岗亭 岗哨	17	2483	0	0	下岗 上岗 站岗
207	高	gāo	117	38662	3	0.000078	高兴 高度 高级	23	24687	2	0.000081	提高 最高 崇高
208	搞	gǎo	5	1960	0	0	搞好 搞活 搞垮	0	0	0	0	
209	告	gào	23	16640	0	0	告诉 告别 告诫	31	15841	0	0	报告 广告 警告
210	革	gé	5	6992	0	0	革命 革新 革职	5	21789	0	0	改革 变革 皮革
211	格	gé	9	3108	0	0	格局 格外 格式	28	22659	3	0.000132	价格 严格 性格

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
212	各	gè	46	47992	6	0.000125	各种 各地 各级	0	0	0	0	
213	根	gēn	21	22070	2	0.000091	根本 根据 根源	33	1268	4	0.003155	扎根 墙根 生根
214	更	gēng	12	1977	0	0	更新 更换 更改	4	430	1	0.002326	变更 五更 打更
215	耕	gēng	8	1921	1	0.000521	耕地 耕作 耕耘	10	321	2	0.006231	春耕 备耕 复耕
216	更	gèng	4	8597	0	0	更加 更为 更是	0	0	0	0	
217	工	gōng	59	102607	6	0.000058	工作 工程 工业	100	28368	6	0.000212	职工 员工 加工
218	供	gōng	16	4518	1	0.000221	供应 供给 供水	2	12536	1	0.000080	提供 储供
219	功	gōng	20	7713	0	0	功能 功夫 功课	34	13562	3	0.000221	成功 武功 气功
220	攻	gōng	17	5864	1	0.000171	攻击 攻势 攻占	16	5209	1	0.000192	进攻 反攻 围攻
221	巩	gǒng	2	1776	0	0	巩固 巩膜	0	0	0	0	
222	共	gòng	25	15523	2	0.000129	共同 共识 共有	10	3620	0	0	公共 一共 总共
223	贡	gòng	6	5413	1	0.000185	贡献 贡品 贡税	2	70	0	0	进贡 朝贡
224	供	gòng	9	554	0	0	供养 供奉 供职	9	280	1	0.003571	逼供 招供 口供
225	沟	gōu	8	2445	2	0.000818	沟通 沟壑 沟渠	16	544	2	0.003676	山沟 鸿沟 水沟
226	构	gòu	9	4723	0	0	构成 构造 构筑	9	19248	3	0.000156	机构 结构 虚构
227	购	gòu	17	4746	0	0	购买 购物 购置	22	4666	0	0	采购 收购 并购
228	估	gū	7	2836	0	0	估计 估算 估价	4	1639	0	0	评估 低估 高估
229	古	gǔ	60	9659	5	0.000518	古代 古老 古人	15	2614	0	0	考古 千古 远古
230	股	gǔ	18	6976	0	0	股票 股份 股市	13	2776	1	0.000360	参股 控股 入股
231	骨	gǔ	28	2926	2	0.000684	骨干 骨肉 骨折	60	1626	7	0.004305	排骨 头骨 肋骨
232	鼓	gǔ	27	6015	3	0.000499	鼓励 鼓舞 鼓吹	24	627	6	0.009569	锣鼓 打鼓 钟鼓
233	固	gù	13	3050	2	0.000656	固定 固然 固执	7	3539	0	0	巩固 牢固 顽固
234	故	gù	28	10042	1	0.000100	故事 故意 故乡	20	3648	1	0.000274	事故 缘故 已故
235	怪	guài	20	1164	0	0	怪异 怪物 怪事	15	4541	1	0.000220	奇怪 难怪 古怪

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
236	冠	guān	4	5276	0	0	冠军 冠名 冠词	1	423	0	0	夺冠
237	官	guān	46	9272	3	0.000324	官员 官兵 官方	43	6083	4	0.000658	军官 法官 长官
238	馆	guǎn	9	579	1	0.001727	馆长 馆内 馆里	24	5374	3	0.000558	宾馆 使馆 旅馆
239	惯	guàn	12	919	4	0.004353	惯例 惯性 惯用	2	4789	0	0	习惯 娇惯
240	冠	guàn	1	12	0	0	冠冕	16	657	0	0	桂冠 花冠 衣冠
241	观	guàn	0	0	0	0		1	21	0	0	道观
242	光	guāng	59	9540	2	0.000210	光荣 光明 光芒	87	18426	7	0.000380	目光 阳光 眼光
243	广	guǎng	18	20231	2	0.000099	广大 广泛 广播	5	2977	2	0.000672	推广 宽广 深广
244	归	guī	39	3383	2	0.000591	归来 归还 归纳	10	1697	1	0.000589	回归 总归 划归
245	规	guī	16	31552	1	0.000032	规定 规模 规划	21	4349	1	0.000230	法规 常规 正规
246	轨	guǐ	5	1486	3	0.002019	轨道 轨迹 轨枕	24	1647	3	0.001821	接轨 转轨 仪轨
247	国	guó	90	112684	9	0.000080	国家 国际 国内	64	80825	3	0.000037	全国 我国 祖国
248	过	guò	57	45684	3	0.000066	过去 过程 过来	36	54130	1	0.000018	通过 经过 不过
249	还	hái	3	26056	0	0	还有 还是 还要	0	0	0	0	
250	孩	hái	2	153	0	0	孩童 孩提	6	9032	0	0	女孩 小孩 男孩
251	害	hài	11	3076	0	0	害怕 害人 害羞	43	11551	4	0.000346	伤害 厉害 损害
252	海	hǎi	100	19489	9	0.000462	海外 海洋 海军	44	5090	3	0.000589	沿海 大海 脑海
253	憾	hàn	1	26	0	0	憾事	2	1758	0	0	遗憾 缺憾
254	行	háng	19	7904	5	0.000633	行业 行列 行长	23	14852	0	0	银行 同行 分行
255	航	háng	21	5728	0	0	航空 航线 航天	22	2012	0	0	民航 通航 导航
256	好	hǎo	43	20386	1	0.000049	好人 好处 好事	37	40513	0	0	友好 不好 良好
257	好	hào	10	1775	0	0	好奇 好客 好色	4	1210	0	0	爱好 嗜好 偏好
258	喝	hē	2	2031	0	0	喝酒 喝茶	1	133	0	0	吃喝
259	合	hé	58	41184	4	0.000097	合作 合同 合理	53	33813	4	0.000118	结合 联合 综合

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
260	和	hé	33	15410	1	0.000065	和平 和谐 和谈	35	4462	1	0.000224	温和 平和 缓和
261	核	hé	23	5513	2	0.000363	核心 核算 核査	13	1958	0	0	考核 审核 稽核
262	和	hè	0	0	0	0		4	299	1	0.003344	附和 应和 唱和
263	喝	hè	6	398	1	0.002513	喝彩 喝道 喝令	1	274	0	0	吆喝
264	贺	hè	15	721	1	0.001387	贺电 贺信 贺卡	9	1590	3	0.001887	祝贺 庆贺 可贺
265	黑	hēi	65	7848	7	0.000892	黑色 黑暗 黑人	18	920	3	0.003261	漆黑 乌黑 黝黑
266	衡	héng	2	691	0	0	衡量 衡器	7	3657	0	0	平衡 均衡 抗衡
267	宏	hóng	9	3033	1	0.000330	宏观 宏伟 宏大	1	66	0	0	恢宏
268	红	hóng	76	8020	9	0.001122	红军 红色 红旗	40	2795	5	0.001789	脸红 通红 大红
269	洪	hóng	13	2252	0	0	洪水 洪峰 洪涝	6	1453	0	0	抗洪 防洪 分洪
270	候	hòu	8	472	1	0.002119	候选 候补 候鸟	18	36976	1	0.000027	时候 气候 问候
271	呼	hū	11	5495	0	0	呼吁 呼吸 呼声	11	2781	2	0.000719	招呼 称呼 欢呼
272	忽	hū	6	5396	0	0	忽然 忽视 忽略	3	335	0	0	疏忽 飘忽 倏忽
273	互	hù	22	6807	2	0.000294	互相 互利 互补	2	4573	0	0	相互 交互
274	户	hù	8	1006	0	0	户口 户外 户籍	40	10226	6	0.000587	客户 用户 窗户
275	护	hù	29	2355	3	0.001274	护士 护照 护理	22	20325	0	0	保护 维护 拥护
276	划	huá	8	304	0	0	划算 划船 划桨	0	0	0	0	
277	化	huà	24	6545	1	0.000153	化工 化学 化石	88	49156	6	0.000122	文化 变化 深化
278	划	huà	14	1737	3	0.001727	划分 划定 划清	11	17707	2	0.000113	计划 规划 策划
279	话	huà	18	5018	1	0.000199	话题 话语 话剧	79	39414	3	0.000076	电话 说话 讲话
280	怀	huái	15	6704	0	0	怀疑 怀里 怀孕	15	2467	2	0.000811	关怀 情怀 满怀
281	坏	huài	11	1974	1	0.000507	坏人 坏事 坏话	9	4737	0	0	破坏 损坏 吓坏
282	欢	huān	23	8093	0	0	欢迎 欢乐 欢呼	7	12857	1	0.000078	喜欢 联欢 狂欢
283	环	huán	20	17466	0	0	环境 环节 环绕	13	1498	1	0.000668	循环 光环 耳环

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
284	还	huán	18	918	1	0.001089	还给 还原 还击	13	1287	0	0	归还 偿还 退还
285	换	huàn	30	1568	2	0.001276	换上 换取 换届	18	4857	0	0	交换 转换 兑换
286	患	huàn	6	2576	0	0	患者 患病 患儿	12	768	2	0.002604	隐患 忧患 水患
287	挥	huī	13	1539	1	0.000650	挥手 挥舞 挥霍	2	11342	0	0	发挥 指挥
288	辉	huī	3	1571	0	0	辉煌 辉映 辉石	4	611	0	0	光辉 生辉 余辉
289	回	huí	80	50094	3	0.000060	回来 回答 回去	30	6038	2	0.000331	返回 收回 来回
290	会	huì	53	41156	5	0.000121	会议 会见 会谈	76	87802	2	0.000023	社会 机会 大会
291	慧	huì	3	129	0	0	慧眼 慧黠 慧心	7	2399	2	0.000834	智慧 聪慧 早慧
292	惠	huì	4	43	1	0.023256	惠及 惠顾 惠存	8	2044	2	0.000978	优惠 实惠 互惠
293	婚	hūn	19	4412	0	0	婚姻 婚礼 婚后	25	8362	0	0	结婚 离婚 新婚
294	魂	hún	2	102	0	0	魂魄 魂灵	22	2100	4	0.001905	灵魂 鬼魂 销魂
295	浑	hún	8	1770	1	0.000565	浑身 浑浊 浑厚	2	75	0	0	雄浑 圆浑
296	混	hún	1	538	0	0	混蛋	0	0	0	0	
297	混	hùn	17	2797	0	0	混乱 混合 混淆	5	187	0	0	含混 鬼混 蒙混
298	活	huó	41	26551	7	0.000264	活动 活力 活跃	37	33991	2	0.000059	生活 灵活 干活
299	和	huó	1	34	0	0	和面	0	0	0	0	
300	火	huǒ	82	10018	9	0.000898	火车 火箭 火山	83	6994	4	0.000572	停火 炮火 大火
301	或	huò	3	13595	0	0	或者 或许 或是	5	415	1	0.002410	抑或 间或 甚或
302	货	huò	28	6354	3	0.000472	货币 货物 货运	55	4224	5	0.001184	期货 交货 存货
303	获	huò	12	14018	1	0.000071	获得 获悉 获胜	17	3638	3	0.000825	收获 荣获 查获
304	基	jī	31	39378	3	0.000076	基础 基本 基地	21	832	1	0.001202	根基 奠基 地基
305	积	jī	28	15180	3	0.000198	积存 积累 积分	16	6035	1	0.000166	面积 堆积 累积
306	机	jī	63	44185	3	0.000068	机会 机构 机关	88	22503	8	0.000356	飞机 危机 手机
307	击	jī	18	2543	0	0	击败 击中 击毙	42	14516	1	0.000069	打击 攻击 袭击

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
308	奇	qí	1	17	0	0	奇数	0	0	0	0	
309	激	jī	24	11587	3	0.000259	激动 激烈 激情	6	3092	0	0	刺激 感激 偏激
310	极	jí	20	7857	0	0	极大 极为 极其	20	13821	3	0.000217	积极 消极 两极
311	集	jí	26	20834	1	0.000048	集团 集中 集体	44	6748	2	0.000296	收集 聚集 召集
312	即	jí	14	10468	1	0.000096	即使 即将 即便	7	7534	0	0	立即 随即 当即
313	级	jí	4	931	0	0	级别 级差 级数	66	23940	1	0.000042	各级 高级 阶级
314	辑	jí	3	28	2	0.071429	辑录 辑入 辑成	6	2517	1	0.000397	编辑 专辑 剪辑
315	给	gěi	4	3893	0	0	给予 给付 给养	5	1130	0	0	供给 自给 补给
316	急	jí	31	5132	1	0.000195	急忙 急剧 急需	14	4391	0	0	着急 紧急 焦急
317	济	jì	1	21	0	0	济困	7	51455	0	0	经济 救济 不济
318	记	jì	23	50489	0	0	记者 记得 记录	38	13503	0	0	书记 忘记 登记
319	际	jì	1	67	0	0	际遇	22	39896	2	0.000050	国际 实际 之际
320	技	jì	9	25551	1	0.000039	技术 技巧 技能	17	14511	0	0	科技 竞技 杂技
321	继	jì	11	16310	0	0	继续 继承 继而	5	1347	1	0.000742	相继 承继 过继
322	计	jì	18	15387	1	0.000065	计划 计算 计较	32	20908	0	0	统计 设计 估计
323	纪	jì	11	7938	0	0	纪念 纪录 纪律	18	13848	1	0.000072	世纪 年纪 违纪
324	绩	jì	1	186	0	0	绩效	14	10166	2	0.000197	成绩 业绩 战绩
325	既	jì	4	3580	1	0.000279	既然 既定 既有	0	0	0	0	
326	季	jì	8	2413	1	0.000414	季度 季节 季军	17	5260	2	0.000380	冬季 夏季 秋季
327	家	jiā	75	33396	4	0.000120	家庭 家里 家长	95	109965	3	0.000027	国家 大家 人家
328	加	jiā	52	45809	2	0.000044	加强 加快 加入	21	37555	1	0.000027	参加 增加 更加
329	佳	jiā	16	1567	0	0	佳节 佳人 佳作	4	2067	0	0	最佳 欠佳 上佳
330	假	jiǎ	36	4656	0	0	假如 假设 假装	16	1254	2	0.001595	虚假 打假 制假
331	价	jià	9	13979	0	0	价格 价值 价钱	97	13003	3	0.000231	评价 代价 物价

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
332	假	jià	3	1081	0	0	假日 假期 假条	20	2388	1	0.000419	度假 暑假 放假
333	间	jiān	3	60	1	0.016667	间距 间架 间脑	41	66844	1	0.000015	时间 中间 期间
334	坚	jiān	21	19803	0	0	坚持 坚决 坚定	2	577	0	0	攻坚 中坚
335	监	jiān	22	11499	2	0.000174	监督 监狱 监测	7	460	0	0	总监 探监 学监
336	艰	jiān	8	4028	0	0	艰难 艰苦 艰巨	1	6	0	0	时艰
337	肩	jiān	7	2663	0	0	肩膀 肩上 肩负	8	654	2	0.003058	并肩 双肩 披肩
338	检	jiǎn	15	9994	4	0.000400	检查 检验 检察	8	284	1	0.003521	体检 商检 抽检
339	减	jiǎn	38	8419	3	0.000356	减少 减轻 减弱	17	1174	1	0.000852	削减 裁减 锐减
340	建	jiàn	31	65251	2	0.000031	建设 建立 建议	36	10973	4	0.000365	组建 创建 兴建
341	间	jiàn	12	1578	0	0	间接 间谍 间断	2	148	0	0	相间 离间
342	件	jiàn	1	25	0	0	件数	41	29836	4	0.000134	条件 事件 文件
343	见	jiàn	22	9446	1	0.000106	见到 见面 见识	67	37261	7	0.000188	意见 看见 会见
344	健	jiàn	16	10048	1	0.000100	健康 健全 健身	8	1652	0	0	保健 稳健 强健
345	监	jiàn	0	0	0	0		1	937	0	0	太监
346	键	jiàn	5	277	1	0.003610	键盘 键位 键板	4	5399	1	0.000185	关键 按键 琴键
347	践	jiàn	3	252	0	0	践踏 践诺 践约	2	5063	0	0	实践 作践
348	渐	jiàn	3	169	0	0	渐进 渐次 渐变	2	3180	0	0	逐渐 日渐
349	荐	jiàn	1	4	0	0	荐举	6	1458	0	0	推荐 举荐 自荐
350	讲	jiǎng	27	10972	1	0.000091	讲话 讲究 讲述	9	3336	0	0	演讲 来讲 开讲
351	强	jiàng	0	0	0	0		1	198	0	0	倔强
352	降	jiàng	22	4377	1	0.000228	降低 降临 降落	12	3860	1	0.000259	下降 空降 起降
353	将	jiàng	8	1058	0	0	将领 将士 将门	31	3819	1	0.000262	名将 上将 中将
354	交	jiāo	72	32760	6	0.000183	交流 交通 交易	37	9029	3	0.000332	外交 建交 提交
355	角	jiǎo	12	3693	1	0.000271	角度 角落 角球	50	4506	5	0.001110	主角 嘴角 三角

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
356	觉	jiào	0	0	0	0		3	2088	0	0	睡觉 午觉 懒觉
357	教	jiào	45	45311	3	0.000066	教育 教授 教师	44	7776	2	0.000257	宗教 邪教 请教
358	校	jiào	10	237	2	0.008439	校正 校对 校验	2	110	0	0	本校 审校
359	叫	jiào	21	3658	1	0.000273	叫做 叫声 叫好	15	2950	0	0	名叫 尖叫 惨叫
360	轿	jiào	2	1289	0	0	轿车 轿夫	2	92	1	0.010870	花轿 驮轿
361	结	jiē	2	11951	0	0	结果 结实	0	0	0	0	
362	接	jiē	49	23696	3	0.000127	接受 接触 接近	19	13033	0	0	直接 迎接 连接
363	阶	jiē	4	11194	0	0	阶段 阶层 阶级	5	1308	0	0	台阶 石阶 音阶
364	街	jiē	18	5261	1	0.000190	街道 街上 街头	11	2167	1	0.000461	大街 上街 临街
365	结	jié	49	45765	3	0.000066	结果 结构 结束	33	10040	3	0.000299	团结 总结 集结
366	解	jiě	48	29754	3	0.000101	解决 解释 解放	41	24475	1	0.000041	了解 理解 缓解
367	姐	jiě	3	1355	0	0	姐妹 姐夫 姐弟	7	8249	1	0.000121	小姐 大姐 表姐
368	解	jiè	1	7	0	0	解送	1	68	0	0	押解
369	介	jiè	9	12235	0	0	介绍 介入 介意	7	2209	1	0.000453	媒介 简介 中介
370	届	jiè	3	652	1	0.001534	届时 届满 届期	6	3667	0	0	本届 上届 历届
371	界	jiè	13	1291	3	0.002324	界限 界定 界线	37	45418	5	0.000110	世界 各界 境界
372	今	jīn	18	60005	0	0	今天 今年 今后	9	10273	0	0	如今 至今 当今
373	金	jīn	60	16208	7	0.000432	金融 金牌 金额	54	17428	4	0.000230	资金 基金 黄金
374	斤	jīn	1	40	0	0	斤两	3	3747	0	0	公斤 千斤 市斤
375	禁	jīn	0	0	0	0		1	1331	0	0	不禁
376	紧	jǐn	16	8912	0	0	紧张 紧急 紧密	13	4568	2	0.000438	赶紧 抓紧 加紧
377	进	jìn	49	79580	3	0.000038	进行 进入 进步	46	32943	3	0.000091	促进 先进 推进
378	近	jìn	30	6627	2	0.000302	近日 近期 近来	24	16512	3	0.000182	最近 附近 接近
379	禁	jìn	27	3727	1	0.000268	禁止 禁毒 禁区	16	1007	2	0.001986	严禁 监禁 囚禁

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
380	经	jīng	40	97456	1	0.000010	经济 经过 经营	35	46907	2	0.000043	已经 曾经 神经
381	睛	jīng	0	0	0	0		2	9737	0	0	眼睛 定睛
382	景	jǐng	13	3714	1	0.000269	景象 景观 景点	42	9188	4	0.000435	背景 前景 情景
383	警	jǐng	46	9547	2	0.000209	警察 警方 警告	30	4392	2	0.000455	武警 民警 干警
384	径	jìng	5	418	0	0	径直 径自 径流	17	4610	0	0	途径 田径 行径
385	净	jìng	16	805	2	0.002484	净化 净增 净值	12	2260	1	0.000442	干净 洁净 纯净
386	竞	jìng	15	9414	1	0.000106	竞争 竞赛 竞选	0	0	0	0	
387	竟	jìng	4	3232	0	0	竟然 竟是 竟敢	4	5935	1	0.000168	毕竟 究竟 未竟
388	静	jìng	19	1036	2	0.001931	静止 静默 静谧	18	7480	0	0	平静 冷静 安静
389	境	jìng	7	4387	0	0	境内 境界 境外	35	20696	1	0.000048	环境 边境 困境
390	镜	jìng	9	1470	1	0.000680	镜头 镜框 镜片	18	1407	5	0.003554	眼镜 墨镜 明镜
391	究	jiū	1	2731	0	0	究竟	10	24187	0	0	研究 讲究 追究
392	久	jiǔ	8	589	0	0	久远 久违 久仰	14	7874	0	0	不久 多久 悠久
393	酒	jiǔ	56	5151	5	0.000971	酒店 酒杯 酒吧	51	5454	3	0.000550	喝酒 啤酒 绍酒
394	就	jiù	27	21200	1	0.000047	就是 就业 就要	13	7290	0	0	成就 早就 造就
395	居	jū	13	10512	0	0	居民 居家 居住	47	4525	4	0.000884	邻居 位居 定居
396	局	jú	12	8431	2	0.000237	局面 局长 局势	39	11616	4	0.000344	当局 格局 大局
397	举	jǔ	23	24696	3	0.000121	举行 举办 举措	23	6916	0	0	选举 高举 推举
398	巨	jù	36	9097	5	0.000550	巨大 巨额 巨人	1	700	0	0	艰巨
399	拒	jù	7	3761	1	0.000266	拒绝 拒收 拒付	2	475	0	0	抗拒 婉拒
400	据	jù	8	4610	0	0	据悉 据说 据点	20	17526	3	0.000171	根据 数据 依据
401	距	jù	1	2653	0	0	距离	12	2477	0	0	差距 相距 间距
402	觉	jué	4	23952	0	0	觉得 觉悟 觉察	19	15771	1	0.000063	感觉 自觉 发觉
403	决	jué	19	27147	0	0	决定 决策 决赛	17	23293	0	0	解决 坚决 判决

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
404	掘	jué	2	45	0	0	掘进 掘开	5	2352	0	0	发掘 挖掘 开掘
405	军	jūn	85	25237	2	0.000079	军队 军事 军人	57	30597	2	0.000065	日军 冠军 将军
406	均	jūn	6	941	0	0	均匀 均衡 均等	5	4157	0	0	平均 不均 人均
407	开	kāi	123	86883	7	0.000081	开始 开发 开展	59	33833	2	0.000059	离开 召开 打开
408	看	kān	7	555	0	0	看守 看管 看门	0	0	0	0	
409	看	kàn	27	38943	0	0	看到 看见 看来	18	6243	1	0.000160	来看 观看 眼看
410	康	kāng	4	544	0	0	康复 康乐 康健	3	7345	0	0	健康 小康 安康
411	抗	kàng	27	7888	4	0.000507	抗战 抗日 抗议	9	2844	0	0	抵抗 对抗 反抗
412	靠	kào	10	1222	0	0	靠近 靠拢 靠山	9	3873	1	0.000258	依靠 可靠 投靠
413	科	kē	15	28826	1	0.000035	科技 科学 科研	26	4046	0	0	学科 外科 本科
414	渴	kě	4	1762	0	0	渴望 渴求 渴盼	7	322	0	0	饥渴 口渴 解渴
415	客	kè	34	15224	0	0	客户 客人 客观	57	8974	3	0.000334	顾客 游客 旅客
416	课	kè	16	4456	0	0	课题 课程 课堂	25	3126	1	0.000320	上课 功课 讲课
417	肯	kěn	3	7292	0	0	肯定 肯干 肯切	4	1133	0	0	不肯 宁肯 中肯
418	空	kōng	55	13239	6	0.000453	空间 空气 空军	40	7860	1	0.000127	航空 天空 上空
419	恐	kǒng	6	5949	0	0	恐怕 恐怖 恐惧	5	617	0	0	惊恐 惶恐 唯恐
420	空	kòng	15	1586	0	0	空白 空地 空闲	5	652	0	0	有空 没空 抽空
421	控	kòng	7	7709	2	0.000259	控制 控股 控诉	20	3184	5	0.001570	调控 指控 监控
422	苦	kǔ	50	4160	4	0.000962	苦笑 苦难 苦恼	27	8490	0	0	痛苦 辛苦 艰苦
423	会	kuài	1	1582	0	0	会计	0	0	0	0	
424	快	kuài	32	9500	3	0.000316	快乐 快速 快餐	21	17256	0	0	加快 很快 尽快
425	款	kuǎn	9	715	3	0.004196	款项 款待 款式	56	12407	6	0.000484	贷款 存款 捐款
426	况	kuàng	1	636	0	0	况且	18	28430	0	0	情况 状况 何况
427	困	kùn	13	9824	0	0	困难 困境 困扰	12	5166	0	0	贫困 特困 脱困

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
428	括	kuò	2	58	0	0	括号 括弧	7	12618	0	0	包括 概括 囊括
429	扩	kuò	14	10332	2	0.000194	扩大 扩张 扩展	0	0	0	0	
430	阔	kuò	8	294	0	0	阔别 阔步 阔气	10	2500	0	0	广阔 开阔 宽阔
431	拉	lā	27	2820	4	0.001418	拉开 拉动 拉扯	2	82	0	0	扒拉 拖拉
432	览	lǎn	1	1	1	1.000000	览胜	12	2493	1	0.000401	展览 游览 浏览
433	劳	láo	24	7611	1	0.000131	劳动 劳务 劳工	17	1856	0	0	疲劳 功劳 勤劳
434	老	lǎo	70	12390	4	0.000323	老人 老头 老家	18	3742	0	0	古老 养老 衰老
435	落	lào	2	26	0	0	落枕 落价	0	0	0	0	
436	乐	lè	9	2825	0	0	乐观 乐趣 乐意	19	6821	1	0.000147	快乐 娱乐 欢乐
437	累	léi	1	81	0	0	累赘	0	0	0	0	
438	累	lěi	7	1892	2	0.001057	累计 累积 累加	6	1901	0	0	积累 拖累 连累
439	类	lèi	7	3762	0	0	类似 类型 类别	40	17137	2	0.000117	人类 各类 分类
440	泪	lèi	9	2108	0	0	泪水 泪珠 泪眼	11	3757	0	0	眼泪 流泪 含泪
441	累	lèi	4	208	0	0	累死 累人 累活	2	270	0	0	劳累 受累
442	冷	lěng	57	7352	3	0.000408	冷静 冷笑 冷淡	13	1403	2	0.001426	寒冷 冰冷 阴冷
443	离	lí	26	11632	2	0.000172	离开 离婚 离去	18	7455	0	0	距离 脱离 分离
444	里	lǐ	12	6271	1	0.000159	里面 里边 里程	56	75342	5	0.000066	这里 心里 那里
445	历	lì	18	23443	3	0.000128	历史 历程 历来	24	7389	0	0	经历 学历 农历
446	力	lì	23	14540	2	0.000138	力量 力度 力气	121	75740	5	0.000066	努力 能力 大力
447	丽	lì	2	50	0	0	丽人 丽日	21	5669	5	0.000882	美丽 华丽 秀丽
448	例	lì	10	3789	1	0.000264	例如 例外 例证	36	7774	3	0.000386	比例 条例 惯例
449	厉	lì	5	2187	1	0.000457	厉害 厉声 厉行	4	1625	1	0.000615	严厉 凌厉 凄厉
450	联	lián	49	23298	10	0.000429	联合 联系 联盟	17	1453	0	0	关联 对联 蝉联
451	连	lián	42	10195	5	0.000490	连续 连忙 连接	13	1793	1	0.000558	接连 相连 牵连

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
452	怜	lián	5	527	0	0	怜悯 怜惜 怜爱	4	2068	0	0	可怜 爱怜 乞怜
453	莲	lián	7	304	1	0.003289	莲子 莲花 莲藕	10	1506	2	0.001328	马莲 雪莲 金莲
454	脸	liǎn	18	10684	0	0	脸上 脸色 脸红	24	3486	1	0.000287	满脸 笑脸 红脸
455	练	liàn	11	1950	0	0	练习 练功 练兵	19	8295	0	0	训练 教练 熟练
456	恋	liàn	9	2353	1	0.000425	恋爱 恋人 恋情	12	1627	0	0	迷恋 初恋 失恋
457	炼	liàn	9	398	0	0	炼钢 炼油 炼铁	9	2406	0	0	锻炼 修炼 提炼
458	量	liáng	4	36	0	0	量度 量具 量器	11	4822	0	0	打量 衡量 测量
459	良	liáng	23	8465	2	0.000236	良好 良心 良性	9	4240	0	0	不良 善良 优良
460	粮	liáng	21	5501	0	0	粮食 粮棉 粮库	27	1097	1	0.000912	口粮 种粮 干粮
461	亮	liàng	10	903	1	0.001107	亮相 亮丽 亮点	26	7174	0	0	发亮 明亮 月亮
462	疗	liáo	4	651	0	0	疗效 疗法 疗养	12	5771	1	0.000173	医疗 治疗 化疗
463	了	liǎo	8	12123	0	0	了解 了事 了结	7	313	0	0	末了 终了 未了
464	列	liè	15	4506	1	0.000222	列车 列入 列为	28	9884	2	0.000202	系列 名列 行列
465	烈	liè	11	1097	0	0	烈士 烈日 烈火	17	11303	0	0	强烈 激烈 热烈
466	裂	liè	11	580	1	0.001724	裂缝 裂开 裂痕	24	3038	3	0.000987	分裂 破裂 断裂
467	临	lín	26	4341	1	0.000230	临时 临床 临近	12	5729	3	0.000524	面临 来临 降临
468	林	lín	21	2953	1	0.000339	林业 林木 林立	28	5810	0	0	森林 山林 造林
469	龄	líng	0	0	0	0		22	3837	3	0.000782	年龄 高龄 老龄
470	灵	líng	26	4403	4	0.000908	灵魂 灵活 灵感	22	3345	4	0.001196	心灵 神灵 精灵
471	领	lǐng	38	34731	1	0.000029	领导 领域 领土	22	8418	0	0	占领 带领 率领
472	另	lìng	5	5800	1	0.000172	另外 另行 另册	0	0	0	0	
473	流	liú	57	16476	3	0.000182	流动 流行 流域	75	16392	3	0.000183	交流 分流 潮流
474	留	liú	34	8513	1	0.000117	留言 留守 留下	24	4269	3	0.000703	保留 停留 遗留
475	露	lòu	7	2717	0	0	露出 露面 露脸	1	215	0	0	泄露

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
476	路	lù	35	9264	3	0.000324	路线 路上 路边	103	23805	8	0.000336	铁路 道路 公路
477	录	lù	8	1789	1	0.000559	录取 录音 录像	26	6350	1	0.000157	纪录 记录 目录
478	露	lù	8	600	0	0	露天 露水 露骨	27	5608	6	0.001070	透露 暴露 流露
479	履	lǚ	4	1888	0	0	履行 履约 履带	1	139	0	0	步履
480	律	lǜ	5	1901	0	0	律师 律动 律己	19	16656	0	0	法律 规律 纪律
481	虑	lù	0	0	0	0		9	8041	0	0	考虑 忧虑 焦虑
482	绿	lǜ	22	4548	3	0.000660	绿色 绿化 绿地	20	452	1	0.002212	碧绿 翠绿 嫩绿
483	乱	luàn	17	738	2	0.002710	乱跑 乱来 乱叫	41	5268	1	0.000190	混乱 胡乱 慌乱
484	论	lùn	21	6056	1	0.000165	论坛 论文 论述	41	32655	2	0.000061	理论 讨论 无论
485	络	luò	0	0	0	0		10	7954	1	0.000126	网络 联络 脉络
486	落	luò	36	10301	4	0.000388	落实 落后 落地	59	7765	3	0.000386	滑落 回落 堕落
487	马	mǎ	49	7160	7	0.000978	马上 马背 马路	49	3233	2	0.000619	人马 上马 骑马
488	卖	mài	24	1721	2	0.001162	卖方 卖淫 卖掉	25	3929	0	0	买卖 拍卖 出卖
489	买	mǎi	16	2559	1	0.000391	买卖 买方 买房	5	2635	0	0	购买 收买 竞买
490	满	mǎn	31	10935	3	0.000274	满足 满意 满脸	20	9093	1	0.000110	充满 不满 圆满
491	漫	màn	11	1809	1	0.000553	漫长 漫画 漫步	5	1964	0	0	浪漫 弥漫 散漫
492	忙	máng	5	1075	0	0	忙碌 忙于 忙活	11	5691	1	0.000176	连忙 急忙 帮忙
493	矛	máo	2	4744	0	0	矛盾 矛头	2	56	1	0.017857	长矛 戈矛
494	毛	máo	37	3240	5	0.001543	毛病 毛巾 毛衣	30	1730	3	0.001734	眉毛 羽毛 羊毛
495	媒	méi	5	3820	0	0	媒体 媒介 媒婆	4	754	0	0	传媒 做媒 说媒
496	美	měi	50	28976	4	0.000138	美满 美丽 美好	31	5690	1	0.000176	完美 优美 审美
497	魅	mèi	1	1333	0	0	魅力	1	53	0	0	鬼魅
498	门	mén	57	10125	6	0.000593	门口 门外 门前	111	33925	14	0.000413	部门 专门 大门
499	盟	méng	10	1467	0	0	盟军 盟国 盟友	8	2971	1	0.000337	联盟 同盟 加盟

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
500	梦	mèng	12	3484	1	0.000287	梦想 梦见 梦境	13	1278	2	0.001565	做梦 噩梦 恶梦
501	密	mì	26	4591	0	0	密切 密集 密码	26	6789	1	0.000147	秘密 紧密 亲密
502	秘	mì	10	4545	0	0	秘密 秘书 秘诀	8	2233	1	0.000448	神秘 奥秘 隐秘
503	免	miǎn	23	2888	3	0.001039	免费 免税 免疫	11	6048	0	0	避免 难免 未免
504	面	miàn	59	31736	3	0.000095	面前 面对 面积	138	81348	5	0.000061	方面 全面 里面
505	灭	miè	8	800	0	0	灭亡 灭绝 灭火	17	3384	0	0	消灭 毁灭 歼灭
506	民	mín	47	29864	2	0.000067	民族 民主 民间	53	64903	2	0.000031	人民 农民 居民
507	敏	mǐn	4	2086	1	0.000479	敏感 敏锐 敏捷	7	544	1	0.001838	过敏 灵敏 机敏
508	明	míng	54	32815	4	0.000122	明白 明显 明确	52	38130	2	0.000052	文明 说明 表明
509	名	míng	74	19800	2	0.000101	名字 名称 名叫	107	19304	7	0.000363	著名 排名 知名
510	模	mó	10	7436	0	0	模式 模仿 模范	7	6135	0	0	规模 劳模 楷模
511	默	mò	12	869	1	0.001151	默契 默认 默然	4	3350	0	0	沉默 幽默 静默
512	漠	mò	2	205	0	0	漠然 漠视	6	2558	0	0	沙漠 冷漠 大漠
513	模	mú	5	1445	1	0.000692	模样 模具 模板	6	17	2	0.117647	字模 铸模 土模
514	母	mǔ	26	10651	5	0.000469	母亲 母女 母子	39	8876	3	0.000338	父母 字母 师母
515	目	mù	15	40088	1	0.000025	目前 目标 目的	52	29060	2	0.000069	项目 节目 题目
516	幕	mù	5	418	0	0	幕后 幕僚 幕布	27	5283	2	0.000379	开幕 屏幕 闭幕
517	男	nán	25	22688	0	0	男人 男子 男孩	2	23	0	0	少男 童男
518	难	nán	33	16548	2	0.000121	难以 难道 难得	15	9984	1	0.000100	困难 艰难 不难
519	难	nàn	3	863	1	0.001159	难民 难友 难侨	30	3765	3	0.000797	灾难 苦难 遇难
520	脑	nǎo	18	3648	1	0.000274	脑袋 脑海 脑筋	16	9433	2	0.000212	电脑 首脑 头脑
521	闹	nào	10	854	1	0.001171	闹事 闹剧 闹市	9	2510	0	0	热闹 喧闹 胡闹
522	内	nèi	83	25720	11	0.000428	内容 内部 内地	39	19989	3	0.000150	国内 在内 境内
523	念	niàn	8	625	0	0	念书 念经 念佛	32	14899	2	0.000134	观念 纪念 概念

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
524	娘	niáng	4	772	0	0	娘家 娘子 娘舅	18	6436	1	0.000155	姑娘 大娘 新娘
525	农	nóng	38	40050	2	0.000050	农业 农民 农村	30	1484	2	0.001348	贫农 工农 富农
526	女	nǚ	49	42345	3	0.000071	女人 女儿 女子	50	21796	6	0.000275	妇女 美女 子女
527	暖	nuǎn	13	764	1	0.001309	暖和 暖气 暖意	9	2239	0	0	温暖 取暖 保暖
528	怕	pà	4	232	0	0	怕人 怕事 怕生	9	10779	0	0	恐怕 害怕 可怕
529	拍	pāi	13	3203	0	0	拍摄 拍卖 拍照	10	468	1	0.002137	球拍 节拍 合拍
530	牌	pái	13	461	2	0.004338	牌照 牌坊 牌楼	46	10471	4	0.000382	金牌 品牌 名牌
531	判	pàn	13	3941	0	0	判断 判决 判处	10	8854	0	0	谈判 批判 审判
532	旁	páng	10	3618	0	0	旁边 旁人 旁听	7	2134	0	0	一旁 身旁 两旁
533	培	péi	5	11145	0	0	培养 培训 培育	2	457	0	0	栽培 代培
534	配	pèi	40	7542	2	0.000265	配合 配套 配置	18	4781	2	0.000418	分配 支配 搭配
535	朋	péng	2	12360	0	0	朋友 朋党	3	103	0	0	亲朋 宾朋 友朋
536	碰	pèng	10	3268	1	0.000306	碰到 碰上 碰见	2	51	0	0	相碰 磕碰
537	批	pī	25	11857	2	0.000169	批评 批准 批判	12	2854	0	0	朱批 审批 特批
538	皮	pí	42	3879	0	0	皮肤 皮鞋 皮包	58	3107	3	0.000966	眼皮 肚皮 调皮
539	脾	pí	4	1548	0	0	脾气 脾胃 脾性	0	0	0	0	
540	否	pǐ	0	0	0	0		1	8	0	0	臧否
541	屁	pì	2	1514	0	0	屁股 屁话	3	293	1	0.003413	狗屁 放屁 马屁
542	片	piàn	13	1730	0	0	片刻 片面 片段	67	12273	6	0.000489	照片 图片 影片
543	漂	piāo	4	446	0	0	漂流 漂浮 漂泊	2	55	0	0	水漂 鱼漂
544	漂	piǎo	2	46	0	0	漂洗 漂白	0	0	0	0	
545	票	piào	18	1024	3	0.002930	票据 票房 票价	56	8023	3	0.000374	股票 邮票 投票
546	漂	piào	1	3575	0	0	漂亮	0	0	0	0	
547	拼	pīn	16	2633	0	0	拼命 拼搏 拼死	3	38	1	0.026316	硬拼 力拼 比拼

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
548	贫	pín	17	6425	0	0	贫困 贫穷 贫富	8	2572	2	0.000778	扶贫 脱贫 清贫
549	品	pǐn	24	7762	0	0	品种 品牌 品质	63	39842	1	0.000025	产品 商品 作品
550	平	píng	80	26589	8	0.000301	平均 平等 平衡	28	28027	2	0.000071	水平 和平 公平
551	评	píng	34	11227	1	0.000089	评价 评论 评估	26	5838	4	0.000685	批评 好评 点评
552	破	pò	28	7509	0	0	破坏 破产 破裂	19	6115	1	0.000164	突破 打破 冲破
553	普	pǔ	13	9242	4	0.000433	普遍 普通 普及	1	621	0	0	科普
554	期	qī	16	12030	1	0.000083	期间 期待 期望	70	33945	3	0.000088	时期 长期 同期
555	妻	qī	7	4461	2	0.000448	妻子 妻妾 妻女	6	2174	1	0.000460	夫妻 前妻 爱妻
556	奇	qí	32	6033	1	0.000166	奇怪 奇迹 奇特	11	4014	0	0	好奇 神奇 惊奇
557	启	qǐ	17	5903	2	0.000339	启动 启明 启示	2	260	0	0	开启 谨启
558	汽	qì	13	8564	0	0	汽车 汽油 汽水	2	123	0	0	蒸汽 水汽
559	弃	qì	5	164	0	0	弃权 弃置 弃儿	17	4919	1	0.000203	放弃 抛弃 摒弃
560	气	qì	65	13931	4	0.000287	气氛 气候 气质	164	30608	5	0.000163	生气 口气 天气
561	器	qì	12	2034	2	0.000983	器材 器官 器械	38	7768	1	0.000129	武器 机器 仪器
562	签	qiān	13	8001	0	0	签署 签订 签字	7	836	0	0	标签 抽签 草签
563	前	qián	76	24830	3	0.000121	前面 前进 前景	40	60436	0	0	目前 面前 以前
564	潜	qián	16	3685	3	0.000814	潜力 潜艇 潜在	2	106	0	0	反潜 挖潜
565	强	qiáng	48	22466	3	0.000134	强调 强烈 强大	20	24063	2	0.000083	加强 增强 坚强
566	强	qiǎng	5	813	0	0	强迫 强求 强逼	2	1012	0	0	勉强 牵强
567	侨	qiáo	15	1286	1	0.000778	侨胞 侨乡 侨务	4	1734	1	0.000577	华侨 归侨 外侨
568	巧	qiǎo	11	925	0	0	巧妙 巧合 巧遇	23	2811	0	0	技巧 恰巧 乖巧
569	切	qiē	18	1330	1	0.000752	切断 切割 切磋	0	0	0	0	
570	切	qiè	9	3333	0	0	切实 切身 切忌	28	20651	4	0.000194	一切 密切 亲切
571	亲	qīn	40	14063	1	0.000071	亲自 亲切 亲戚	41	20158	4	0.000198	父亲 母亲 乡亲

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
572	侵	qīn	13	5536	0	0	侵略 侵犯 侵华	1	526	0	0	入侵
573	清	qīng	89	22633	4	0.000177	清楚 清醒 清理	21	2609	1	0.000383	认清 弄清 澄清
574	轻	qīng	39	7504	0	0	轻松 轻易 轻声	3	6710	0	0	年轻 减轻 看轻
575	倾	qīng	16	3778	2	0.000529	倾向 倾听 倾斜	2	297	0	0	右倾 左倾
576	情	qíng	52	37996	5	0.000132	情况 情绪 情感	117	49675	8	0.000161	事情 感情 爱情
577	请	qǐng	19	4394	0	0	请求 请问 请示	18	7207	0	0	邀请 申请 聘请
578	庆	qìng	5	2556	0	0	庆祝 庆典 庆幸	10	1100	1	0.000909	国庆 喜庆 欢庆
579	求	qiú	35	2813	1	0.000355	求职 求助 求得	26	34118	1	0.000029	要求 追求 需求
580	球	qiú	36	5767	5	0.000867	球员 球队 球迷	65	16310	5	0.000307	全球 足球 地球
581	区	qū	19	6018	3	0.000499	区域 区别 区分	83	45958	10	0.000218	地区 社区 山区
582	屈	qū	9	768	1	0.001302	屈辱 屈服 屈膝	6	1806	1	0.000554	委屈 不屈 冤屈
583	渠	qú	2	1970	1	0.000508	渠道 渠网	9	179	2	0.011173	水渠 沟渠 明渠
584	取	qǔ	32	18803	3	0.000160	取得 取消 取代	49	21051	0	0	采取 争取 听取
585	趣	qù	4	605	0	0	趣味 趣事 趣闻	24	7454	2	0.000268	兴趣 有趣 乐趣
586	全	quán	81	80604	2	0.000025	全国 全面 全部	13	22871	0	0	完全 安全 健全
587	权	quán	20	11657	2	0.000172	权利 权力 权益	50	15094	2	0.000133	人权 主权 产权
588	券	quàn	2	39	0	0	券商 券种	8	2290	1	0.000437	证券 债券 胜券
589	缺	quē	22	7301	0	0	缺乏 缺少 缺点	13	1286	1	0.000778	短缺 欠缺 残缺
590	确	què	11	15285	1	0.000065	确定 确实 确保	5	14362	0	0	正确 明确 准确
591	群	qún	25	17805	4	0.000225	群众 群体 群岛	20	2963	3	0.001012	人群 成群 羊群
592	然	rán	4	21100	1	0.000047	然后 然而 然则	123	88741	15	0.000169	虽然 当然 突然
593	染	rǎn	10	391	0	0	染上 染发 染指	13	4500	0	0	污染 感染 传染
594	绕	rào	7	154	1	0.006494	绕道 绕行 绕嘴	9	2698	1	0.000371	围绕 环绕 缠绕
595	热	rè	53	15313	5	0.000327	热情 热烈 热闹	42	3064	5	0.001632	亲热 狂热 炎热

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
596	人	rén	83	120492	3	0.000025	人民 人员 人家	229	133532	9	0.000067	女人 别人 男人
597	认	rèn	21	48561	0	0	认为 认识 认真	10	7793	0	0	承认 确认 否认
598	容	róng	15	9945	1	0.000101	容易 容量 容忍	39	15485	0	0	内容 笑容 形容
599	融	róng	12	2595	1	0.000385	融资 融合 融入	7	7275	1	0.000137	金融 交融 通融
600	荣	róng	13	2637	0	0	荣誉 荣获 荣幸	6	4059	1	0.000246	繁荣 光荣 虚荣
601	柔	róu	15	1895	1	0.000528	柔软 柔和 柔道	6	1782	0	0	温柔 轻柔 娇柔
602	入	rù	62	6302	1	0.000159	入学 入侵 入手	80	49058	5	0.000102	进入 收入 投入
603	软	ruǎn	32	3356	6	0.001788	软件 软弱 软化	12	1075	0	0	柔软 疲软 松软
604	润	rùn	6	125	0	0	润泽 润滑 润湿	13	2958	0	0	利润 湿润 滋润
605	赛	sài	19	4913	2	0.000407	赛事 赛场 赛车	20	22998	1	0.000043	比赛 决赛 参赛
606	色	sè	22	3008	4	0.001330	色彩 色差 色调	119	28755	6	0.000209	特色 脸色 绿色
607	森	sēn	3	2342	0	0	森林 森严 森然	1	87	0	0	阴森
608	沙	shā	44	3585	6	0.001674	沙漠 沙滩 沙化	14	1383	1	0.000723	治沙 风沙 泥沙
609	色	shǎi	1	5	0	0	色酒	0	0	0	0	
610	山	shān	95	11808	10	0.000847	山区 山林 山上	44	5482	6	0.001094	火山 登山 上山
611	善	shàn	20	4086	2	0.000489	善于 善良 善意	11	11090	0	0	改善 完善 妥善
612	伤	shāng	23	7164	0	0	伤害 伤心 伤亡	44	4947	0	0	受伤 悲伤 忧伤
613	商	shāng	50	24984	1	0.000040	商品 商业 商量	32	12219	0	0	外商 工商 协商
614	赏	shǎng	11	497	0	0	赏识 赏赐 赏脸	13	3862	2	0.000518	欣赏 赞赏 观赏
615	少	shào	13	5266	1	0.000190	少年 少女 少爷	0	0	0	0	
616	舍	shě	5	634	0	0	舍得 舍弃 舍身	3	308	0	0	施舍 取舍 割舍
617	社	shè	17	45599	0	0	社会 社区 社长	18	2482	1	0.000403	报社 公社 分社
618	设	shè	17	21818	0	0	设施 设备 设计	24	32762	1	0.000031	建设 开设 假设
619	涉	shè	12	3883	2	0.000515	涉及 涉外 涉嫌	5	1990	0	0	干涉 交涉 牵涉

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
620	摄	shè	10	2892	1	0.000346	摄影 摄入 摄像	2	1354	0	0	拍摄 统摄
621	舍	shè	6	50	1	0.020000	舍下 舍监 舍人	18	2228	3	0.001346	宿舍 校舍 农舍
622	参	shēn	1	1	1	1.000000	参茸	9	296	1	0.003378	人参 海参 丹参
623	身	shēn	38	30135	2	0.000066	身上 身体 身边	86	25181	2	0.000079	自身 本身 转身
624	深	shēn	54	20856	3	0.000144	深入 深刻 深化	11	1742	0	0	加深 纵深 资深
625	申	shēn	13	4039	1	0.000248	申请 申办 申报	2	1237	0	0	重申 引申
626	伸	shēn	11	3965	1	0.000252	伸手 伸出 伸展	1	832	0	0	延伸
627	神	shén	66	10764	5	0.000465	神秘 神情 神经	39	19043	5	0.000263	精神 眼神 留神
628	审	shěn	25	9013	1	0.000111	审议 审查 审计	26	1435	2	0.001394	评审 终审 一审
629	升	shēng	26	3299	4	0.001212	升级 升起 升高	21	5119	3	0.000586	上升 提升 回升
630	生	shēng	86	82438	0	0	生活 生产 生命	102	86546	6	0.000069	发生 先生 学生
631	声	shēng	39	13780	4	0.000290	声音 声明 声称	83	16692	10	0.000599	大声 掌声 笑声
632	胜	shèng	18	6032	1	0.000166	胜利 胜负 胜过	17	5158	1	0.000194	战胜 获胜 取胜
633	剩	shèng	5	2672	1	0.000374	剩下 剩余 剩菜	3	348	2	0.005747	过剩 吃剩 残剩
634	施	shī	16	3593	2	0.000557	施工 施行 施加	4	22976	0	0	实施 措施 设施
635	失	shī	70	17396	2	0.000115	失去 失败 失望	21	9815	0	0	损失 消失 流失
636	诗	shī	31	4322	1	0.000231	诗人 诗歌 诗词	14	893	1	0.001120	史诗 唐诗 古诗
637	时	shí	56	86028	2	0.000023	时候 时间 时代	80	81555	6	0.000074	同时 当时 小时
638	实	shí	57	64452	4	0.000062	实现 实施 实际	55	44974	4	0.000089	其实 现实 事实
639	识	shí	8	1053	0	0	识别 识字 识破	17	28665	2	0.000070	认识 知识 意识
640	食	shí	27	8293	3	0.000362	食品 食物 食用	54	8640	4	0.000463	粮食 饮食 吸食
641	石	shí	66	6586	6	0.000911	石油 石器 石块	67	3961	9	0.002272	化石 岩石 钻石
642	拾	shí	7	105	2	0.019048	拾掇 拾遗 拾荒	3	1688	1	0.000592	收拾 捡拾 撇拾
643	史	shǐ	17	1720	1	0.000581	史料 史学 史诗	29	20250	4	0.000198	历史 党史 文史

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
644	始	shǐ	8	4486	0	0	始终 始建 始祖	9	28935	1	0.000035	开始 原始 伊始
645	士	shì	8	2854	0	0	士兵 士气 士绅	49	18555	2	0.000108	人士 战士 博士
646	示	shì	6	2246	0	0	示范 示威 示意	23	29627	0	0	表示 显示 展示
647	世	shì	27	52072	2	0.000038	世界 世纪 世人	40	5682	2	0.000352	去世 逝世 问世
648	市	shì	37	37261	3	0.000081	市容 市长 市民	47	25367	4	0.000158	城市 全市 都市
649	势	shì	6	2933	1	0.000341	势力 势能 势利	59	22706	3	0.000132	优势 形势 趋势
650	事	shì	29	45771	0	0	事情 事业 事件	134	44083	7	0.000159	故事 军事 从事
651	试	shì	36	6234	3	0.000481	试验 试点 试图	17	5792	0	0	考试 尝试 测试
652	视	shì	19	4865	1	0.000206	视为 视察 视野	44	23195	1	0.000043	重视 电视 忽视
653	适	shì	17	12576	0	0	适应 适合 适当	6	2767	0	0	合适 舒适 不适
654	收	shōu	70	25054	3	0.000120	收入 收到 收费	36	8414	3	0.000357	税收 吸收 丰收
655	首	shǒu	40	18884	4	0.000212	首先 首都 首脑	28	2864	3	0.001047	为首 元首 榜首
656	售	shòu	12	769	2	0.002601	售价 售票 售出	16	8767	3	0.000342	销售 出售 零售
657	殊	shū	3	193	1	0.005181	殊荣 殊死 殊为	2	4242	0	0	特殊 悬殊
658	输	shū	17	2248	1	0.000445	输入 输出 输送	5	3895	0	0	运输 传输 灌输
659	蔬	shū	1	2040	0	0	蔬菜	1	25	0	0	菜蔬
660	舒	shū	8	2900	0	0	舒服 舒适 舒畅	1	11	0	0	宽舒
661	熟	shú	22	4410	2	0.000454	熟悉 熟人 熟练	13	3303	0	0	成熟 娴熟 早熟
662	属	shǔ	9	4175	0	0	属于 属性 属下	31	6082	2	0.000329	家属 下属 金属
663	数	shǔ	3	193	0	0	数落 数数 数说	3	103	1	0.009709	悉数 尽数 历数
664	署	shǔ	3	327	1	0.003058	署名 署长 署理	13	5414	0	0	签署 部署 行署
665	术	shù	4	345	2	0.005797	术语 术科 术式	28	38855	0	0	技术 艺术 学术
666	束	shù	1	559	0	0	束缚	8	9902	0	0	结束 约束 装束
667	述	shù	5	459	0	0	述职 述评 述说	31	10150	2	0.000197	上述 描述 阐述

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
668	数	shù	26	14128	1	0.000071	数字 数量 数据	105	15676	8	0.000510	人数 多数 少数
669	树	shù	33	4860	3	0.000617	树立 树木 树上	44	2882	6	0.002082	植树 大树 果树
670	率	shuài	5	2666	0	0	率领 率先 率真	23	5982	2	0.000334	效率 利率 汇率
671	双	shuāng	41	18380	5	0.000272	双方 双手 双边	0	0	0	0	
672	水	shuǐ	176	30968	15	0.000484	水平 水果 水利	145	19294	6	0.000311	泪水 洪水 茶水
673	税	shuì	24	4200	1	0.000238	税收 税务 税费	37	2848	3	0.001053	利税 关税 纳税
674	睡	shuì	16	3258	2	0.000614	睡觉 睡眠 睡衣	10	1069	0	0	入睡 沉睡 熟睡
675	说	shuì	0	0	0	0		1	157	0	0	游说
676	顺	shùn	31	7180	2	0.000279	顺利 顺便 顺手	16	766	2	0.002611	理顺 孝顺 温顺
677	瞬	shùn	5	1432	1	0.000698	瞬间 瞬时 瞬息	2	189	0	0	一瞬 转瞬
678	说	shuō	28	22751	2	0.000088	说话 说明 说道	45	28566	1	0.000035	来说 小说 听说
679	司	sī	8	5927	1	0.000169	司机 司法 司令	7	35593	0	0	公司 土司 上司
680	思	sī	19	26173	1	0.000038	思想 思考 思维	24	10495	2	0.000191	意思 心思 反思
681	私	sī	48	4911	3	0.000611	私人 私营 私下	14	2730	1	0.000366	走私 自私 隐私
682	死	sǐ	43	6710	2	0.000298	死亡 死去 死刑	36	3611	1	0.000277	生死 杀死 气死
683	似	sì	0	0	0	0		11	4486	0	0	类似 相似 看似
684	松	sōng	37	1840	2	0.001087	松开 松弛 松懈	20	3896	2	0.000513	轻松 放松 宽松
685	送	sòng	27	2957	4	0.001353	送终 送行 送礼	34	3922	1	0.000255	赠送 运送 输送
686	诉	sù	6	1284	0	0	诉讼 诉说 诉苦	21	15627	2	0.000128	告诉 投诉 起诉
687	速	sù	15	5352	2	0.000374	速度 速滑 速写	37	12833	2	0.000156	迅速 快速 加速
688	肃	sù	6	554	0	0	肃穆 肃反 肃清	2	2288	0	0	严肃 整肃
689	算	suàn	16	4066	1	0.000246	算术 算计 算命	37	13021	3	0.000230	打算 计算 预算
690	虽	suī	3	14132	0	0	虽然 虽说 虽则	0	0	0	0	
691	随	suí	24	11036	0	0	随便 随后 随时	5	1348	1	0.000742	伴随 跟随 追随

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
692	损	sǔn	7	5884	1	0.000170	损失 损害 损坏	16	1814	2	0.001103	亏损 受损 破损
693	索	suǒ	11	1311	0	0	索性 索赔 索取	19	5626	1	0.000178	探索 线索 思索
694	它	tā	0	0	0	0		1	3257	0	0	其它
695	拓	tà	2	40	0	0	拓片 拓本	0	0	0	0	
696	抬	tái	9	1896	2	0.001055	抬头 抬高 抬杠	1	36	0	0	哄抬
697	态	tài	2	6960	0	0	态度 态势	38	15017	0	0	状态 生态 心态
698	谈	tán	14	9779	0	0	谈判 谈话 谈论	34	10192	3	0.000294	会谈 交谈 访谈
699	坛	tán	0	0	0	0		36	4301	3	0.000698	论坛 文坛 讲坛
700	坦	tǎn	11	1754	2	0.001140	坦率 坦白 坦然	2	337	0	0	平坦 舒坦
701	探	tàn	33	6842	2	0.000292	探索 探讨 探亲	14	1478	0	0	勘探 试探 侦探
702	讨	tǎo	14	8635	1	0.000116	讨论 讨厌 讨好	8	3094	0	0	探讨 研讨 检讨
703	套	tào	32	777	4	0.005148	套话 套用 套路	28	2807	2	0.000713	配套 外套 圈套
704	特	tè	53	42724	1	0.000023	特别 特点 特色	3	2775	0	0	独特 模特 奇特
705	题	tí	16	2731	1	0.000366	题目 题材 题词	30	71098	1	0.000014	问题 标题 话题
706	提	tí	51	61299	2	0.000033	提高 提出 提供	7	2431	1	0.000411	前提 别提 手提
707	体	tǐ	42	30542	2	0.000065	体育 体制 体系	105	46327	9	0.000194	身体 具体 集体
708	天	tiān	95	18273	6	0.000328	天下 天气 天空	67	66929	7	0.000105	今天 每天 明天
709	条	tiáo	24	18097	0	0	条件 条约 条例	51	3000	3	0.001000	纸条 线条 面条
710	调	tiáo	39	13083	2	0.000153	调整 调控 调节	7	4537	0	0	协调 空调 烹调
711	挑	tiǎo	9	3966	0	0	挑战 挑起 挑衅	1	44	0	0	高挑
712	铁	tiě	79	7550	9	0.001192	铁路 铁道 铁锤	21	1960	3	0.001531	钢铁 地铁 废铁
713	听	tīng	24	14512	1	0.000069	听说 听到 听见	18	3417	1	0.000293	打听 倾听 好听
714	庭	tíng	5	412	1	0.002427	庭院 庭长 庭审	10	10077	0	0	家庭 法庭 开庭
715	停	tíng	34	6789	2	0.000295	停止 停留 停下	5	692	0	0	暂停 不停 调停

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
716	通	tōng	79	37584	7	0.000186	通过 通知 通信	31	16627	1	0.000060	交通 普通 沟通
717	童	tóng	19	1427	2	0.001402	童年 童话 童心	16	5594	1	0.000179	儿童 女童 神童
718	统	tǒng	21	17291	1	0.000058	统一 统计 统治	11	30096	0	0	总统 传统 系统
719	痛	tòng	23	5863	0	0	痛苦 痛快 痛楚	25	2795	5	0.001789	疼痛 头痛 悲痛
720	投	tóu	43	32266	2	0.000062	投资 投入 投降	4	145	0	0	空投 相投 远投
721	头	tóu	54	8690	2	0.000230	头发 头脑 头顶	174	31939	9	0.000282	回头 点头 摇头
722	透	tòu	19	4082	1	0.000245	透露 透明 透过	10	1138	2	0.001757	渗透 看透 浸透
723	突	tū	16	18212	1	0.000055	突然 突出 突破	2	3588	0	0	冲突 奔突
724	图	tú	36	6171	3	0.000486	图片 图书 图案	38	5940	3	0.000505	企图 试图 地图
725	途	tú	4	2722	0	0	途径 途中 途经	18	3998	1	0.000250	前途 长途 沿途
726	土	tǔ	69	12322	13	0.001055	土地 土司 土壤	52	7915	5	0.000632	领土 出土 国土
727	团	tuán	29	10946	3	0.000274	团结 团体 团长	32	13290	1	0.000075	集团 师团 兵团
728	推	tuī	47	25295	6	0.000237	推动 推进 推出	3	76	0	0	助推 公推 类推
729	退	tuì	42	5474	1	0.000183	退休 退出 退回	19	2409	0	0	撤退 后退 衰退
730	脱	tuō	43	3020	4	0.001325	脱离 脱贫 脱困	12	3159	1	0.000317	摆脱 解脱 挣脱
731	拓	tuò	3	1362	0	0	拓展 拓宽 拓荒	1	1453	0	0	开拓
732	完	wán	20	30158	0	0	完全 完成 完善	2	82	0	0	没完 售完
733	亡	wáng	5	286	0	0	亡国 亡灵 亡命	17	5626	2	0.000355	死亡 伤亡 身亡
734	网	wǎng	25	12104	1	0.000083	网络 网上 网站	33	3349	2	0.000597	上网 联网 电网
735	忘	wàng	9	3081	1	0.000325	忘记 忘掉 忘却	5	1105	0	0	难忘 遗忘 淡忘
736	望	wàng	8	219	0	0	望见 望族 望去	51	31858	5	0.000157	希望 愿望 渴望
737	危	wēi	15	8444	2	0.000237	危险 危机 危害	7	553	0	0	濒危 安危 垂危
738	微	wēi	33	5246	3	0.000572	微笑 微妙 微弱	13	1526	0	0	稍微 轻微 细微
739	威	wēi	16	4818	1	0.000208	威胁 威力 威严	14	2388	0	0	权威 示威 国威

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
740	为	wéi	17	8668	0	0	为主 为期 为止	49	94869	9	0.000095	认为 成为 作为
741	围	wéi	30	5154	2	0.000388	围绕 围棋 围墙	15	11188	0	0	范围 周围 包围
742	唯	wéi	7	4462	0	0	唯一 唯有 唯独	0	0	0	0	
743	违	wéi	16	6480	0	0	违法 违反 违背	2	141	1	0.007092	久违 有违
744	伟	wěi	6	4796	0	0	伟大 伟人 伟业	5	1049	0	0	宏伟 雄伟 奇伟
745	未	wèi	19	9617	0	0	未来 未能 未必	3	3371	0	0	尚未 从未 并未
746	位	wèi	9	6809	1	0.000147	位置 位于 位居	85	33037	2	0.000061	单位 地位 岗位
747	味	wèi	5	2342	0	0	味道 味精 味觉	50	7578	3	0.000396	滋味 香味 意味
748	慰	wèi	5	1129	0	0	慰问 慰藉 慰劳	7	2340	0	0	安慰 欣慰 抚慰
749	温	wēn	29	8044	2	0.000249	温暖 温柔 温度	18	2373	1	0.000421	气温 高温 体温
750	文	wén	75	61563	8	0.000130	文化 文学 文明	83	9385	7	0.000746	论文 人文 散文
751	闻	wén	5	705	0	0	闻名 闻讯 闻到	21	8199	1	0.000122	新闻 见闻 传闻
752	稳	wěn	11	11416	1	0.000088	稳定 稳步 稳妥	5	1578	0	0	平稳 沉稳 安稳
753	问	wèn	21	61928	0	0	问题 问道 问候	30	16794	1	0.000060	访问 顾问 询问
754	握	wò	7	1344	0	0	握手 握住 握紧	5	6437	1	0.000155	掌握 把握 紧握
755	污	wū	13	3762	1	0.000266	污染 污水 污秽	10	1216	2	0.001645	贪污 排污 油污
756	屋	wū	13	3112	2	0.000643	屋里 屋顶 屋内	20	2893	2	0.000691	房屋 小屋 进屋
757	无	wú	96	37617	6	0.000160	无法 无论 无数	4	2576	0	0	毫无 不无 虚无
758	午	wǔ	14	1703	2	0.001174	午饭 午餐 午夜	6	13143	0	0	下午 上午 中午
759	伍	wǔ	0	0	0	0		6	5839	0	0	队伍 入伍 落伍
760	武	wǔ	36	10297	4	0.000388	武器 武装 武警	12	1040	1	0.000962	文武 威武 比武
761	舞	wǔ	27	4602	4	0.000869	舞台 舞蹈 舞厅	14	2585	1	0.000387	鼓舞 跳舞 歌舞
762	物	wù	26	10010	1	0.000100	物质 物资 物价	72	30525	5	0.000164	人物 动物 文物
763	误	wù	23	2074	4	0.001929	误会 误解 误区	14	7055	3	0.000425	错误 失误 耽误

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
764	晤	wù	3	16	1	0.062500	晤面 晤对 晤谈	1	1664	0	0	会晤
765	希	xī	2	17499	0	0	希望 希冀	0	0	0	0	
766	析	xī	2	17	1	0.058824	析出 析疑	8	7190	1	0.000139	分析 剖析 解析
767	惜	xī	4	39	0	0	惜别 惜败 惜贷	10	4183	0	0	可惜 珍惜 不惜
768	晰	xī	0	0	0	0		2	1397	0	0	清晰 明晰
769	席	xí	10	695	2	0.002878	席位 席卷 席间	30	20216	3	0.000148	主席 出席 首席
770	习	xí	13	5586	1	0.000179	习惯 习俗 习性	21	17255	2	0.000116	学习 演习 练习
771	媳	xí	1	1439	0	0	媳妇	4	257	0	0	儿媳 弟媳 婆媳
772	袭	xí	4	1457	0	0	袭击 袭扰 袭用	15	1402	2	0.001427	空袭 侵袭 偷袭
773	喜	xǐ	29	15924	7	0.000440	喜欢 喜爱 喜悦	18	2616	1	0.000382	惊喜 可喜 欣喜
774	戏	xì	21	2175	1	0.000460	戏剧 戏曲 戏院	25	4410	2	0.000454	游戏 演戏 把戏
775	系	xì	6	14448	1	0.000069	系统 系列 系数	31	42890	5	0.000117	关系 联系 体系
776	细	xì	47	6151	5	0.000813	细节 细胞 细致	16	4507	1	0.000222	仔细 详细 精细
777	峡	xiá	2	309	0	0	峡谷 峡湾	4	1576	0	0	海峡 巫峡 山峡
778	夏	xià	20	3152	3	0.000952	夏天 夏季 夏日	14	1229	0	0	华夏 仲夏 西夏
779	先	xiān	36	29289	2	0.000068	先生 先进 先后	18	14522	1	0.000069	首先 领先 率先
780	鲜	xiān	19	4027	0	0	鲜花 鲜明 鲜血	6	2666	1	0.000375	新鲜 生鲜 海鲜
781	鲜	xiǎn	2	76	0	0	鲜见 鲜有	0	0	0	0	
782	显	xiǎn	21	14832	2	0.000135	显得 显示 显然	5	5937	0	0	明显 浅显 突显
783	现	xiàn	34	64820	1	0.000015	现在 现代 现象	22	67467	0	0	发现 出现 实现
784	限	xiàn	13	4294	1	0.000233	限制 限度 限期	18	5573	0	0	有限 无限 期限
785	线	xiàn	25	2122	6	0.002828	线路 线索 线条	107	13017	14	0.001076	路线 战线 航线
786	陷	xiàn	6	2713	0	0	陷入 陷阱 陷于	10	1298	0	0	缺陷 沦陷 诬陷
787	乡	xiāng	22	5523	1	0.000181	乡镇 乡村 乡长	29	5807	1	0.000172	家乡 城乡 故乡

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
788	降	xiáng	2	23	0	0	降服 降伏	6	1141	0	0	投降 受降 劝降
789	详	xiáng	12	1729	2	0.001157	详细 详尽 详情	4	562	0	0	安详 端详 不详
790	想	xiǎng	12	18218	0	0	想到 想起 想法	34	29952	0	0	思想 理想 梦想
791	响	xiǎng	13	2396	2	0.000835	响起 响应 响亮	21	16573	3	0.000181	影响 反响 声响
792	享	xiǎng	7	4644	0	0	享受 享有 享用	3	1177	0	0	分享 共享 安享
793	向	xiàng	9	4001	0	0	向前 向导 向上	42	20943	2	0.000095	方向 走向 面向
794	销	xiāo	14	7369	3	0.000407	销售 销毁 销量	46	5693	3	0.000527	营销 促销 撤销
795	小	xiǎo	165	45644	13	0.000285	小时 小镇 小说	28	6365	1	0.000157	微小 弱小 缩小
796	校	xiào	29	7090	2	0.000282	校长 校园 校门	40	17408	5	0.000287	学校 院校 党校
797	效	xiào	15	10409	1	0.000096	效果 效益 效率	29	11635	2	0.000172	有效 成效 高效
798	笑	xiào	22	5191	3	0.000578	笑容 笑话 笑声	43	8775	2	0.000228	微笑 玩笑 可笑
799	协	xié	14	19506	1	0.000051	协议 协会 协调	1	570	0	0	妥协
800	胁	xié	2	67	0	0	胁迫 胁从	2	2734	0	0	威胁 裹胁
801	写	xiě	17	4265	0	0	写作 写信 写道	30	4061	1	0.000246	描写 撰写 编写
802	械	xiè	1	74	0	0	械斗	8	2457	0	0	机械 器械 缴械
803	欣	xīn	8	3129	0	0	欣赏 欣慰 欣喜	1	60	0	0	欢欣
804	新	xīn	92	22097	3	0.000136	新闻 新鲜 新型	18	14096	1	0.000071	重新 创新 最新
805	信	xìn	46	24189	3	0.000124	信息 信心 信任	49	19985	6	0.000300	相信 通信 自信
806	星	xīng	22	3144	0	0	星期 星座 星球	43	6142	5	0.000814	卫星 明星 火星
807	行	xíng	52	29009	2	0.000069	行动 行为 行政	93	98222	8	0.000081	进行 举行 实行
808	形	xíng	17	30114	0	0	形成 形式 形势	49	5618	3	0.000534	情形 无形 变形
809	型	xíng	5	319	0	0	型号 型钢 型材	43	14107	5	0.000354	大型 典型 类型
810	省	xǐng	4	53	0	0	省亲 省悟 省察	3	724	0	0	反省 自省 内省
811	醒	xǐng	6	952	0	0	醒来 醒目 醒悟	11	4822	1	0.000207	提醒 清醒 惊醒

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
812	幸	xìng	10	6386	0	0	幸福 幸运 幸好	6	2845	0	0	不幸 庆幸 有幸
813	性	xìng	19	8334	2	0.000240	性格 性质 性能	97	20743	7	0.000337	女性 男性 个性
814	姓	xìng	2	851	0	0	姓名 姓氏	11	1479	0	0	百姓 贵姓 同姓
815	兄	xiōng	4	3827	0	0	兄弟 兄妹 兄长	12	816	1	0.001225	弟兄 师兄 仁兄
816	雄	xióng	29	2018	2	0.000991	雄厚 雄伟 雄风	9	3086	1	0.000324	英雄 称雄 群雄
817	休	xiū	16	4530	0	0	休息 休闲 休假	13	2183	0	0	退休 不休 离休
818	修	xiū	37	7551	4	0.000530	修改 修建 修复	26	2476	1	0.000404	维修 装修 进修
819	宿	xiù	0	0	0	0		1	20	0	0	星宿
820	秀	xiù	8	977	0	0	秀丽 秀美 秀才	10	5427	1	0.000184	优秀 新秀 清秀
821	需	xū	3	22946	0	0	需要 需求 需水	8	1725	0	0	急需 必需 内需
822	序	xù	11	916	0	0	序幕 序列 序言	15	6531	1	0.000153	秩序 程序 有序
823	绪	xù	2	14	0	0	绪论 绪言	8	4480	1	0.000223	情绪 思绪 就绪
824	续	xù	13	449	0	0	续集 续建 续签	9	25362	0	0	继续 持续 连续
825	宣	xuān	20	14926	3	0.000201	宣布 宣传 宣言	0	0	0	0	
826	选	xuǎn	51	23299	3	0.000129	选择 选手 选举	37	7720	1	0.000130	当选 竞选 评选
827	券	xuàn	0	0	0	0		1	3	0	0	拱券
828	学	xué	58	65233	2	0.000031	学生 学习 学校	77	64034	1	0.000016	大学 科学 同学
829	询	xún	1	1180	0	0	询问	7	2800	0	0	咨询 查询 质询
830	迅	xùn	6	6204	1	0.000161	迅速 迅猛 迅即	0	0	0	0	
831	训	xùn	10	5003	2	0.000400	训练 训斥 训导	13	7093	1	0.000141	培训 教训 集训
832	讯	xùn	3	225	0	0	讯息 讯问 讯号	16	4089	0	0	通讯 简讯 资讯
833	压	yā	26	6870	3	0.000437	压力 压抑 压迫	35	1973	6	0.003041	镇压 血压 积压
834	研	yán	8	23385	1	0.000043	研究 研制 研讨	3	4084	0	0	科研 调研 钻研
835	严	yán	28	20585	0	0	严重 严格 严肃	9	3016	0	0	尊严 从严 庄严

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
836	言	yán	14	2367	1	0.000422	言论 言语 言行	64	16448	2	0.000122	语言 发言 禁言
837	颜	yán	4	1926	0	0	颜色 颜料 颜面	8	371	2	0.005391	容颜 红颜 汗颜
838	沿	yán	9	3259	1	0.000307	沿海 沿途 沿线	8	829	1	0.001206	前沿 床沿 炕沿
839	眼	yǎn	52	29803	2	0.000067	眼睛 眼前 眼光	65	6578	3	0.000456	心眼 着眼 亲眼
840	验	yàn	11	927	0	0	验收 验证 验放	21	17106	2	0.000117	经验 实验 试验
841	阳	yáng	17	3808	1	0.000263	阳光 阳台 阳气	14	3893	0	0	太阳 阴阳 夕阳
842	仰	yǎng	11	745	0	0	仰慕 仰望 仰面	6	1648	0	0	信仰 瞻仰 敬仰
843	养	yǎng	30	4426	2	0.000452	养老 养成 养殖	41	11242	2	0.000178	培养 营养 修养
844	样	yàng	11	925	1	0.001081	样品 样式 样板	37	86150	3	0.000035	这样 一样 那样
845	要	yāo	3	20061	0	0	要求 要挟 要约	0	0	0	0	
846	邀	yāo	4	3339	1	0.000299	邀请 邀功 邀集	3	848	0	0	应邀 特邀 诚邀
847	摇	yáo	14	2875	1	0.000348	摇头 摇晃 摇篮	4	1016	0	0	动摇 招摇 手摇
848	药	yào	56	4382	11	0.002510	药品 药物 药材	59	5131	2	0.000390	医药 中药 农药
849	爷	yé	0	0	0	0		13	3352	1	0.000298	大爷 少爷 王爷
850	业	yè	10	8921	0	0	业务 业绩 业余	80	125537	7	0.000056	企业 农业 工业
851	医	yī	25	18851	1	0.000053	医院 医生 医疗	25	2357	0	0	中医 太医 军医
852	衣	yī	27	8538	0	0	衣服 衣裳 衣着	60	4075	7	0.001718	大衣 黑衣 上衣
853	疑	yí	17	2371	2	0.000844	疑惑 疑问 疑虑	18	6209	3	0.000483	怀疑 无疑 质疑
854	遗	yí	51	9264	3	0.000324	遗址 遗憾 遗产	4	54	0	0	拾遗 子遗 梦遗
855	移	yí	13	4608	1	0.000217	移动 移民 移植	13	2723	0	0	转移 迁移 推移
856	已	yǐ	5	37092	0	0	已经 已然 已故	5	7426	0	0	早已 而已 不已
857	意	yì	27	36225	2	0.000055	意见 意义 意思	104	47594	3	0.000063	注意 愿意 同意
858	忆	yì	5	79	0	0	忆起 忆苦 忆及	3	4163	0	0	回忆 记忆 追忆
859	谊	yì	0	0	0	0		6	2443	0	0	友谊 情谊 联谊

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
860	异	yì	35	4086	3	0.000734	异常 异性 异议	21	4324	0	0	差异 诧异 优异
861	役	yì	2	20	0	0	役使 役畜	18	2194	1	0.000456	战役 退役 服役
862	因	yīn	12	55724	0	0	因为 因此 因素	15	11078	2	0.000181	原因 基因 起因
863	音	yīn	44	5328	6	0.001126	音乐 音像 音响	68	11120	6	0.000540	声音 录音 话音
864	银	yín	43	13951	6	0.000430	银行 银牌 银花	6	666	0	0	收银 金银 白银
865	引	yǐn	40	16386	2	0.000122	引起 引进 引导	11	4726	0	0	吸引 指引 援引
866	印	yìn	25	4887	2	0.000409	印象 印刷 印制	42	1405	5	0.003559	打印 烙印 脚印
867	应	yīng	8	24241	1	0.000041	应该 应当 应有	1	312	0	0	理应
868	营	yíng	22	5954	0	0	营养 营造 营业	26	14536	1	0.000069	经营 私营 民营
869	迎	yíng	15	3837	1	0.000261	迎接 迎来 迎面	6	5176	1	0.000193	欢迎 喜迎 逢迎
870	赢	yíng	4	1612	1	0.000620	赢得 赢利 赢余	2	192	0	0	输赢 稳赢
871	影	yǐng	23	16979	3	0.000177	影响 影片 影视	45	12191	6	0.000492	电影 摄影 身影
872	应	yìng	29	6572	2	0.000304	应用 应付 应邀	21	17491	1	0.000057	适应 反应 答应
873	映	yìng	9	308	2	0.006494	映照 映出 映衬	13	5798	1	0.000172	反映 放映 掩映
874	拥	yōng	10	7923	0	0	拥有 拥护 拥抱	2	242	0	0	簇拥 蜂拥
875	永	yǒng	10	7510	1	0.000133	永远 永恒 永久	1	31	0	0	隽永
876	泳	yǒng	6	140	1	0.007143	泳装 泳衣 泳坛	9	1721	1	0.000581	游泳 蝶泳 蛙泳
877	勇	yǒng	10	3597	1	0.000278	勇气 勇敢 勇于	4	631	0	0	英勇 奋勇 骁勇
878	用	yòng	48	12833	1	0.000078	用途 用户 用人	97	62807	2	0.000032	作用 利用 使用
879	优	yōu	25	20900	0	0	优势 优秀 优惠	9	254	0	0	择优 名优 创优
880	幽	yōu	27	2246	1	0.000445	幽默 幽灵 幽静	2	44	0	0	清幽 探幽
881	油	yóu	77	4358	4	0.000918	油田 油画 油气	67	7549	12	0.001590	石油 酱油 原油
882	友	yǒu	11	9382	0	0	友好 友谊 友人	44	17265	5	0.000290	朋友 战友 女友
883	右	yòu	17	3545	3	0.000846	右派 右手 右翼	3	5585	0	0	左右 反右 极右

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
884	舆	yú	2	1894	0	0	舆论 舆情	0	0	0	0	
885	愉	yú	2	1803	0	0	愉快 愉悦	1	47	0	0	欢愉
886	语	yǔ	24	7778	0	0	语言 语气 语文	75	5286	9	0.001703	话语 言语 外语
887	予	yǔ	1	1677	0	0	予以	12	6295	2	0.000318	给予 授予 赋予
888	育	yù	10	423	0	0	育种 育龄 育苗	20	33207	0	0	教育 体育 培育
889	域	yù	1	95	0	0	域名	18	13729	3	0.000219	领域 区域 流域
890	与	yù	1	1135	0	0	与会	1	6077	0	0	参与
891	遇	yù	8	5467	1	0.000183	遇到 遇见 遇难	13	5686	1	0.000176	机遇 待遇 遭遇
892	欲	yù	3	1391	1	0.000719	欲望 欲念 欲火	14	968	1	0.001033	性欲 食欲 情欲
893	裕	yù	0	0	0	0		6	1609	1	0.000622	富裕 充裕 宽裕
894	员	yuán	4	5214	0	0	员工 员额 员外	95	47386	22	0.000464	人员 委员 成员
895	援	yuán	12	2600	2	0.000769	援助 援引 援建	13	2450	1	0.000408	支援 增援 救援
896	远	yuǎn	43	4714	4	0.000849	远处 远离 远景	19	9921	1	0.000101	永远 长远 遥远
897	院	yuàn	16	5306	0	0	院长 院士 院校	38	17078	2	0.000117	医院 学院 法院
898	愿	yuàn	3	8636	1	0.000116	愿意 愿望 愿心	21	4877	1	0.000205	自愿 意愿 祝愿
899	约	yuē	12	2415	0	0	约束 约定 约会	46	10111	2	0.000198	签约 条约 制约
900	乐	yuè	20	1707	1	0.000586	乐队 乐器 乐团	18	4187	0	0	音乐 民乐 古乐
901	阅	yuè	5	1757	0	0	阅读 阅历 阅兵	16	957	0	0	检阅 查阅 翻阅
902	跃	yuè	11	385	3	0.007792	跃居 跃入 跃进	7	2630	1	0.000380	活跃 飞跃 踊跃
903	允	yǔn	3	3073	1	0.000325	允许 允诺 允当	2	137	0	0	应允 公允
904	杂	zá	40	4138	1	0.000242	杂志 杂交 杂技	18	4368	2	0.000458	复杂 夹杂 嘈杂
905	仔	zǎi	0	0	0	0		4	123	1	0.008130	牛仔 猪仔 虎仔
906	载	zǎi	3	191	0	0	载入 载文 载明	9	2115	1	0.000473	记载 转载 下载
907	再	zài	20	10820	0	0	再次 再说 再生	2	5435	0	0	不再 一再

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
908	暂	zàn	9	2923	0	0	暂时 暂行 暂停	1	665	0	0	短暂
909	遭	zāo	9	4597	0	0	遭到 遭受 遭遇	4	271	1	0.003690	惨遭 免遭 未遭
910	早	zǎo	34	13899	3	0.000216	早点 早就 早晨	11	2332	0	0	尽早 清早 迟早
911	造	zào	29	11645	5	0.000429	造成 造就 造型	26	20795	0	0	创造 改造 制造
912	则	zé	0	0	0	0		21	16569	0	0	原则 否则 规则
913	择	zé	6	312	1	0.003205	择优 择业 择偶	2	8356	0	0	选择 抉择
914	责	zé	9	8492	0	0	责任 责备 责令	22	10066	2	0.000199	负责 职责 指责
915	增	zēng	40	37349	1	0.000027	增长 增加 增强	12	1868	0	0	新增 递增 净增
916	曾	zēng	2	30	0	0	曾祖 曾孙	0	0	0	0	
917	炸	zhá	5	82	0	0	炸鸡 炸鱼 炸糕	0	0	0	0	
918	炸	zhà	7	1182	0	0	炸弹 炸毁 炸药	2	2364	0	0	爆炸 轰炸
919	择	zhái	1	6	0	0	择菜	0	0	0	0	
920	宅	zhái	4	140	0	0	宅院 宅门 宅邸	11	1489	1	0.000672	住宅 民宅 私宅
921	展	zhǎn	24	9393	1	0.000106	展开 展示 展览	29	85818	2	0.000023	发展 开展 进展
922	占	zhàn	12	5424	1	0.000184	占领 占有 占据	10	1514	0	0	攻占 侵占 抢占
923	战	zhàn	61	34618	2	0.000058	战略 战争 战士	59	15597	2	0.000128	挑战 作战 抗战
924	站	zhàn	17	1684	2	0.001188	站住 站立 站点	23	3204	1	0.000312	网站 车站 电站
925	章	zhāng	8	665	0	0	章程 章节 章法	37	8510	3	0.000353	文章 宪章 规章
926	张	zhāng	13	2120	0	0	张嘴 张开 张大	19	10713	1	0.000093	紧张 主张 扩张
927	掌	zhǎng	15	6120	1	0.000163	掌握 掌声 掌柜	14	1606	0	0	巴掌 鼓掌 手掌
928	障	zhàng	1	1634	0	0	障碍	8	5618	1	0.000178	保障 故障 屏障
929	找	zhǎo	11	5542	1	0.000180	找到 找回 找寻	5	3305	0	0	寻找 查找 自找
930	召	zhào	6	6903	1	0.000145	召开 召集 召唤	4	935	1	0.001070	号召 感召 征召
931	照	zhào	33	9415	1	0.000106	照片 照顾 照样	36	10359	3	0.000290	按照 依照 护照

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
932	哲	zhé	3	2626	0	0	哲学 哲理 哲人	2	41	0	0	先哲 贤哲
933	者	zhě	0	0	0	0		60	70762	14	0.000198	记者 或者 学者
934	针	zhēn	17	3204	4	0.001248	针对 针灸 针线	27	4552	5	0.001098	方针 指针 打针
935	真	zhēn	40	20097	3	0.000149	真正 真实 真是	17	10073	1	0.000099	认真 天真 当真
936	阵	zhèn	15	3425	0	0	阵地 阵容 阵线	18	722	2	0.002770	上阵 对阵 方阵
937	镇	zhèn	13	1675	0	0	镇压 镇定 镇上	15	5050	1	0.000198	城镇 乡镇 小镇
938	正	zhēng	2	283	1	0.003534	正月 正旦	0	0	0	0	
939	争	zhēng	25	9231	2	0.000217	争取 争夺 争论	10	20065	0	0	战争 竞争 斗争
940	征	zhēng	32	3820	2	0.000524	征服 征求 征收	12	4754	0	0	特征 象征 长征
941	整	zhěng	41	19648	2	0.000102	整个 整体 整理	14	10551	0	0	调整 完整 平整
942	正	zhèng	90	38685	7	0.000181	正式 正确 正是	28	17987	2	0.000111	真正 反正 公正
943	证	zhèng	14	10821	0	0	证明 证券 证据	42	12275	1	0.000081	保证 论证 签证
944	知	zhī	22	56366	1	0.000018	知道 知识 知名	23	15590	0	0	不知 通知 得知
945	织	zhī	10	344	0	0	织女 织成 织造	9	24193	1	0.000041	组织 纺织 编织
946	只	zhī	2	158	0	0	只身 只字	2	379	0	0	船只 舰只
947	职	zhí	15	18191	0	0	职工 职业 职务	54	5206	3	0.000576	辞职 任职 求职
948	植	zhí	11	2293	1	0.000436	植物 植树 植被	9	2649	1	0.000378	种植 移植 培植
949	指	zhǐ	33	29540	1	0.000034	指出 指导 指挥	16	2736	1	0.000365	手指 戒指 食指
950	只	zhǐ	13	28577	0	0	只要 只能 只有	2	174	0	0	不只 仅只
951	纸	zhǐ	40	1628	4	0.002457	纸条 纸张 纸片	43	4536	5	0.001102	报纸 白纸 造纸
952	止	zhǐ	10	639	1	0.001565	止住 止境 止痛	21	12793	2	0.000156	停止 防止 禁止
953	址	zhǐ	0	0	0	0		19	3979	0	0	遗址 地址 旧址
954	治	zhì	30	9513	2	0.000210	治理 治疗 治安	23	25075	0	0	政治 自治 统治
955	至	zhì	14	11566	2	0.000173	至少 至于 至今	13	16055	1	0.000062	甚至 乃至 截至

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
956	制	zhì	38	30758	1	0.000033	制度 制定 制造	83	33102	7	0.000211	控制 体制 机制
957	置	zhì	9	1125	1	0.000889	置于 置身 置换	20	10837	3	0.000277	位置 设置 安置
958	秩	zhì	1	2652	0	0	秩序	0	0	0	0	
959	智	zhì	13	4168	2	0.000480	智慧 智力 智能	16	2055	2	0.000973	理智 明智 机智
960	中	zhōng	116	52155	12	0.000230	中央 中心 中国	42	49258	2	0.000041	其中 集中 心中
961	终	zhōng	24	11547	3	0.000260	终于 终身 终究	9	9196	0	0	最终 始终 临终
962	钟	zhōng	11	1110	0	0	钟声 钟头 钟表	20	5810	2	0.000344	分钟 秒钟 时钟
963	种	zhǒng	18	1524	3	0.001969	种类 种族 种苗	68	20832	4	0.000192	各种 多种 某种
964	中	zhòng	12	868	0	0	中标 中意 中暑	10	1427	0	0	击中 看中 选中
965	种	zhòng	12	2630	0	0	种植 种田 种地	16	969	0	0	播种 耕种 栽种
966	众	zhòng	8	3823	2	0.000523	众多 众人 众望	14	24783	0	0	群众 观众 公众
967	重	zhòng	51	55353	2	0.000036	重要 重点 重大	53	26856	1	0.000037	严重 尊重 注重
968	周	zhōu	34	11618	3	0.000258	周年 周围 周末	13	2214	1	0.000452	四周 每周 上周
969	主	zhǔ	99	77055	10	0.000130	主要 主席 主任	58	13591	5	0.000368	民主 为主 自主
970	助	zhù	24	2521	3	0.001190	助理 助手 助学	22	16173	2	0.000124	帮助 援助 协助
971	注	zhù	19	15068	0	0	注意 注重 注册	22	5668	2	0.000353	关注 专注 赌注
972	祝	zhù	10	2618	1	0.000382	祝贺 祝愿 祝福	4	1591	0	0	庆祝 预祝 恭祝
973	著	zhù	8	7665	0	0	著名 著作 著称	16	3504	1	0.000285	显著 专著 名著
974	筑	zhù	6	185	1	0.005405	筑路 筑坝 筑巢	6	5242	0	0	建筑 构筑 修筑
975	抓	zhuā	16	8103	0	0	抓住 抓好 抓紧	4	485	1	0.002062	狠抓 主抓 力抓
976	转	zhuǎn	70	19296	4	0.000207	转变 转身 转移	20	3156	0	0	好转 运转 旋转
977	传	zhuàn	3	418	0	0	传记 传略 传主	10	344	0	0	列传 自传 外传
978	转	zhuàn	15	696	0	0	转动 转悠 转圈	5	131	0	0	倒转 自转 公转
979	装	zhuāng	34	6663	1	0.000150	装备 装置 装饰	65	11422	1	0.000088	武装 服装 包装

序列	单字	注音	前字					后字				
			型频	例频	孤词型频	能产指数	例子	类频	例频	孤词类频	能产指数	例子
980	追	zhuī	37	8429	6	0.000712	追求 追究 追问	4	105	0	0	围追 猛追 穷追
981	桌	zhuō	9	2763	0	0	桌上 桌面 桌前	17	1444	2	0.001385	餐桌 饭桌 书桌
982	著	zhuó	0	0	0	0		1	267	0	0	执著
983	咨	zī	2	2097	0	0	咨询 咨文	0	0	0	0	
984	资	zī	18	33580	1	0.000030	资源 资金 资料	39	32470	3	0.000092	投资 外资 工资
985	字	zì	33	2463	1	0.000406	字样 字眼 字母	43	14000	4	0.000286	名字 文字 数字
986	宗	zōng	10	4957	0	0	宗教 宗旨 宗族	8	901	2	0.002220	祖宗 大宗 正宗
987	综	zōng	5	5248	0	0	综合 综述 综艺	1	22	0	0	错综
988	总	zǒng	59	48283	0	0	总统 总理 总裁	5	353	0	0	汇总 副总 一总
989	走	zǒu	34	9987	1	0.000100	走向 走私 走过	27	3764	2	0.000531	行走 带走 临走
990	祖	zǔ	20	8107	1	0.000123	祖国 祖先 祖宗	14	562	0	0	太祖 高祖 祭祖
991	组	zǔ	22	31332	3	0.000096	组织 组成 组建	20	7089	2	0.000282	小组 重组 改组
992	族	zú	8	290	0	0	族人 族群 族长	23	16831	2	0.000119	民族 各族 家族
993	嘴	zuǐ	8	4228	0	0	嘴里 嘴唇 嘴角	38	1729	2	0.001157	张嘴 快嘴 插嘴
994	最	zuì	12	38791	0	0	最后 最近 最高	0	0	0	0	
995	罪	zuì	12	3077	0	0	罪犯 罪行 罪恶	31	5826	4	0.000687	犯罪 得罪 无罪
996	尊	zūn	13	5985	1	0.000167	尊重 尊严 尊敬	4	477	0	0	自尊 独尊 令尊
997	作	zuō	1	176	0	0	作坊	0	0	0	0	
998	昨	zuó	4	4751	0	0	昨天 昨晚 昨日	0	0	0	0	
999	左	zuǒ	20	7543	1	0.000133	左右 左手 左边	3	116	1	0.008621	极左 相左 中左
1000	坐	zuò	27	2735	3	0.001097	坐下 坐落 坐标	15	891	2	0.002245	乘坐 端坐 静坐

## 附录二 三音节词表(示例)

### 举例

者	消费者	工作者	投资者	与会者	经营者	志愿者	劳动者	爱好者	受害者	管理者	领导者	侵略者	组织者	获得者	旅游者
	所有者	统治者	研究者	第三者	吸毒者	决策者	使用者	建设者	幸存者	生产者	佼佼者	参与者	失败者	支持者	目击者
	竞争者	求职者	创业者	胜利者	保持者	主办者	旁观者	爱国者	参加者	参观者	受益者	教育者	来访者	成功者	失业者
	获奖者	感染者	示威者	死难者	革命者	创造者	创作者	表演者	设计者	后来者	打工者	参赛者	演讲者	摄影者	肇事者
	受访者	崇拜者	先行者	责任者	独裁者	合作者	发明者	开拓者	实践者	申请者	制造者	追随者	购买者	策划者	围观者
	当权者	制作者	缔造者	被害者	无产者	先驱者	当事者	发言者	出资者	投机者	流浪者	优胜者	特许者	挑战者	受难者
	偷猎者	从业者	受伤者	扮演者	执行者	上访者	创立者	收入者	倡导者	出版者	有志者	年长者	改革者	开创者	偷渡者
	幸运者	接收者	献血者	终结者	创建者	拓荒者	初学者	演唱者	叛乱者	学习者	勘探者	剥削者	造假者	罹难者	探路者
	集邮者	经商者	举报者	发送者	首倡者	资产者	报考者	违反者	观光者	观赏者	告密者	迟到者	鼓吹者	编撰者	遇害者
	受礼者	监管者	剧作者	驾车者	有产者	无聊者	带菌者	送礼者	入境者	垄断者	瞻仰者	有功者	首创者	受试者	倡议者
	逃亡者	主创者	控制者	饰演者	跟随者	参会者	编选者	体现者	护照者	作伪者	引进者	健在者	到会者	发行者	注意者
	来信者	初来者	满分者	掌握者	救人者	鉴定者	为政者	睿智者	失聪者	信教者	审判者	租用者	持有者	逝世者	滑冰者
	贩毒者	购物者	访问者	报喜者	破坏者	词作者	提出者	胜任者	失足者	廉洁者	要求者	钓鱼者	选举者	演奏者	拥戴者
	释放者	看海者	自豪者	廉政者	旅行者	评选者	良好者	服药者	嫌少者	求援者	食用者	私有者	持旗者	捍卫者	知情者
	殖民者	吸烟者	不幸者	张贴者	遇难者	离业者	盗版者	违纪者	持球者	编图者	写信者	入选者	执法者	违法者	创始者
	编者	零散者	讨论者	盗卖者	伤亡者	甩花者	勘察者	鉴赏者	执针者	发起者	羡慕者	源头者	卖官者	祭扫者	复仇者
	排污者	诠释者	冬泳者	丧生者	象牙者	拥有者	用血者	嫌多者	求购者	窥探者	吹箫者	入侵者	倡言者	作假者	登山者
	盗码者	勤政者	演出者	违章者	传播者	泡沫者	冠名者	圣战者	报忧者	错误者	捐款者	经销者	从政者	说情者	飘泊者
	毒品者	推动者	运送者	占有者	为官者	借钱者	利益者	大象者	淘粪者	功力者	守护者	操作者	同情者	代销者	开户者
	来稿者	审问者	承担者	持枪者	贡献者	仿制者									
人	领导人	负责人	年轻人	发言人	主持人	残疾人	外国人	当事人	老年人	候选人	犹太人	青年人	其他人	陌生人	经纪人

外地人 普通人 成年人 家里人 现代人 嫌疑人 代理人 城里人 创始人 中年人 村里人 接班人 机器人 当地人 带头人  
 继承人 西方人 乡下人 责任人 读书人 纳税人 聪明人 正常人 全家人 文化人 年青人 外星人 申请人 同龄人 代言人  
 制片人 被告人 受害人 生意人 东方人 心上人 承运人 债权人 投标人 借款人 被害人 发起人 有心人 局外人 召集人  
 监护人 自己人 全村人 有情人 明白人 合伙人 手下人 委托人 知情人 庄稼人 债务人 山里人 好心人 见证人 明眼人  
 介绍人 存款人 奠基人 土著人 都市人 掌门人 投资人 外乡人 举报人 过来人 中间人 调解人 自然人 领头人 过路人  
 出资人 牧羊人 当代人 植物人 少年人 提案人 付款人 直立人 承租人 后来人 贵妇人 策划人 热心人 贴心人 收信人  
 公诉人 外省人 贷款人 稻草人 主办人 辩护人 外来人 撰稿人 木头人 客家人 聋哑人 发明人 出租人 娘家人 骑车人  
 意中人 行为入 创建人 保证人 侵权人 文明人 经办人 担保人 买卖人 同路人 公证人 发行人 保护人 决策人 承包人  
 联系人 正经人 写信人 掌舵人 伤残人 尼基人 保管人 主婚人 手艺人 管理人 庄户人 调停人 故乡人 证婚人 作案人  
 剧中人 正派人 中介人 异乡人 白发人 经手人 带兵人 寄信人 阴阳人 太空人 报案人 受话人 负心人 知心人 壮年人  
 办报人 相对人 中国人 戍边人 推车人 外族人 寄卡人 违章人 写稿人 肇事人 应选人 两性人 克隆人 监票人 自诉人  
 放贷人 竞投人 申诉人 收发人 订货人 购买人 拍片人 哈尼人 出品人 白种人 主讲人 拍卖人 机械人 欠款人 受贿人  
 鉴定人 涉案人 自家人 古代人 色目人 采药人 领军人 邻里人 赞助人 盗车人 回回人 驾车人 出家人 领路人 引路人  
 持旗人 犯罪人 制作人 投保人 监管人 外行人 捐资人 引水人 带路人 寄件人 发运人

运动员 通讯员 服务员 研究员 评论员 司令员 教练员 飞行员 指导员 警卫员 业务员 营业员 驾驶员 裁判员 指挥员  
 观察员 管理员 推销员 技术员 辅导员 售票员 教导员 审判员 保安员 侦查员 监督员 列车员 工作员 采购员 协调员  
 饲养员 辩论员 仲裁员 侦察员 讲解员 操作员 护林员 销售员 勤务员 监察员 联络员 办事员 保管员 打字员 咨询员  
 程序员 促销员 伤病员 检察员 交易员 战斗员 值班员 宣传员 解说员 公证员 助理员 通信员 卫生员 担架员 潜水员  
 情报员 预审员 检查员 航天员 调解员 试飞员 信息员 接待员 广播员 理发员 普查员 实验员 招待员 安全员 信贷员  
 书记员 报告员 收款员 保育员 接线员 演奏员 传销员 协理员 治安员 护理员 调度员 放牧员 投递员 公安员 导游员  
 调查员 会计员 记录员 检验员 出纳员 巡视员 交通员 放映员 接生员 统计员 创作员 组成员 电话员 储蓄员 质检员  
 预报员 化验员 训导员 调研员 校对员 参政员 分析员 客运员 航空员 宣讲员 录入员 资料员 督察员 发行员 向导员  
 评判员 机要员 作业员 审讯员 联防员 购销员 传真员 撰述员 审核员 巡查员 送货员 福利员 收费员 枪械员 直销员  
 速记员 报单员 办理员 主办员 考勤员 操纵员 引水员 稽核员 补票员 急救员 计时员 时评员 调配员 评审员 税务员  
 清洁员 任教员 核算员 组织员 打包员 簿记员 谍报员 报道员 从业员 发球员 救生员 看护员 给养员 测量员 警戒员  
 看守员 翻译员 见习员 保洁员 维护员 训练员 扫描员 剪票员 社会员 审计员 复核员 撰稿员 护航员 运送员 绘图员

记分员 品种员 保险员 工艺员 外销员 维修员 记工员 滑翔员 协办员 执行员 供销员 估价员 评议员 水管员 送报员  
导航员 接收员 动态员 收发员 播音员 献血员 罚款员 唱票员 考察员 科技员 兜售员 退货员 代理员 记账员 户籍员  
构成员 军队员 计票员 采访员 处理员 检测员 清扫员 制作员 特务员 连接员 表演员 摄像员 编制员 干事员 代办员  
结算员 关口员 设计员 农机员 视察员 押运员 监理员 中介员 跳伞员 拓展员 预算员 具有员 护卫员 查票员 搜集员  
弟子员 教养员 体育员 墙报员 航飞行员 传达员 司法员 讲评员 新教员 监管员 故事员 品酒员 说明员 教务员 舞蹈员  
会议员 灭鼠员 编码员 药剂员 审察员 领队员 征收员 交涉员 验收员 工具员 驯养员 字幕员 评估员 迎宾员 保护员  
管护员 材料员 贷款员 事务员 收购员 鉴定员 编写员 提审员 团委员 户口员 引导员 督导员 职务员 配送员 控制员  
供应员 增加员 计划员 招聘员 声纳员 省道员 领导员 清理员 电脑员 网络员 软件员 打井员 板报员 跑腿员 轰炸员  
警备员 审查员

师 工程师 设计师 会计师 建筑师 摄影师 化妆师 驯兽师 魔术师 理发师 农艺师 经济师 按摩师 美容师 拍卖师 摄像师  
气功师 中医师 营养师 审计师 钢琴师 估价师 药剂师 精算师 麻醉师 美术师 工艺师 化妆师 针灸师 策略师 灯光师  
舞美师 净化师 雕刻师 发型师 畜牧师

家 科学家 企业家 艺术家 政治家 哲学家 资本家 评论家 分析家 音乐家 革命家 文学家 数学家 观察家 书法家 思想家  
小说家 作曲家 教育家 史学家 美术家 理论家 摄影家 军事家 美食家 化学家 歌唱家 实业家 收藏家 剧作家 批评家  
探险家 指挥家 银行家 雕塑家 书画家 演奏家 外交家 舞蹈家 汉学家 发明家 漫画家 野心家 活动家 法学家 慈善家  
戏剧家 旅行家 实干家 钢琴家 散文家 出版家 航海家 阴谋家 美学家 改革家 鉴赏家 冒险家 预言家 战略家 农学家  
投资家 金融家 设计家 雕刻家 政论家 道学家 编辑家 神学家 儒学家 篆刻家 油画家 词作家 编导家 纵横家 制造家  
阴阳家 空想家 林学家 演唱家 空谈家 谋略家 理想家 经销家 园艺家 声学家 集邮家 独奏家 电影家 翻译家 表演家

手 一把手 刽子手 主攻手 狙击手 红旗手 削球手 吹鼓手 拳击手 二传手 多面手 二把手 得分手 机械手 爆破手 突击手  
刀斧手 快攻手 农机手 三把手 甩花手 操作手 骗子手

## 参考文献

### 中文书目

#### 专书:

- 王力 (1944). 中国语法理论. 上海:商务印书馆.
- 吕叔湘 (1979). 汉语语法问题分析. 北京:商务印书馆.
- 朱德熙 (1982). 语法讲义. 北京:商务印书馆.
- 陆志韦 (1937). 北京话单音词词汇. 陆志韦语言学著作集(三). 北京:中华书局,1990.
- 陆志韦 (1964). 汉语的构词法(修订本). 北京:科学出版社.
- 陈光磊 (1994). 汉语词法论. 上海:学林出版社.
- 陈保亚 (1999). 20 世纪中国语言学方法论. 济南:山东教育出版社.
- 陈保亚 (2009). 当代语言学. 北京:高等教育出版社.
- 张普 (2009). 动态语言知识更新研究. 北京:商务印书馆.
- 赵元任 (1968). A Grammar of Spoken Chinese. Los Angeles: University of California Press. 赵元任全集(第3卷). 北京:清华大学出版社, 2004.
- 徐通锵 (1997). 语言论. 沈阳:东北师范大学出版社.
- 董秀芳 (2004). 汉语的词库和词法. 北京:北京大学出版社.

#### 期刊论文:

- 马庆株 (1995). 现代汉语词缀的性质、范围和分类. 中国语言学报, (6):101-137. 著名中年语言学家自选集:马庆株卷. 合肥:安徽教育出版社, 2002: 42-91.
- 王洪君 (1994). 从字和字组看词和短语. 中国语文, (2): 102-111.
- 王洪君 (2001). 《信息处理用现代汉语分词词表》的内部构造和汉语的结构特点. 语言文字应用, (4): 90-97.
- 王洪君, 富丽 (2005). 试论现代汉语的类词缀. 语言科学, (5): 3-17.
- 吕叔湘 (1962). 说“自由”和“黏着”. 中国语文, (1): 1-6.
- 张普 (2004). 基于动态流通语料库的语感模拟和新词语提取研究. 外国语言文学研究, (2): 23-31.
- 郭良夫 (1983). 现代汉语的前缀和后缀. 中国语文, (4): 250-256.
- 郭锐 (1996). 汉语语法单位及其相互关系. 汉语学习, (1): 8-14.
- 徐通锵 (1991). 语义句法刍议——语言的结构基础和语法研究的方法论初探. 语言教学与研究, (3): 38-62.
- 徐通锵 (1994).“字”和汉语研究的方法论. 世界汉语教学, (3): 1-14.

#### 学位论文:

梁源 (1999). 二字结构凝固度分级研究——兼论汉语词和非词的界限问题[硕士学位论文]. 北京: 北京大学中文系。

## 英文书目

### Books:

- Aitchison, J. (2012). *Words in the mind: An introduction to the mental lexicon* (4<sup>th</sup> ed.). Oxford: John Wiley & Sons.
- Anderson, S. R. (1992). *A-morphous morphology*. Cambridge: Cambridge University Press.
- Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, MA: MIT Press.
- Bauer, L. (1983). *English word-formation*. Cambridge: Cambridge university press.
- Bauer, L. (2001). *Morphological productivity*. Cambridge: Cambridge University Press.
- Bauer, L. (2004). *A glossary of morphology*. Edinburgh: Edinburgh University Press.
- Bauer, L., Lieber, R., & Plag, I. (2013). *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.
- Baker, M. (1988). *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.
- Bloomfield, L. (1933). *Language*. Holt, Rinehart and Winston.
- Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford: Oxford University Press.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row, Publishers.
- Crystal, D. (2008). *A dictionary of linguistics and phonetics* (6<sup>th</sup> ed.). Blackwell.
- Daelemans, W., & Bosch, A. V. D. (2005). *Memory-based language processing*. Cambridge: Cambridge University Press .
- Fertig, D. (2013). *Analogy and morphological change*. Edinburgh: Edinburgh University Press.
- Greenberg, J. (1966). *Language universals: With special reference to feature hierarchies*. The Hague: Mouton.
- Harman, G. & Kulkarni, S. (2012). *Reliable reasoning: Induction and statistical learning theory*. Cambridge, MA: MIT Press.
- Harris, Z. S. (1951). *Methods in structural linguistics*. Chicago: The University of Chicago Press.
- Haspelmath, M. & Sims, A. (2010). *Understanding morphology* (2<sup>nd</sup> ed.). London and New York: Routledge.
- Hay, J. (2003). *Causes and Consequences of word structures*. London and New York: Routledge.
- Hockett, C. (1958). *A course in modern linguistics*. New York: Macmillan.

- Hockett, C. (1987). *Refurbishing our foundations: Elementary linguistics from an advanced point of view*. Amsterdam: John Benjamins.
- Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Jespersen, O. (1942). *A modern English grammar on historical principles. Part VI. Morphology*. London and Copenhagen: George Allen & Unwin and Ejnar Munksgaard.
- Kurzweil, R. (2012). *How to create a mind: The secret of human thought revealed*. Penguin.
- Langacker, R. (1987). *Foundations of cognitive grammar: Vol.1. Theoretical prerequisites*. Stanford, CA: Stanford university press.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. Harlow, England: Longman.
- Lees, R. B. (1960). *The grammar of English nominalizations*. The Hague: Mouton.
- Lieber, R. (2009). *Introducing morphology*. Cambridge: Cambridge University Press.
- Inkelas, S. (2014). *The interplay of morphology and phonology*. Oxford: Oxford University Press.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
- Marchand, H. (1969). *The categories and types of present-day English word-formation: A synchronic-diachronic approach* (2<sup>nd</sup> ed.). Beck.
- Matthew, P. (1991). *Morphology* (2<sup>nd</sup> ed.). Cambridge: Cambridge University Press.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Packard, J. L. (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.
- Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: Basic Books.
- Plag, I. (1999). *Morphological Productivity: structural constraints on English derivation*. Berlin and New York: Mouton de Gruyter.
- Sapir, E. (1921). *Language: An introduction to the study of speech*. New York: Harcourt Brace.
- Skousen, R. (1989). *Analogical modeling of language*. Dordrecht: Kluwer.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Taylor, J. R. (2012). *The mental corpus: how language is represented in the mind*. Oxford: Oxford University Press.
- Trask, T. L. (2007). *Language and linguistics: the key concepts*. Tylor & Francis.
- Xiao, R., Rayson, P., & McEnery, T. (2009). *A frequency dictionary of Mandarin Chinese*. London and New York: Routledge.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.
- Zipf, G. K. (1965). *The psycho-biology of language*. Cambridge, MA: MIT Press.

Articles:

- Abney, S. (1996). Statistical methods and linguistics. In J. L. Klavans and P. Resnik (Eds.), *The balancing act: Combining symbolic and statistical approaches to language* (pp. 1-26). Cambridge, MA: MIT Press.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89(4), 369-406.
- Antic, E. (2012). Relative frequency effects in Russian morphology. In S. T. Gries & D. Divjak (Eds.), *Frequency effects in language learning and processing* (Vol.1, pp.83 - 107). Berlin and New York: Walter de Gruyter.
- Arcodia, G. F., & Basciano, B. (2012). On the productivity of the Chinese suffixes 儿-r, -化-hua, -头-tou. *Taiwan Journal of Linguistics*, 10(2), 89-118.
- Aronoff, M. (1983). Potential words, actual words, productivity and frequency. *Proceedings of the 13th International Congress of Linguists*, 163-171.
- Bauer, L. (1979). Against word-based morphology. *Linguistic Inquiry*, 10, 508-9.
- Bauer, L. (1998). When is a sequence of two nouns a compound in English? *English Language and Linguistics*, 2(01), 65-86.
- Bauer, L. (1992). Scalar productivity and -ly adverbs. In G. Booij and J. V. Marle (Eds.), *Yearbook of morphology 1991*(pp. 185-191). Dordrecht: Kluwer.
- Baayen, H. (1992). Statistical models for word frequency distributions: A linguistic evaluation. *Computers and the Humanities*, 26(5-6), 347-363.
- Baayen, H. (1993). On frequency, transparency and productivity. In G. Booij and J. van Marle (Eds.), *Yearbook of Morphology 1992* (pp. 181-208). Dordrecht: Kluwer.
- Baayen, H. (2003). Probabilistic approaches to morphology. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 229-287). Cambridge, MA: MIT Press.
- Baayen, H., & Lieber, R. (1991). Productivity and English derivation: a corpus-based study. *Linguistics*, 29, 801-844.
- Baayen, H., & Renouf, A. (1996). Chronicling the Times: productive lexical innovations in an English newspaper. *Language*, 72, 69-96.
- Baayen, H., & Schreuder, R. (1999). War and peace: Morphemes and full forms in a noninteractive activation parallel dual-route model. *Brain and language*, 68(1), 27-32.
- Bloomfield, L. (1926). A set of postulates for the science of language. *Language*, 2, 153-164.
- Bolinger, D. L. (1948). On defining the morpheme. *Word*, 4(1), 18-23.
- Bybee, J. (2005). Restrictions on phonemes in affixes: A crosslinguistic test of a popular hypothesis. *Linguistic Typology*, 9, 165-222.
- Bybee, J. (2011). Usage-based theory. In P. C. Hogan (Ed.). *The Cambridge encyclopedia of the language sciences* (pp. 889-891). Cambridge: Cambridge University Press.
- Bybee, J., & Slobin, D. (1982). Rules and schemas in the development and use of the English Past Tense. *Language*, 58, 265-289.

- Carstairs-McCarthy, A. (2006). Affixation. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (Vol.1, pp. 84-88). Amsterdam: Elsevier.
- Chomsky, N. (1970). Remarks on nominalizations. In R.A.J. Rosenbaum, & P. S. Rosenbaum (Eds.), *Readings in English transformational grammar* (pp. 184-221). Ginn.
- Colé P., Segui, J., & Taft, M. (1997). Words and morphemes as units for lexical access. *Journal of Memory and Language*, 37(3), 312-330.
- Downing, P. (1977). On the creation and use of English compound nouns. *Language*, 53, 810-842.
- Ellis, N. (2012). What can we count in language, and what counts in language acquisition, cognition, and use? In S. T. Gries & D. Divjak (Eds.), *Frequency effects in language learning and processing* (Vol.1, pp. 7-33). Berlin and New York: Walter de Gruyter.
- Fenk-Oczlon, G. (2001). Familiarity, information flow, and linguistic form. In J. Bybee and P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 431-448). Amsterdam: John Benjamins.
- Frauenfelder, U. H., & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In G. Booij and J. V. Marle (Eds.), *Yearbook of Morphology 1991* (pp. 165-183). Dordrecht: Kluwer.
- Goldsmith, J. (2009). Morphological analogy: Only a beginning. In J. P. Blevins, & J. Blevins (Eds.), *Analogy in grammar: Form and acquisition* (pp. 137-163). Oxford: Oxford University Press.
- Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universal of language* (pp. 59-91). Cambridge, MA: MIT Press.
- Harwood, F. W., & Wright, A. M. (1956). Statistical study of English word formation. *Language*, 32,260-273.
- Hay, J. (2001). Lexical frequency in morphology: is everything relative?. *Linguistics*, 39(6), 1041-1070.
- Hockett, C. (1960). The origin of speech. *Scientific American*, 203, 88-96.
- Hockett, C. (1961). Linguistic elements and their relations. *Language*, 37, 29-53.
- Horn, L. R. (2008). Pragmatics and the lexicon. In P. V. Sterkenburg (Ed.), *Unity and diversity of languages* (pp. 29-41). Amsterdam: John Benjamins.
- Jakobson, R. (1965). Quest for the essence of language. *Diogenes*, 13, 21-37.
- Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 39-95). Cambridge, MA: MIT Press.
- Kiparsky, P. (1982). Lexical phonology and morphology. In The Linguistic Society of Korea (Ed.), *Linguistics in the morning calm* (pp. 1-91). Seoul: Hashin Publishing.
- Krott, A., Schreuder, R., Baayen, H. (2001). Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics*, 1, 51-93.
- Krott, A., Schreuder, R., Baayen, H., & Dressler, W. (2007). Analogical effects on linking elements in German compound words. *Language and cognitive processes*, 22(1), 25-57.
- Legendre, G., Miyata, Y., & Smolensky, P. (1990). Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In *Proceedings of the*

*Twelfth Annual Conference of the Cognitive Science Society*, 388–395.

- Libben, G. (2006). Why study compound processing? An overview of the issues. In G. Libben & G. Jarema (Eds.), *The representation and processing of compound words* (pp.1-22).Oxford: Oxford University Press.
- Nishimoto, E. (2003). Measuring and comparing the productivity of Mandarin Chinese suffixes. *Computational Linguistics and Chinese Language Processing*, 8(1), 49-76.
- Nosofsky , R. M . (1992). Exemplar-based approach to relating categorization, identification, and recognition. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 363-93). Hillsdale, NJ : Lawrence Erlbaum.
- MacWhinney, B. & Leinbach, J. (1991). Implementations are not conceptualizations: revising the verb learning model. *Cognition*, 40, 121-157.
- Pierrehumbert, J. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 177-228). Cambridge, MA: MIT Press.
- Pinker, S. (1991). Rules of language. *Science*, 153, 530–535.
- Pinker, S. & Prince, A. (1988). On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28, 73-193.
- Plag, I. (2006). The variability of compound stress in English: Structural, semantic, and analogical factors. *English Language & Linguistics*, 10(01), 143-72.
- Skousen, R. & Chandler, S. (2011). Exemplar theory. In P. C. Hogan (Ed.). *The Cambridge encyclopedia of the language sciences* (pp. 300-302). Cambridge, NY: Cambridge University Press.
- Rumelhart, D. & McClelland, J. (1986). On learning the past tenses of English verbs. In J. McClelland, D. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing* (Vol.2, pp. 216–271). Cambridge, MA: MIT Press.
- Smolensky, P. (1999). Connectionist Approaches to Language. In R. A. Wilson, & F. C. Keil (Eds.). *The MIT encyclopedia of the cognitive sciences* (pp.188-190). Cambridge, MA: MIT press.
- Sproat, R., & Shih, C. (1996). A corpus-based analysis of Mandarin nominal root compound. *Journal of East Asian Linguistics*, 5(1), 49-71.
- Sproat, R., & Shih, C. (2002). *Corpus-based methods in Chinese morphology and phonology*. Tutorial at the 19th COLING.
- Thorndike, E. L. (1943). Derivation ratios. *Language*, 19, 27-37.
- Dissertations:
- Dai, X. L. (1992). *Chinese morphology and its interface with syntax* (Doctoral dissertation, The Ohio State University).
- Yang, X. (2013). *Morphological productivity in Chinese [AN]: A corpus-based study*.(Unpublished master's thesis, National University of Singapore.)