

现代汉语词长弹性的量化研究*

黄丽君 (四川大学外国语学)
端木三 (密歇根大学语言学系)

提要 本文先讨论对汉语词长弹性的两种看法, 确认词长弹性是汉语的一大特点。然后根据《现代汉语词典》, 对词长弹性进行全面的量化考察。对词典收录的单音节词条的 1/10 进行抽样, 针对每一词义逐条考察其弹性特征。结果发现, 汉语近 70% 的词有弹性长度, 其中名词的比例最高, 超过 80%, 动词的比例略低于名词。文体对词长弹性比例也有影响。

关键词 词长弹性; 词条; 词义; 词类; 词汇; 复合词

1 引言

汉语不少词有长短两种形式, 如: 煤炭-煤、技术-技、老虎-虎、鸭子-鸭、学习-学、等。这种现象郭绍虞 (1938) 称之为词长的弹性, 我们也可称之为长短词。词长弹性又引出词长搭配现象 (如: 技术+工人、技术+工、*技+工人、技+工) 和词序变化现象 (如: 大型汉语词典、*汉语大型词典、汉语大词典、? 大汉语词典)。“*”表示不好用例, “?”表示不常用例。关于词长搭配和词序变化, 学术界已有不少讨论 (如: 吕叔湘 1963, 端木三 1999; 周荐 1999; 刘晓梅 2004; 冯胜利 2005; 周韧 2007; Duanmu 2012, 等), 本文不多重复。本文的重点, 是探讨词长弹性本身。

关于词长弹性, 有两种不同看法。一种认为, 一个词可以有两个语音形式, 一长一短, 两者基本同义, 所以是同一词。郭绍虞 (1938), 潘文国 (1997) 持这一观点。赵元任、杨联陞也持这一观点, 比如, 在他们所编的《国语字典》里 (Chao and Yang 1947), 长短词列于同一词义之下 (如: ‘清’、‘清楚’归同一词义, 给一个标注; ‘清’ (清楚)、‘清’ (清白) 则归不同词义, 分别标注), 语义完全不同的词则一律分列, 如‘仿’ (模仿) 和‘仿’ (仿佛) 分列于两个词条, ‘供’ (提供) 和‘供’ (供奉) 也分列与两个词条。

第二种看法认为, 语音的不同会产生语义的不同, 因此, 长短词属于不同的词。持这种观点的学者有李临定 (1990), 刘凤樾 (Liu 1992), 王灿龙 (2002), 吴为善 (2006), 柯航 (2007) 等。比如, 李临定 (1990)、王灿龙 (2002) 认为, 单音节语义较具体, 双音节语义较抽象 (或者数量较大)。李临定 (1990)、Liu (1992) 又认为, 单音节有动词倾向, 双音节有名词倾向。吴为善 (2006)、柯航 (2007) 认为, 在节奏上, 1+2 和 2+1 的紧密度不同, 进而会产生语义的不同。比如, 柯航 (2007) 举例说, ‘学校店’指以学生教师为顾客的商店 (但不一定是学校拥有), 而‘校商店’指学校拥有的商店 (但顾客不一定以学生教师为主)。

以上两种看法, 不一定非得完全对立。比如, 刘智伟 (2007)、冯胜利 (2010) 认为, 单音节和双音节有语体上的差别, 单音节比较口语化, 双音节比较书面化, 特别是动词, 但两者基本同义。我们认为, 如果因为长短词在语体上或其他方面的某些区别, 就否定语音或韵律的作用, 把它们当作不同词, 恐怕会有不少问题。比如, 王灿龙 (2002) 认为, 词长搭配跟韵律无关, 只跟语义有关。他认为, 1+2 的名+名结构没有什么不好, 并列举了一些例子。但 Duanmu (2012) 根据语料统计发现, 1+2 的名+名出现率只有 1% 左右, 远远低于 1+1, 2+1, 及 2+2 的出现率。而且, 1+2 的例子多限于两种结构, 即‘材料’+名词 (‘皮手套’, ‘棉大衣’, 等), 或‘所有格’+名词 (‘党代表’, ‘系主任’, 等), 王

灿龙的反例，多数也属于这两类。如果将少数的、特殊的例子和大量的、普通的例子混为一谈，就可能忽略问题的本质。王灿龙还认为，2+2 的名词组不能指具体概念，因此，‘煤炭商店’、‘技术工人’、‘手表工厂’等皆“不是地道汉语”，这种说法，尚未见到证据（如：中国人是否一概不用，或多数中国人是否将这些表达判断为错）。又如，只从语体出发，也不能完全解释词长搭配。比如，词+词是不是应该要么 2+2（正式），要么 1+1（非正式）？是不是 1+2 和 2+1 皆有语体混淆，两者应该都少？若是，为什么名+名结构的 2+1 优于 1+2，而动+宾结构却相反，1+2 优于 2+1 呢？还有，‘技术工人’和‘技术工’，哪个语体更好，我们根据什么来判断？我们还必须肯定，有的表达有习惯性，比如，‘种花’常说，‘种草’、‘种花草’却不常说，‘油田’常说，‘石油田’却不常说，我们不能因此推断后者有语法问题（王灿龙 2002）或有节奏问题。

我们认为，长短词可以有某些语体上或使用上的不同倾向，但在多数情况下，其基本语义仍然相同或相近（即所指相同），这点恐怕有公认。而且，无论长短词算一词还是两词，它们之间的关系与其他词之间的关系有着系统的不同，因此，长短词应该作为一个独立的现象来研究。长短词之间的关系，具体有以下几点：

(1) 长短词之间的关系

- a. 两词有相同语素（同源，同音，同词类）
- b. 基本语义相同或相近，如：所指相同，词典互注
- c. 长词的多余部分不提供明确语义（即：语义结构是 XX，0X，X0）
- d. 两词在某些场合可以同义互换

以上（1a）排除了没有相同语素的词对，如英语的 *happy* 和 *glad*（皆指‘高兴’）。（1b）排除了同音但不同义的词对，如‘仿’（模仿）和‘仿’（仿佛），其中词典互注指长词往往用来标注短词，如《现代汉语词典》里‘煤’标注为‘也叫煤炭’，‘需’标注为‘需要’。很明显，词典互注从客观上反映了长短词的同义性。（1c）说明，长短词不同于真正的复合词。真正的复合词，语义关系是 XY，即两词的语义不同，如‘书包’，‘镜框’，等。而长短词的语义结构，要么是重复性的（即 XX），如‘技术’、‘煤炭’、‘学习’、‘需要’等，要么上字无实际意义（即 0X），如‘老虎’，要么下字无实际意义（即 X0），如‘鸭子’。（1d）要求长短词在某些场合可以同义互换，比如，‘养鸭子’和‘养鸭’基本同义（从而得长短词‘鸭子-鸭’），‘技术工’和‘技工’基本同义（从而得长短词‘技术-技’）。值得注意的是，我们不要求长短词在任何场合都可互换，比如，‘技术’和‘技’在‘工人’前不能互换（不说‘技工人’），‘树’和‘树木’在‘公园里有很多’后可以互换，在‘成林’前却不能互换（不说‘树成林’）。这是因为词长跟节奏有关，词长搭配要受节奏的影响。还值得注意的是，有的语义结构，看上去是 XY 实际上是 X0。比如，‘法国’看上去是 XY，因为‘法’和‘国’分别有实际意义，但在‘日本，加拿大，巴西，法国’里，‘国’的语义实际上是多余的，它的出现完全出于节奏的考虑，所以‘法国’可以算 X0。

符合（1）中四条关系的长短词，汉语里很多。英语里好像也有，如 *math-mathematics* ‘数学’，*lab-laboratory* ‘实验室’，但其往往不符合（1a）或（1c），而且其数量比汉语少得多。因此，长短词是汉语的一大特征。

那么，汉语到底有多少词有弹性长度？不同词类的弹性比例有何不同？如：名词的弹性比例是

否高于动词？虽然有学者进行过估计，如潘文国（1997：140）认为汉语“几乎所有的词”都有弹性长度），但目前尚无系统的量化研究。Duanmu（2011）对汉语单字词条进行了抽样考察，认为汉语 80-90%的词条有弹性长度，不过，其抽样数量较少（共 60 字，84 词条），而且没有细分各词义的弹性规律。本文根据《现代汉语词典》第五版（中国社会科学院语言研究所词典编辑室 2005，下称《现汉》）对汉语词长弹性进行全面的量化考察。通过统计数据，探讨不同文体以及不同词类对现代汉语词长弹性的影响。

2 字、词、词组

无论英语汉语，要给词下定义都不容易。Bloomfield (1926) 把词定义为“最小的自由形式”(minimum free form)。但这一定义把有些我们通常认定为词的单位排除在外了，如 a, the, my, to, of 等功能词。另外，派生词给词典编撰带来一个问题。比如，英语电子词典 CELEX (Baayan et al. 1995) 提供了三种不同的词汇量。一是词形词典 (word-form lexicon)，收录英语词汇 16 万左右。词形词典收录的词条包括所有曲折变化后的词形，如 call, calls, calling, called 都作为单独词条分别收录。二是基本词典 (lemma lexicon)，收录基本词目 5 万余条，不收录曲折变化的词形，比如，只收 call 作为词条，不收 calls, calling, called 等。三是单语素词汇，即只有一个语素的词，共收录 7 千多条。

汉语要区分字和词也不容易。汉语中的字，是一个单音节，但汉语的词，在形式上很难分辨。正如吕叔湘（1984：491-492）所说，“词在两头都有划界问题：一头是如何区别单独成词的语素和单独不成词的语素；另一头是如何决定什么样的语素组合只是一个词，什么样的语素组合构成一个短语”。像“店”、“技”、“工”这类不单用的算不算词，学界各说不一。比如“店”不单用，那它就不算一个词，但是，“店”也不是一个词缀，因为它有其指称意义，而词缀没有确切的指称意义；同时，“店”在另一个词旁边的位置不固定，可左可右，如“商店”、“店家”等。英语中的词缀通常只能附于词根的某一边，要么是前缀，要么是后缀，不能有时前有时后，如 un- 是前缀，只能附于词根的左边，如 undo, unhappy 等，而 -ness 是后缀，只能附于词根的右边，如 goodness, loneliness 等。另外，“店”还有一个双音节形式“商店”，而“商店”是可以单用的。也就是说，汉语一个概念往往有两个语音形式，一个长形，一个短形，长形能单用，短形往往不单用。如果这样看，单音节其实也可以代表词，因为该词还有一个双音节的形式。

《现汉》第五版的一个新特征是在区分词与非词的基础上给词标注词类。这有助于我们考察不同词类的词长特征。如词条“青”根据词义标注词类如下（文中带圆圈的数字如①②③等表示词义项，下同）：

- (2) 青①<形> 蓝色或绿色 ~天
- ②<形> 黑色 ~布
- ③ 青草或没有成熟的庄稼 踏~
- ④ 比喻年轻 ~年
- ⑤ 指青年 ~工（青年工人）
- ⑥<名>姓

“青”可用作形容词，表示“蓝色或绿色”或“黑色”，也可用作名词，表示姓名。“青”还可用作不成词语素，同别的语素构成复合词（如③④⑤）。不难看出，《现汉》判断词与非词的标准是能否单用。这样的区分，就把前文讨论的短形排除在外了。如“青⑤”，指青年，那么“青”就是“青年”的短形，二者表达的意思相同，且具有相同的构词功能：“青工”—“青年工人”。在我们的分析中，“青”和“青年”就是一对长短词。同样，“青③”的“青”和“青草”也是一对长短词。

“青”字下面的多字条目同样在区分词与非词的基础上标注词类，如青帮<名>、青菜<名>、青菜头<名>、青草<名>、青茶<名>、青霉素<名>、青春<名>、青春期<名>、青春痘<名>、青葱<形>、等等。没有标明词类的如：青出于蓝、青翠欲滴、青红皂白、青梅竹马、青面獠牙、青天白日、青天霹雳、青铜时代、青云直上、等等。这些都是由“青”字构成的词组(包括成语、熟语)，所以《现汉》也不标注词类。

总之，《现汉》将词目分三类：（1）一般不单用的单字语素，（2）可以单用的单词及复合词，（3）词组。《现汉》给（2）标注词类，（1）、（3）不标。我们认为（1）也可以标，理由有两点：第一，（1）往往有一个长形，即可单用的双音节（或多音节）形式。第二，英语的语素都可以标注词类，比如 CELEX 给所有的词和语素，包括所有的词缀，一律标注词类。

3 词条和词义

《现汉》收录“词条”(entries) 6 万 7 千余条，包括单音节词条和多音节词条。每一条再细分“词义”(senses)若干。形同而音、义不同的，则分立词条，如“奇”qí 和“奇”jī:

- (3) 奇 qí ①罕见的，特殊的，非常的：~闻
②出人意料的，令人难测的：~兵
③惊异：不足为~
④<名>姓
奇 jī ①单的，不成对的（跟“偶”相对）：~数
②<书>零数：五十有~

形同音同而意义明显不同的，也分立条目，如“美”有两个词条，每一条又细分各词义：

- (4) 美①<形>美丽，好看（跟“丑”相对）：这小姑娘长得真~
②使美丽：~容
③<形>令人满意；好：价廉物~
④美好的事物；好事：~不胜收
⑤<方><形>得意：老师夸了他几句，他就~得了不得
⑥<名>姓
美<名>①指美洲：南~
②指美国：~籍华人

词长弹性，既可以指词条的长度有弹性，又可以指词义的长度有弹性，而后者更加精确。如果

只考察词条的词长弹性，会出现一些难以避免的问题。

首先，一词条下通常收录多条词义，有的词义有词长弹性，有的词义没有，那么这一词条到底算有弹性还是算没有，就很难做判断。如例（4）中的“美”，形同义不同，因而分立两个词条。在第一个词条下，第一个词义有弹性特征（美—美丽），而其他词义没有，那么“美”到底算有弹性还是无弹性？如果判定为无，似乎不合理，因为“美—美丽”有弹性。如果判定为有，似乎也不合理，因为并不是所有的词义项都有弹性。考察词义就可以避免这一问题，针对每项词义，我们分别统计。同时，对有弹性的词，我们逐一加注其长式。如前例（4）中第一词条下的“美”，只有一个词义项有弹性，其他词义项均算作无弹性。而“美”的第二词条下的两个词义项都有弹性（美—美洲；美—美国）。值得指出的是，长短词的长式往往直接来自词典的注释，如（4）所示。

其次，如果只考察词条的弹性，很难对不同词类的弹性做精确的统计。因为一个词条下收录的词义项分属不同的词类，如“学”可用作动词，“学-学习”，也可以用作名词，“学-学问”，那么词条“学”算动词还是名词呢？如果词条的词类不分开，很难统计不同词类的词长特征。

4 方法

我们的考察对象是简单词，即单语素的词。主要原因是《现汉》收录的复合词多为 1+1 式，即由两个单音节词构成，如：书包、鸡蛋、签名、狂欢，环保，货运等。部分复合词无弹性，如“书包”、“鸭蛋”等，而部分有弹性的复合词其弹性来自于其构成成分的弹性，如“环保-环境保护”、“货运-货物运输”，考察简单词的弹性也就知道复合词的弹性。另外，所有构成复合词的简单词已经作为单独词条收录在《现汉》中了。为了避免重复统计，我们以《现汉》中的简单词为重点。

简单词主要有三类。一类是单音节词，如“天”、“地”、“人”、“走”、“跑”、“跳”等。一类是双音节的简单词，如“老虎”、“兔子”、“书籍”、“技术”、“大蒜”等，它们看似复合词，其实是简单词，因为它们语义结构不是复合性的 XY，而是同义/近义重复 XX（如“书籍”、“技术”），或是其中一构词成分不具实质意义（如“兔子”属于 X0、“老虎”属于 0X）。这类词其实是单音节词的长式，它们同相应的短式“虎”、“兔”、“书”、“技”、“蒜”基本同义，而这些单音节词已作为单独词条收录在《现汉》中了。第三类是双/多音节单语素词，如“钥匙”、“玛瑙”、“尴尬”等，其构词成分基本不单用。这类简单词都列在单字义项里，如“钥”见“钥匙”，“尴”见“尴尬”。出于以上考虑，我们的主要考察对象是《现汉》收录的单音节词条。

《现汉》收录单音节词条 1 万多，共计 2 万余项词义。我们对其中 1/10 进行抽样分析，具体步骤如下：

- 1) 去掉少量带难检字的词条（难检字在 Windows Office 文档中显示为空白）；
- 2) 去掉相同词条，即标注为“<古>又同”词条：如“份”<古>又同“彬”bīn，表示“份”在古代其用法同“彬”，表示文雅有礼貌的样子，但现代汉语中没有这个用法，去掉该词条；如果一词义项被标注为“同”（如“叭”bā 同“吧”bā），在词类栏标注“同”。在《现汉》中，如果 A 被标注为“同 B”，而“B”未被标注为“同 A”，可能的原因是 B 更常用，如“晖②”同“辉”，但在“辉”词条下没有词义项同“晖”。这样，我们的研究去掉冗余，保留相关词项；

- 3) 按顺序, 从每 10 个词条中抽取 1 个词条;
- 4) 把每个词条的每个词义单行列出: 如词条“败”下有 9 个词义①<动>在战争或竞赛中失败(跟“胜”相对); ②<动>使失败, 打败(敌人或对手); ③(事情)失败(跟“成”相对); ④毁坏, 搞坏(事情); ⑤排除; 消除; ⑥破旧; 腐烂; ⑦<动>凋谢; 枯萎; ⑧<动>败落; ⑨<动>使败落; 9 个词义项单列 9 行, 分别考察其词长, 词类, 文体特征;
- 5) 人工标注词长: 如果一个词义项具有一个同义或近义的双音节形式, 在词长栏给例(如: 技-技术、虎-老虎、鸭-鸭子、分-分开、搬-搬动、败-失败、等)。同/近义的双音节形式给例, 多数直接来自词典标注。除此以外, 还有两条标准。一条是, 在某个语境下, 双音节形式可以替换单音节形式, 且词义基本不变, 如: 技工-技术工, 搬家具-搬动家具, 败了-失败了, 等。另一条是, 双音节形式的语义结构要么含一个无具体词义的音节(如: 0X 类的“老虎”和 X0 类的“鸭子”等); 要么两个音节词义重复(即 XX 类的“技术”、“失败”等)。如果一词义项没有双音节形式, 在词长栏标注“1”, 如“饭”、“床”等可以单用的词, 以及只能跟别的词连用的词如“醇”(“胆固醇”, “乙醇”)等;
- 6) 提取或加注词类, 包括普通词类(如: 名、动、形、副、量、代、介、助、等)、<姓>、<见>、<同>, 等; 如: “环”③<名>环节, 表示“环”的第三个词义项作名词, 提取<名>填注在词类栏; “带”⑦带动, “带”的第七个词义项没标注词类, 加注<动>在词类栏; “晖”②同“辉”, 表示“晖”的第二个词义项同“辉”, 提取<同>并填注在词类栏;
- 7) 注明文体, 如<书>、<方>、<口>、<古>、<叹>、<译>、<名姓>等: <书>表示书面上的文言词语, 如“掎”, “搦”, “薙”等; <方>表示方言, 如“孖”, “厖”等; <口>表示口语, 如“蹠”, “色”(shǎi)等; <古>表示古代用法, 如“莱”②古时指郊外轮休的天地; <叹>表示感叹词, 多为表感叹的助词, 如“啊”、“咳”等; <译>表音译外来词, 如“酚”、“矾”等; <名姓>表示只用于人名的词, 如“单”(“单于”)、“姆”等;

5 统计结果

5.1 我们考察的词义共计 2016 项。其中有的条目转指其他条目(词典标注为‘见’、‘见下’、‘同’、等), 共 143 项, 皆属重复, 排除不计。另外, 姓氏词义有 181 项, 如: 敖③名姓, 闭④名姓, 段④名姓, 等。姓氏项比较特殊, 本身一般只单不双(一例除外, “单-单于”), 但在实际使用时, 前面可加“老”、“小”“姓”等, 后面可加“兄”、“姐”、“君”等, 所以又可以说姓氏词义项都有弹性。如果排除转指重复项和姓氏项, 其他词义项统计数据如下:

(5) 排除姓氏后词义项统计

词长	数量	比例	举例
只单	600	35.5%	“铐”②
弹性	1032	61.0%	“龟”①-乌龟
只双	60	3.5%	“猓”-“猓猓”
共计	1692	100%	

“只单”表示只单不双, 也就是没有固定的双/多音节形式的词义项。如: “铐”②把犯人铐起来;

“好”①好人，好东西，好事情，好脾气，等。“弹性”即有长短两式的词义项，其长式由人工标注（一般直接取自词义标注）。如：龟①-乌龟、环③-环节、交⑤-结交、贫①-贫穷、绝⑦-绝对、澳①-澳洲、澳②-澳大利亚、等。“只双”表示只双不单的词条，即只用于双/多音节形式而没有其他用法的词条，如：匐-匍匐、迺-迺迺、徊-徘徊、蹒-蹒跚、龊-龊龊、甚-桑甚、等。

结果说明，汉语简单词汇里的‘只单’比例占 35%，高于 Duanmu（2011）所估计的 10-20%。只单比例跟词类也有关系，统计结果见（6）。

（6） 词类及词长弹性统计（排除姓氏）

	只单	只双	弹性	总数	总数比	只单比
名	195	35	503	733	43%	27%
动	183	5	383	571	34%	32%
形	102	18	131	251	15%	41%
副	32	0	4	36	2%	89%
量	30	0	4	34	2%	88%
代	15	0	1	16	1%	94%
介	15	0	0	15	1%	100%
助	12	0	0	12	1%	100%
拟	4	2	4	10	1%	40%
缀	4	0	0	4	0%	100%
连	2	0	2	4	0%	50%
数	3	0	0	3	0%	100%
疑	3	0	0	3	0%	100%
合计	600	60	1032	1692	100%	35%

结果显示，名、动、形三大词类占总数的 92%。其他词类不但数量少，而且多数无弹性，只单不双。副词有弹性的如：仅-仅仅、偏-偏偏、绝-绝对、等。数量词弹性的多为外来词简称，如：亨-亨利、安-安培、吨-英吨、开-开尔文、等。

5.2 普通文体词类的词长弹性比例

特殊文体包括书面体、方言、口语、古、感叹词、音译词等，其他称普通文体。普通文体的统计结果如下：

（7） 普通文体的词类及词长弹性统计

	只单	只双	弹性	总数	总数比	只单比
名	92	27	358	477	41%	19%
动	111	1	322	434	38%	26%
形	61	3	92	156	14%	39%
副	19	0	3	22	2%	86%
量	23	0	2	25	2%	92%
代	9	0	1	10	1%	90%
介	11	0	0	11	1%	100%
助	2	0	0	2	0%	100%

拟	3	1	4	8	1%	38%
缀	4	0	0	4	0%	100%
连	0	0	2	2	0%	0%
数	2	0	0	2	0%	100%
疑	1	0	0	1	0%	100%
合计	338	32	784	1154	100%	29%

比较（6）和（7）的结果，不难看出，普通文体名、动、形三大类的只单比例都有所下降。名词下降的幅度最明显，由27%下降到19%，动词由32%下降到26%，形容词类由41%下降到39%。产生这种变化的原因是特殊文体义项的弹性词长比例大大低于平均值，如：书面体义共计327词，有弹性的仅占50%，方言语体共计59词，有弹性的仅17项。

5.3 常用三千字的统计数据

我们考察的2016个词义项中，属于常用三千字（Da 2004）的共计930义项（排除姓氏、转指重复项），其统计结果如下：

（8）常用三千字的词类及词长弹性统计

	只单	只双	弹性	总数	总数比	只单比
名	63	1	276	340	37%	19%
动	89	0	283	372	40%	24%
形	53	0	87	140	15%	38%
副	18	0	3	21	2%	86%
量	22	0	2	24	3%	92%
代	9	0	1	10	1%	90%
介	11	0	0	11	1%	100%
助	2	0	0	2	0%	100%
拟	0	0	2	2	0%	0%
缀	4	0	0	4	0%	100%
连	0	0	2	2	0%	0%
数	1	0	0	1	0%	100%
疑	1	0	0	1	0%	100%
合计	273	1	656	930	100%	29%

对比（7）和（8），名、动、形三大词类的“只单”比例变化不大。可是，名词的总数却由第一下降到第二，低于动词。这个结果，跟前人的结论有一定出入。比如，尹斌庸（1986）根据统计，提出词类的分布是个“不变量”，无论什么年代，什么文体，名词的数量都是领先，高于动词。那么，本文和前人的出入来自何处呢？我们发现，尹斌庸统计的是词条，有多种词类的词条只选其中之一，比如‘报告’只算动词，不算名词。而本文统计的是词义，‘报告’（动）、‘报告’（名）属于不同词义，一律分开统计。因此，本文为汉语的词类研究，提供了一个不同视角。

6 结语

前人对汉语的词长弹性有过不少观察，但系统的量化研究很少。我们对《现汉》（《现代汉语词典》第五版）简单词的各项词义项做了系统考察，发现有 60-70%的词有弹性长度，其中名词的弹性比例最高，超过 80%。动词的弹性也高于平均比例，但比名词低。

词长弹性研究有助于解释为什么汉语中词很难划界。吕叔湘（1984：422）说，现代汉语的“双音节是占优势的基本语音段落”。不难看出，很多双音节是从单音节扩充而成，因此也能收缩回单音节，可单可双，且意义基本不变，如“虎-老虎”、“店-商店”。长短式之间的区别在于，长式能单用，短式往往不能单用。同时，长式与复合词也不太容易区分，如“老虎”、“煤炭”，看上去像复合词，但是它们的词汇意义不是复合性的（“老虎”≠“老”+“虎”、“煤炭”≠“煤”+“炭”），而是跟相应的短形（“虎”和“煤”）基本一样，因此不是真正的复合词。正确理解汉语的弹性词长，有助于理解为什么“词在两头都有划界问题”（吕叔湘 1984：491-492）。

很多具有弹性长度的词，其长短形都收列在《现汉》中。如“虎”、“店”、“租”、“举”等列在单音节词条下，它们相应的长形“老虎”、“商店”、“出租”、“列举”列在多音节词条下。知道了词长弹性的比例，就可以更精确估算《现汉》中词汇的重复计数。这为研究汉语的词汇，特别是汉语基本词汇提供了方便。

本文也对汉语的词类分布做了统计。我们发现，名词和动词的数量差别没有前人所说那么大。这是因为前人统计的是词条，多词类词条只选其中之一（如‘报告’算动词，不算名词，见尹斌庸 1986）。而本文统计的是词义，不同词类属于不同词义，一律分开统计，如‘报告’（动）、‘报告’（名）是两个词义，分别统计。因此，本文为汉语的词类研究提供了新的视角。

* 本文得到匿名审稿人的有益指点，在此致谢。本文部分内容曾在国际中国语言学学会第 20 届年会、四川大学、香港中文大学、密执安大学等处宣读，得益于与会者的不少建议，也在此表示感谢。

参考文献

- 端木三 2007 重音、信息和语言的分类，《语言科学》第 5 期，3-16 页。
- 冯胜利 2005 《汉语韵律语法研究》，北京：北京大学出版社。
- 冯胜利 2010 论语体的机制及其语法属性，《中国语文》第 5 期，401-412 页。
- 郭绍虞 1938 中国词语之弹性作用，《燕京学报》第 24 期，1-34 页。
- 何元建 王玲玲 2005 汉语真假复合词，《语言教学与研究》第 5 期，11-21 页。
- 柯航 2007 《现代汉语单双音节搭配研究》，中国社会科学院语言研究所博士论文。
- 李临定 1990 动词分类研究说略。《中国语文》第 4 期，248-257 页。
- 刘晓梅 2004 当代汉语新词语的词长考察，《吉林师范大学学报（人文社会科学版）》第 4 期，71-73 页。
- 刘智伟 2007 舍同一语素的同义词；单双音节动词；语体色彩。《语言文字应用》2007.2: 96-104.
- 吕叔湘 1963 现代汉语双音节问题初探，《中国语文》第 1 期，11-23 页。
- 吕叔湘 1984 《汉语语法论文集》，北京：商务印书馆。
- 潘文国 1997 《汉英语对比纲要》，北京：北京语言文化大学出版社。
- 王灿龙 2002 句法组合中单双音节选择的认知解释，《语法研究和探索》第 11 辑，中国语文杂志社编，151-168 页。北京：商务印书馆。

- 王洪君 富丽 2005 试论现代汉语的类词缀, 《语言科学》第 5 期, 3-17 页。
- 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》, 第 4 期, 90-97 页。
- 吴为善 2003 双音化、语法化和韵律词的再分析, 《汉语学习》 2003 年 2 期: 8-14.
- 吴为善 2006 《汉语韵律句法探索》. 上海: 学林出版社.
- 尹斌庸 1986 汉语词类的定量研究. 《中国语文》 1986 年第 6 期: 428-436.
- 中国社会科学院语言研究所词典编辑室 2005 《现代汉语词典》, 北京: 商务印书馆。
- 周荐 1999 从词长看词典语汇单位的确定, 《辞书研究》第 2 期, 34-39 页。
- 周韧 2007 信息量原则与汉语句法组合的韵律模式, 《中国语文》第 3 期, 208-222 页。

- Baayen, R. H, R. Piepenbrock, and L. Gulikers. 1995. *The CELEX lexical database: release 2* (CD-ROM). Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Bloomfield, L. 1926. A set of postulates for the science of language, *Language* 2.3:153-164.
- Chao, Yuen Ren, and Lien Sheng Yang. 1947. *Concise Dictionary of Spoken Chinese*. Cambridge, Mass.: Harvard University Press. [Sixth Printing by Oxford 1961]
- Da, J. 2004 *Chinese text computing*, Murfreesboro: Department of Foreign Languages and Literatures, Middle Tennessee State University, <http://lingua.mtsu.edu/chinese-computing/>.
- Duanmu, S. 2007. *The phonology of Standard Chinese (the 2nd edition)*. Oxford: Oxford University Press.
- Duanmu, S. 2011. How many Chinese words have elastic length? To appear in *Festschrift in honor of Prof. William S-Y. Wang's 80th birthday*. ed. Shi Feng and Feng Gang. .
- Duanmu, S. 2012. Word-length preferences in Chinese: a corpus study, *Journal of East Asian Linguistics* 21.1: 89-114.
- Liu, Feng-hsi. 1992. Verb and syllable in Chinese. Paper presented at the 25th International Conference on Sino-Tibetan Languages and Linguistics, Berkeley.

A Quantitative Study of Elastic Word Length in Modern Chinese

Abstract This paper reports a quantitative study of elastic word length in modern Chinese, using *Xiandai Hanyu Cidian* (A Dictionary of Modern Chinese, fifth edition, 2009). 1/10 of non-compound word entries are sampled, totaling 2,016 word senses. It is found that about 60-70% of all Chinese words have elastic lengths, with the percentage of nouns being the highest, at over 80%, and the percentage of verbs being slightly lower than that of nouns. Stylistics also influences the percentage of elastic word length.

Keywords elastic word length; word entry; word sense; part of speech; lexicon; compound

作者简介:

端木三, 1981-1986 年任教于复旦大学外文系, 1990 年获美国麻省理工学院博士, 1991 年至今任教于美国密歇根大学语言学系, 现为教授, 主要研究音系学, 著有 *The phonology of standard Chinese* (牛津大学出版社 2007 年第二版) 和 *Syllable Structure: The Limits of Variation* (牛津大学出版社 2008 年)。

黄丽君, 女, 1972 年 6 月生, 重庆人, 四川大学外国语学院副教授, 博士在读, 主要从事普通语言学研究, 已发表论文数篇。2011. 8-2012. 8 在美国密歇根大学中国研究中心访学。