现代汉语词类、词长、词义的数量特征分析*

张永伟

【摘　　要】本文以《现代汉语词典》第7版收录的词目为现代汉语基本词汇的代表，统计这些词语的词类、词长、词义的数量特征并进一步探讨这些特征之间的内在关系。从词长来看，基本词类以双音节为主，而叹词、助词、量词等以单音节为主；从词义来看，汉语以单义词为主；从词类来看，名词、动词、形容词是主体，占比超过95%。无论是否区分词性，词长和词量、词长和词义数、词义数和词量均大致呈反比关系。

【关 键 词】《现代汉语词典》　词类　词长　词义　数量特征

【作者简介】张永伟，文学博士，中国社会科学院大学文学院教授，中国社会科学院语言研究所研究员。

【中图分类号】H03　　【文献标识码】A

【文章编号】2097－1125（2024）05－0064－17

词类、词长、词义均是词汇的基本特征，基于统计数据对这些特征进行规律分析，探讨它们的相互关系，有助于加深对汉语词汇整体面貌及其分布规律的认识。同定性研究相比，汉语词汇数量特征分析研究的起步较晚，《现代汉语频率词典》① 的附录统计了多种类型的词长分布，是早期的词汇数量特征分析研究成果之一。权威语文辞书，尤其是《现代汉语词典》②

* 本文系国家社会科学基金一般项目“基于《现代汉语词典》的词汇计量研究”（20BYY170）、国家社会科学基金重大项目“辞书编纂用大型多功能语料库建设与研究”（23&ZD314）的阶段性成果。

① 参见北京语言学院语言教学研究所编：《现代汉语频率词典》，北京语言学院出版社1986年版。

② 参见中国社会科学院语言研究所词典编辑室编：《现代汉语词典》第7版，商务印书馆2016年版。

（以下简称《现汉》）成为很多研究的主要材料。尹斌庸为《现汉》词目标注词类，较早地统计并揭示了汉语词类的数量特征，[①] 但限于当时的条件，在统计时每个词目均只计一个词类，这在一定程度上影响了结论的可靠性。[②] 苏新春专门对《现汉》第2版和第3版的收词、注音、释义、词语性质等进行数量特征分析，[③] 描绘了汉语词汇的基本面貌，但《现汉》第2版距今已出版41年，《现汉》第3版距今已出版28年，其收词已不能反映汉语词汇的最新面貌。王惠利用《现汉》第5版和文本语料对多义词的词长、词义、词频之间的关系进行系统研究，进行了一些很有启发性的分析，[④] 但未将单义词作为考察对象，也未将词类作为考察对象。黄丽君、端木三抽取《现汉》中1/10的单音词对词长弹性进行了数量特征考察，认为弹性词长是汉语的普遍现象，[⑤] 但未对词和语素进行区分，不能不说是一个遗憾。另外，郭锐利用《现代汉语语法信息词典》中43330个词的材料，对各类词的数量特征以及词类在真实语料库中的实际分布进行了考察。[⑥]

《现汉》收录的是现代汉语的基本词、常用词，有极广的流传面和极高的权威性，是致力于反映现代汉语词汇系统的词典，其收录的词目在很大程度上反映着现代汉语词汇的构成与概貌。[⑦]《现汉》从第5版开始标注词性，并区分义项中的词义和语素义，此后又经过两个版本的修订，对收词立目、注音、词类标注、释义等进行完善。[⑧]《现汉》在收词、词类标注、释义等方面的工作堪称典范。[⑨]

本文以《现汉》第7版收录的词目为基本词汇，将词类、词长、词义作

① 参见尹斌庸：《汉语词类的定量研究》，《中国语文》1986年第6期，第428～436页。

② 对兼类词，尹斌庸仅按“本来的词性”进行标注并计数。例如，“报告”既是动词也是名词，而在统计时只计动词，不计名词。

③ 参见苏新春：《关于〈现代汉语词典〉词汇计量研究的思考》，《世界汉语教学》2001年第4期，第39～47页；苏新春等：《汉语词汇计量研究》，厦门大学出版社2002年版；苏新春：《词典与词汇的计量研究》，上海辞书出版社2013年版。

④ 参见王惠：《词义·词长·词频——〈现代汉语词典〉（第5版）多义词计量分析》，《中国语文》2009年第2期，第120～130页。

⑤ 参见黄丽君、端木三：《现代汉语词长弹性的量化研究》，《语言科学》2013年第1期，第8～16页。

⑥ 参见郭锐：《现代汉语词类研究》，商务印书馆2018年版，第311～335页。

⑦ 参见苏新春：《关于〈现代汉语词典〉词汇计量研究的思考》，《世界汉语教学》2001年第4期，第40页。

⑧ 参见江蓝生：《〈现代汉语词典〉第6版概述》，《辞书研究》2013年第2期，第1～19页；谭景春：《词典在不断修订中逐步完善——谈〈现代汉语词典〉第7版条目修订》，《辞书研究》2017年第2期，第1～9页。

⑨ 参见王惠：《词义·词长·词频——〈现代汉语词典〉（第5版）多义词计量分析》，《中国语文》2009年第2期，第120页。

为基本特征，全面分析词类、词长、词义的数量特征及其内在关系。本文与前人研究的不同之处有以下四点：第一，笔者以《现汉》第 7 版为考察材料，希望能更全面、更准确地反映汉语词汇的最新面貌；第二，在基本定词规则中，除了参考词性，根据《现汉》的编纂特点，对异形词、用其他条目或义项做注解的条目或义项等进行处理，使统计的结果更贴近真实情况；第三，在考察时，多义词往往存在多种合理的统计方法，本文相应地给出每种统计结果并加以分析；第四，本文对词类、词长、词义之间的关系进行了更多角度、更细致的统计分析，得出一些新的结论。

一、定词规则

（一）基本定词规则

这里的基本定词规则指从《现汉》中提取词时依据的规则。《现汉》收录的单字条目除了词，还包含不成词的语素和非语素字；多字条目除了词，还包括词组、成语和其他熟语。经统计，《现汉》第 7 版明确收录单字条目 11166 个，多字条目 58700 个，合计 69866 个。《现汉》从第 5 版开始为词标注词类，因此可以通过观察条目是否有义项标注了词类（姓氏义的名词除外）[①] 来判断条目是否是词。这样的基本定词规则与王惠的研究成果一致。[②]

（二）对异形词的处理

《现汉》收录了大量的异形词，规范词形大多正常独立出条，非推荐词形有的不明确出条，有的出条但难以根据基本定词规则准确识别为词。据张永伟统计，《现汉》第 7 版收录“已有国家试行标准”[③] 的异形词 337 组（以下简称规范内异形词）、“国家标准未做规定”的异形词 687 组（以下简称规范外异形词），合计 1024 组。[④] 规范内和规范外的非推荐词形均需要经过一定处理后，才可以使用基本定词规则准确识别。

1. 对规范内异形词的处理

对规范内异形词，《现汉》第 7 版凡例指出，“以推荐词形立目并做注

① 《现汉》将姓氏标注为名词，但只对姓氏用法的现代常用汉字进行标注，并不全面。参见王楠：《中小型语文辞书中姓氏义的收录原则及呈现方式研究》，《辞书研究》2020 年第 4 期，第 16 页。姓氏在用法功能上也明显有别于其他名词，因此本文在统计时忽略姓氏义的名词。

② 参见王惠：《词义·词长·词频——〈现代汉语词典〉（第 5 版）多义词计量分析》，《中国语文》2009 年第 2 期，第 121～122 页。

③ 异形词国家试行标准指《第一批异形词整理表》。

④ 参见张永伟：《〈新华字典〉中异形词收录及使用情况分析》，《辞书研究》2018 年第 2 期，第 30～31 页。

解，非推荐词形加括号附列于推荐词形之后；在同一大字头下的非推荐词形不再出条，不在同一大字头下的非推荐词形如果出条，只注明见推荐词形”。例如：

【搭讪】（搭赸、答讪）dā·shàn 动 为了想跟人接近或把尴尬的局面敷衍过去而找话说。

【答讪】dā·shàn 见 230 页〖搭讪〗。

“搭讪—搭赸、答讪”是规范内异形词。其中，“搭讪”是推荐词形，独立出条并做注解。“搭赸”和“答讪”是非推荐词形，二者均附列于“搭讪”后的括注内，但前者和“搭讪”在同一大字头下，没有独立出条，后者和“搭讪”在不同大字头下，独立出条。《现汉》这样的处理有助于节约篇幅，但有可能导致有的非推荐词形可以被准确识别为词，有的不可以。为避免这种情况，在推荐词形后括注内的所有非推荐词形均应独立出条，并将推荐词形的注解作为非推荐词形的注解。例如，将“搭赸”“答讪”的注解处理为：

【答讪】dā·shàn 动 为了想跟人接近或把尴尬的局面敷衍过去而找话说。见 230 页〖搭讪〗。

【搭赸】dā·shàn 动 为了想跟人接近或把尴尬的局面敷衍过去而找话说。

“答讪”是原有条目，仅补充了注解。“搭赸”是新增条目，注解使用“搭讪”的注解。此时“答讪”“搭赸”均可按基本定词规则被准确识别为动词。

2. 规范外异形词的处理

对规范外异形词，《现汉》第 7 版凡例指出，“以推荐词形立目并做注解，注解后加‘也作某’……非推荐词形如果出条，只注同推荐词形”。比如：

【绿荫】lǜyīn 名 指树荫：～蔽日。也作绿阴。

【绿阴】lǜyīn 同“绿荫”。

“绿荫—绿阴”是规范外异形词，其中，“绿荫”是推荐词形，独立出条并做注解；“绿阴”是非推荐词形，虽然独立出条，但没有标注词类，只注“同‘绿荫’”。根据基本定词规则，“绿阴”未标词类，无法被识别为词。为避免这种情况，非推荐词形的注解应与推荐词形的注解保持一致。例如，将“绿阴”的注解改写为：

【绿阴】lǜyīn 名 指树荫：～蔽日。

在改写后，“绿阴”被标记为名词，此时使用基本定词规则可以准确将其识别为名词。

（三）对用其他条目/义项做注解的条目或义项的处理

在《现汉》中有的条目或义项使用了其他的条目或义项做注解。例如：

楔 xiē❶（～儿）名楔子①②。❷同“揳”。

“楔”字包括两个义项，第 1 个义项的注解使用了“楔子”的前两个义项，第 2 个义项使用了“揳”字条目。“楔子”和“揳”的完整注解如下：

【楔子】xiē·zi 名❶插在木器的榫子缝儿里的木片，可以使接榫的地方不活动。❷钉在墙上挂东西用的木钉或竹钉。❸杂剧里加在第一折前头或插在两折之间的片段；近代小说加在正文前面的片段。

揳 xiē 动把楔子、钉子等捶打到物体里面：榫子缝儿里～上个楔子｜墙上～个钉子｜往地里～根橛子。

“楔子”有 3 个名词义项，“揳”只有一个动词义项。在直接使用基本定词规则时，“楔”将被误识别为只有 1 个名词义项，而结合“楔子”“揳”的注解不难看出，“楔”实际上有 2 个名词义项，1 个动词义项。为了准确识别“楔”的全部义项，在应用基本定词规则前，需将“楔”的注解改写，例如：

楔 xiē❶（～儿）名插在木器的榫子缝儿里的木片，可以使接榫的地方不活动。❷（～儿）名钉在墙上挂东西用的木钉或竹钉。❸动把楔子、钉子等捶打到物体里面：榫子缝儿里～上个楔子｜墙上～个钉子｜往地里～根橛子。

上述例子仅涉及义项和词类的识别。有时，不对注解进行这样的改写，将影响词的识别，例如：

原[3] yuán❶宽广平坦的地方：平～｜高～｜草～｜～野。❷同“塬”。

塬 yuán 名我国西北黄土高原地区因流水冲刷而形成的一种地貌，呈台状，四周陡峭，顶上平坦。

《现汉》第 7 版中“原[3]”没有义项标注词类，因此将不会被识别为词。然而，仔细分析后发现，义项 2 使用“同‘塬’”做注解，而“塬”是名词。因此“原[3]”的义项 2 也应直接使用“塬”的注解全文，从而在应用基本定词规则时，应当被识别为词。

（四）对多义词及同形异义词的识别

同一条目的不同义项，意义不相同但有联系，因此包含多个成词义项的条目，视为多义词。与多义词相关的，是同形异义词。根据凡例，《现汉》对形同音不同的词分立条目，对形同音同但意义需要分别处理的也分立条目。例如：

【草本】[1] cǎoběn 形属性词。有草质茎的（植物）。

【草本】[2] cǎoběn 名文稿的底本。

两个“草本”形同音同，但在意义上需要分别处理，分立为两个条目。

因此，两个“草本”应视为不同的词，而不是多义词。

（五）对多音词的识别

《现汉》为绝大多数词形相同而读音不同的词分立不同条目，它们在使用基本定词规则时，也将被识别为不同的词。例如：

【作乐】zuòlè 动 取乐：寻欢～｜苦中～。

【作乐】zuòyuè 动 ❶制定乐律：制礼～。❷奏乐。

但是，《现汉》为极个别词形相同、词义相同但有不同读音的词只出一个条目，并同时注明不同读音。对这样的词，本文将之视为一个词。例如：

掴（摑）guāi 又 guó 动 用巴掌打：～了他一记耳光。

二、词的基本数量分布

对《现汉》第7版收录的条目及义项进行相应处理后，统计得知《现汉》第7版收录单字条目11166个、多字条目58878个，合计70044个。对这些条目使用基本定词规则进行识别，共得到55812个词、71572个标注了词类的义项，接下来对这些词的词长、词类和词义的基本数量分布进行分析。

（一）词长的基本数量分布

词长通常指词的音节数，[①] 其基本数量分布主要通过词量和比例来体现。词量指词的数量，对多义词而言，每个词的词量既可以是1，也可以是义项数。前者以词为基本单位统计，与义项无关；后者以义项为基本单位统计，与词义相关。在按义项统计时，有的义项还包含子义项，子义项也应视为独立义项。此外，名词“姓”的义项和未标注词类的非词义项均忽略不计。采用两种方法分别统计，词长的基本数量分布情况见表1。

表1　词长的基本数量分布情况

词长	按词统计		按义项统计	
	词量	比例（%）	词量	比例（%）
1	4372	7.83	8247	11.52
2	44964	80.56	56095	78.38
3	6058	10.85	6760	9.45
4	400	0.72	452	0.63
5	15	0.03	15	0.02

① 如果词目包含儿化，则在计算词长时，儿化长度按1计算。

续表

词长	按词统计		按义项统计	
	词量	比例（%）	词量	比例（%）
6	3	0.01	3	0*
合计	55812	100	71572	100

资料来源：《现代汉语词典》第 7 版。下同。

* 实际值为 0.00004192，保留小数点后两位为 0。

两种统计方法均显示《现汉》第 7 版的词长介于 1 到 6 之间，二字词数量最多，三字词和单字词数量次之，三者合计占比超过 99%。长度大于 3 的词，数量极少，合计占比不足 1%。此外，对多字词而言，词长和词量呈反比关系，即词长越长、词量越小，反之亦然。

两种统计方法的主要不同在于，在以词为基本单位统计时，三字词数量多于单字词；而在以义项为基本单位统计时，单字词数量多于三字词。造成这种差异的原因是单字词的义项数量远多于三字词，故在以义项为基本单位统计时，义项多的单字词比义项少的三字词更多次地被重复统计。

（二）词类的基本数量分布

《现汉》标注了 12 种词类，分别为名词、动词、形容词、数词、量词、代词、副词、介词、连词、助词、叹词、拟声词，与《中学教学语法系统提要（试用）》的分类相同。①

假设多义词有 n 个义项词性相同，则在统计时该词性对应的词量可以是 1，也可以是 n，前者以词形为基本单位统计，与 n 的大小无关；后者以相同词性的义项为基本单位统计，与 n 的大小相关。当多义词标有多种不同词性时，每种词性的词量均应单独统计。采用两种方法分别统计，词类的基本数量分布情况见表 2。

表 2　词类的基本数量分布情况

词类	按词形统计		按义项统计	
	词量	比例（%）	词量	比例（%）
名词	30753	51.76	36292	50.71
动词	19684	33.13	24386	34.07
形容词	6157	10.36	7602	10.62

① 参见人民教育出版社中学语文室编：《中学教学语法系统提要（试用）》，人民教育出版社 1984 年版。

续表

词类	按词形统计		按义项统计	
	词量	比例（%）	词量	比例（%）
副词	1254	2.11	1452	2.03
量词	500	0.84	543	0.76
拟声词	270	0.45	286	0.40
连词	225	0.38	243	0.34
代词	181	0.30	264	0.37
介词	112	0.19	147	0.21
助词	105	0.18	163	0.23
叹词	97	0.16	105	0.15
数词	72	0.12	89	0.12
合计	59410	99.98	71572	100.01

注：由于四舍五入的原因，表中所有百分比之和可能不等于100.00%、与100.00%存在微小差异。这种差异不影响数据的有效性和结论的可靠性，特此说明。表3同。

可以看到，两种方法统计的不同词类的词量分布情况较为一致：不同词类的词量差异很大，名词、动词、形容词是主体，占比超过95%，其他9类词不足5%。两种统计结果的不同之处在于按义项统计时，代词的词量大于连词，助词的词量大于介词，而按词形统计时恰好相反。代词和连词的词量大小存在差异，是由于代词和连词的词量比较接近，而代词的义项数比连词多。助词和介词的词量大小差异也是由同样原因导致的。

（三）词义的基本数量分布

词义的基本数量分布主要通过不同词义数的词量和比例来体现，词义数按词的义项数来统计。经统计，词义的基本数量分布情况见表3。

表3　词义的基本数量分布情况

词义数	词量	比例（%）	占多义词比例（%）
1	44195	79.19	—
2	9173	16.44	78.96
3	1637	2.93	14.09
4	429	0.77	3.69
5	184	0.33	1.58
6	81	0.15	0.70
7	46	0.08	0.40

续表

词义数	词量	比例（%）	占多义词比例（%）
8	26	0. 05	0. 22
9	13	0. 02	0. 11
≥10	28	0. 05	0. 24
合计	55812	100. 01	99. 99

可以看到，在汉语中 79. 19% 的词都是单义的，也就是说，大部分的词只表示一个意义。而多义词中又以两个义项的多义词为主，占多义词的 78. 96% 。所以，尽管《现汉》第 7 版收录的义项数比词的数目多了 15760 个，但这大部分是由双义词贡献的。一个词承载的意义不宜过多，这才能保证词在语境中被迅速识别和理解。此外，基于表 3 的结果进一步统计得到，汉语词的平均词义数为 1. 28，单义词和多义词的数量比值为 3. 8。表 3 同时显示，词义数越大，词量越小，二者呈反比关系。

三、词类、词长、词义的关系

（一）词类、词长的数量关系

不同词类、不同词长的词量统计，依然有按词统计和按义项统计两种方法，不同方法的统计结果分别见表 4 和表 5。

表 4　不同词类、不同词长的词量（按词统计）

词类	词长							
	1	2	3	4	5	6	合计	均值
名词	1875	23387	5221	253	14	3	30753	2. 13
形容词	446	5207	388	115	1		6157	2. 03
动词	1852	17435	388	9			19684	1. 93
副词	285	868	85	16			1254	1. 87
拟声词	92	161	2	15			270	1. 78
连词	56	162	7				225	1. 78
代词	73	85	20	3			181	1. 74
介词	74	38					112	1. 34
数词	52	18	2				72	1. 31
量词	387	91	22				500	1. 27
助词	80	24	1				105	1. 25
叹词	83	13	1				97	1. 15
合计	5355	47489	6137	411	15	3	59410	—

“汉语词以双音节为主”已经是学界的共识，但通过对不同词类词长的分析可以看到，不同词类的词长还是存在一定差异的。表 4 和表 5 同时显示，词长均值最接近双音节的是名词、形容词和动词，而叹词、助词和量词的词长均值更接近于单音节。由于名词、形容词、动词是汉语词类的主体，占比 95% 以上，故而它们的词长决定了汉语词长的基本面貌。所以，更准确地说，汉语的基本词类以双音节为主，而叹词、助词、量词等则仍以单音节为主。

表 5　不同词类、不同词长的词量（按义项统计）

词类	词长							
	1	2	3	4	5	6	合计	均值
名词	2520	27773	5711	271	14	3	36292	2. 1
形容词	779	6239	448	135	1		7602	1. 99
动词	3470	20455	452	9			24386	1. 88
副词	379	964	90	19			1452	1. 83
拟声词	94	175	2	15			286	1. 78
连词	66	170	7				243	1. 76
代词	115	122	24	3			264	1. 68
介词	105	42					147	1. 29
数词	66	21	2				89	1. 28
量词	427	94	22				543	1. 25
助词	136	26	1				163	1. 17
叹词	90	14	1				105	1. 15
合计	8247	56095	6760	452	15	3	71572	—

在三个基本词类中，名词不仅数量最多，而且平均词长明显大于其他词类，同时也是三字词数量多于单字词的唯一词类。从词长的角度看，长度大于 2 的词，以名词为主，占比超过半数；长度大于 4 的，形容词有 1 个（可惜了$_{儿}$的），名词有 17 个（哀的美敦书、爱斯基摩人、布尔什维克、叽里旮旯$_{儿}$、纪事本末体、居民身份证、柯尔克孜族、克里姆林宫、萨噶达娃节、手指头肚$_{儿}$、乌孜别克族、无后坐力炮、小手工业者、伊斯兰教历、阿尔茨海默病、达摩克利斯剑、英特纳雄耐尔）。名词表示事物，世界上的事物千差万别，新的事物也在不断产生。事物的不同特征需要反映在名词中，这样表示 A 事物的名词才能与表示 B 事物的名词区分开来，不影响人们交际。

动词的三字词含儿化数量较多，有 170 个。另外，“双音名词 + 词缀”的构词方式也产生了不少三字动词，仅以“化”结尾的动词就有 33 个，如“城镇化、概念化、人格化、沙漠化”等。此外，将“V 得 + 补”“V 不 + 补”作为动词的数量也较多，分别有 24 个、49 个，如“吃得来、对得住”

“吃不来、对不住”等。

形容词的三字词、四字词大多是表描摹的状态词。重叠是用来表示描摹的重要手段，是造成形容词三字词、四字词数量较多的主要原因。例如，在按词统计的 388 个三字词中，状态词有 228 个，而在状态词中 ABB 模式的有 218 个，如“碧油油、沉甸甸、乐滋滋、香喷喷”等。在按词统计的 115 个四字形容词中，有 90 个是状态词，而在状态词中 AABB 模式的有 63 个，如“沸沸扬扬、轰轰烈烈、堂堂正正、熙熙攘攘”等。一般的成语词典也收录了这些词，所以它们是词还是语其实是可以讨论的。

（二）词类、词义的数量关系

假设多义词有 n 个义项词性相同，总义项数有 m 个（1≤n，n≤m，等号不同时成立），则在统计时该词性对应的词义数可以是 1、n 或者 m。在按 1 计算时，某个词的词类对应的词义数始终为 1，词类与词义的数量分布等同于词类在按词统计时的数量分布；在按 n 计算时，词义数等于与该词性相同的义项数，即排斥异类义项；在按 m 计算时，词义数等于整个词包含的所有义项数，即包含异类义项。例如，“楔”有 2 个名词义项，1 个动词义项，总义项数为 3，则在统计名词的义项数时，上述三种统计方法分别对应 1、2、3。后两种统计方法的区别在于，当多义词只有一种词性时（n = m），统计结果完全一致；当多义词为兼类词时，按 m 计算的词义数大于按 n 计算的词义数，多出的是其他词性对应的义项数。由此可见，当考察词类与词义的关系时，只统计与该词类相同的义项数更为合理。

1. 不同词类的词义数和词量呈反比关系，当词义数和词类数均较大时尤为明显。

统计不同词类、不同词义数的词量，有助于详细分析每个词类的词义数量分布。不同词类、不同词义数的词量详情见表 6。

表 6　不同词类、不同词义数的词量

词类	词义数										
	1	2	3	4	5	6	7	8	9	≥10	合计
名词	26162	3875	562	107	30	9	4	3		1	30753
动词	16277	2680	476	127	54	25	23	7	5	10	19684
形容词	5046	878	170	46	8	2	3	3	1		6157
副词	1105	124	15	4	1	3	1	1			1254
量词	462	34	3	1							500
拟声词	257	11	1	1							270

续表

词类	词义数										
	1	2	3	4	5	6	7	8	9	≥10	合计
连词	210	13	1	1							225
代词	130	33	9	6	2		1				181
介词	91	14	3	2	1	1					112
助词	77	14	6	3	3	1	1				105
叹词	91	4	2								97
数词	63	6	1	1			1				72
合计	49971	7686	1249	299	99	41	34	14	6	11	59410

表 6 最末行数据显示，在区分词性统计时，词义数和词量之间依然整体呈现反比关系，词义数越大，词量越小。表 6 最末行数据和表 3 数据略有不同，前者在统计时区分词性，后者不区分。但无论是否区分词性，统计结果都显示了词义数和词量之间的反比关系，说明这种反比关系具有稳定性。但是，具体而言，并非所有词类都满足这种反比关系。例如，表 6 显示，形容词义项数为 6 的词少于义项数为 7 和 8 的词，副词义项数为 5 的词少于义项数为 6 的词等。因此，词义数与词量的关系，还应补充以下两点：一是名词、量词等的词义数和词量呈反比关系，词义数越小，反比关系越明显、越稳定；二是当词义数较大且词量较小时，词义数和词量之间的反比关系就会受到挑战。一种关系的呈现需要一定的数据量，而当词义数较大时，各词类的词量通常都非常小，以个位数为主，上述这种反比关系只能大致呈现，难以验证。

2. 汉语各词类均以单义词为主

将不同词类的多义词作为整体同单义词进行对比，有助于考察不同词类的单义词、多义词整体数量差异。不同词类的单义词、多义词数量及二者的比值见表 7。

表 7　不同词类的单义词、多义词数量及二者的比值

词类	单义词数量	多义词数量	二者的比值
拟声词	257	13	19.77
叹词	91	6	15.17
连词	210	15	14.00
量词	462	38	12.16
副词	1105	149	7.42
数词	63	9	7.00
名词	26162	4591	5.70

续表

词类	单义词数量	多义词数量	二者的比值
动词	16277	3407	4. 78
形容词	5046	1111	4. 54
介词	91	21	4. 33
助词	77	28	2. 75
代词	130	51	2. 55
合计	49971	9439	5. 29

表 7 显示，所有词类的单义词数量均大于多义词数量，这进一步明确了汉语各词类均以单义词为主的事实。其中，尽管代词的单义词与多义词的数量比值最小，但单义词数量依然是多义词数量的 2. 55 倍。拟声词、叹词、连词、量词的单义词数量甚至是多义词数量的 10 倍以上。按单义词与多义词的数量比值从大到小排列，名词、动词、形容词分别居于第 7、8、9 位，处在中后段，比值在 4. 54 ~ 5. 7 区间内。

需要指出的是，表 7 统计得到的单义词与多义词的数量比值为 5. 29，而根据前文统计得到单义词与多义词的数量比值为 3. 8。这是因为，前文在统计时不区分词类，而此处在统计时区分词类，只考虑和目标词类完全相同的义项。对非兼类词和单义词而言，两种方法的统计结果相同，但对兼类的多义词而言，前者统计的多义词数量大于后者，因为非目标词类的义项也被统计，导致多义词的数量增加，从而单义词与多义词的数量比值降低。

3. 不同词类的平均词义数最小为 1. 06，最大为 1. 55

对词类平均词义数的分析有助于考察不同词类的词义数的整体差异。将各词类的平均词义数按从大到小排列，结果见图 1。

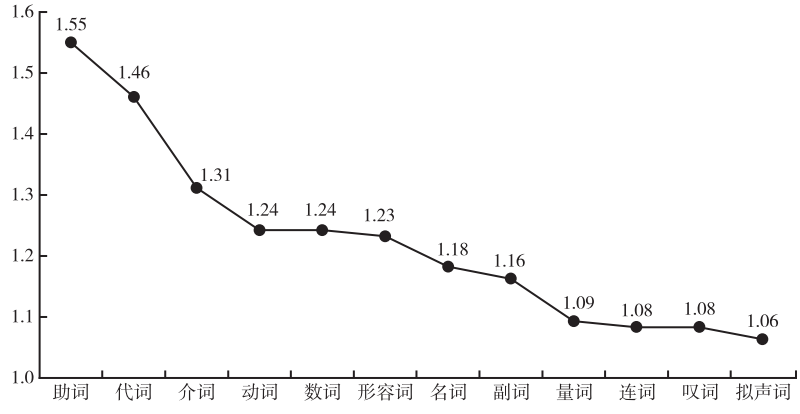

图 1　各词类的平均词义数

图 1 显示，不同词类的平均词义数分布不均匀。助词、代词、介词的平均词义数较大，均在 1.3 以上，但平均词义数最大的助词的该数值也只有 1.55；拟声词、叹词、连词、量词的平均词义数较小，均小于 1.1；动词、数词、形容词、名词、副词的平均词义数居中，在 1.1～1.3 区间内。

需要指出的是，根据前文统计得到汉语词平均有 1.28 个词义，那是在不区分词性的情况下统计得到的，而图 1 的词义数在统计时只考虑和目标词类完全相同的词义，忽略和目标词类不同的词义。如果在统计时考虑词性因素，则汉语词的平均词义数为 1.22，这个数值主要是由动词、形容词、名词决定的，因为三者的词量占比 95% 以上，而且其平均词义数居于中段，在 1.1～1.3 区间内。

4. 汉语词的多义性经常以不同词类呈现

统计词的兼类数和词义数，有助于考察二者之间的关联，详细数据见表 8。

表 8　不同兼类数和不同词义数的词量

兼类数	词义数											兼类数和词义数相等的词占比（%）
	1	2	3	4	5	6	7	8	9	≥10	合计	
1	44195	6926	1033	220	69	26	13	8	3	5	52498	84.18
2		2247	517	159	76	31	17	11	5	7	3070	73.19
3			87	45	33	19	12	4	3	8	211	41.23
4				5	6	4	3	2	2	4	26	19.23
5						1	1	1		4	7	0
合计	44195	9173	1637	429	184	81	46	26	13	28	55812	83.38

兼类数大于 1 的为兼类词。表 8 显示，兼类数为 2、词义数为 2 的词，即兼类数和词义数均为 2 的兼类词为 2247 个，占所有兼 2 个词类的词的 73.19%，即 73.19% 的双义词有不同词类表现。更全面地，兼类数和词义数相等的兼类词共计 2339 个，占全部兼类词（3314 个）的 70.58%，即超过 70% 的多义词均有相等数量的词类表现，这充分说明兼类对词的多义贡献是非常大的。

此外，随着兼类数的增加，兼类数和词义数相等的词占相同兼类数的词的比例逐渐降低。换言之，一个多义词的词类种数越多，出现相同词性词义的可能性也就越大。

（三）词长、词义的数量关系

统计不同词长、不同词义数的词量，可以更细致地考察不同词长的词义数的数量分布以及不同词义数下词长的数量分布。不同词长、不同词义数的词量见表 9。

表 9　不同词长、不同词义数的词量

词长	词义数											
	1	2	3	4	5	6	7	8	9	≥10	合计	均值
1	2687	802	364	220	124	68	40	26	13	28	4372	1.89
2	35715	7751	1217	203	59	13	6				44964	1.25
3	5422	578	51	6	1						6058	1.12
4	353	42	5								400	1.13
5	15										15	1
6	3										3	1
合计	44195	9173	1637	429	184	81	46	26	13	28	55812	1.28
均值	2.08	1.98	1.81	1.50	1.33	1.16	1.13	1	1	1	—	—

表 9 中词义数大于等于 10 的 28 个词均是单字词，词义数的最大值为 24。更详细地，词义数为 10～18 的单字词数量分别为 7、7、3、3、3、3、0、0、1 个，词义数为 24 的单字词只有 1 个，即“打”。

由前文可知，汉语多字词的词长和词量整体上呈反比关系，表 9 显示这一关系不受词义数的影响。换言之，不同词义数的多字词的词长和词量整体上呈反比关系。同样由前文可知，汉语词的词义数和词量也呈反比关系，表 9 显示这一关系不受词长的影响。换言之，不同词长的词义数和词量均呈反比关系。例如，长度为 3 的词，单义词为 5422 个，数量最多，而当词义数逐步增加到 2、3、4、5 时，词量逐步减少为 578、51、6、1 个。

此外，表 9 显示，词长越长，词的平均词义数越小，二者大致呈反比关系，例外的是词长为 4 的词的平均词义数为 1.13，比词长为 3 的词的平均词义数 1.12 大。在所有长度的词中，只有单字词的平均词义数大于平均值，为 1.89。词长为 5 和 6 的词，平均词义数最小，均只有 1 个词义。表 9 还显示，词义数越大，平均词长越短，二者呈反比关系。单义词的平均词长最长，为 2.08；超过 7 个词义的词的平均词长最短，全部为单字词。

（四）不同词类的词长、词义数分布

前文指出，词长和词的平均词义数大致呈反比关系。接下来将按不同词类分别统计不同词长的平均词义数，以观察哪些词类的词长和平均词义数呈反比关系，哪些没有呈现，统计详情见表 10。

表 10　不同词类、不同词长的平均词义数

词类	词长						
	1	2	3	4	5	6	均值
名词	1.34	1.19	1.09	1.07	1	1	1.18
动词	1.87	1.17	1.16	1			1.24

续表

词类	词长						
	1	2	3	4	5	6	均值
形容词	1.75	1.2	1.15	1.17	1		1.23
数词	1.27	1.17	1				1.24
量词	1.1	1.03	1				1.09
代词	1.58	1.44	1.2	1			1.46
副词	1.33	1.11	1.06	1.19			1.16
介词	1.42	1.11					1.31
连词	1.18	1.05	1				1.08
助词	1.7	1.08	1				1.55
叹词	1.08	1.08	1				1.08
拟声词	1.02	1.09	1	1			1.06
平均词义数均值	1.89	1.25	1.12	1.13	1	1	1.28

表10显示，名词、动词、数词、量词、代词、介词、连词、助词、叹词的词长和平均词义数之间均呈严格的反比关系，而形容词、副词、拟声词是例外。词长为4的形容词、副词的平均词义数大于词长为3的，这是导致这两类词的词长和平均词义数未呈严格反比关系的主要原因。如前文所述，这部分条目是词还是语是可以讨论的。

（五）不同词类的词义、平均词长分布

前文发现词义数和平均词长呈反比关系，本部分将按不同词类分别统计不同词义数的平均词长，以观察是否所有词类的词义数和平均词长之间均呈反比关系，统计详情见表11。

表11　不同词类、不同词义数的平均词长

词类	词义数										
	1	2	3	4	5	6	7	8	9	≥10	均值
名词	2.15	2.05	1.91	1.72	1.6	1.56	1.25	1		1	2.13
形容词	2.05	1.97	1.82	1.57	1.25	1	1	1	1		2.03
动词	1.95	1.88	1.64	1.33	1.24	1.08	1	1	1	1	1.93
副词	1.89	1.76	1.4	1.25	1	1	2	1			1.87
连词	1.8	1.62	1	1							1.78
拟声词	1.77	2	1	2							1.78
代词	1.81	1.64	1.44	1.33	1.5		2				1.74
介词	1.38	1.14	1.33	1	1	1					1.34
数词	1.3	1.5	1	1			1				1.31
量词	1.29	1.09	1	1							1.27
助词	1.31	1.14	1	1	1	1	1				1.25
叹词	1.15	1.25	1								1.15
平均词长均值	2.08	1.98	1.81	1.50	1.33	1.16	1.13	1	1	1	—

表 11 显示，虽然词义数和平均词长在整体上呈反比关系，但具体到各个词类，只有名词、形容词、动词、副词、连词、量词、助词的词义数和平均词长呈严格的反比关系，其余 5 个词类（拟声词、代词、介词、数词、叹词）的词义数和平均词长只是大致呈反比关系，均存在例外情况。但是，由于名词、动词、形容词占比高达 95% 以上，故而词义数和平均词长在整体上的关系主要受它们的影响，依然呈反比关系。

四、结语

本文在对《现汉》词语的词类、词长和词义等基本特征进行统计的基础上，分析了三者之间的内在关系，主要得出以下五点结论。

第一，汉语词整体上以双音节为主，其中基本词类以双音节为主，而叹词、助词、量词等以单音节为主。无论是否区分词性，多字词的词长和词量整体上均呈反比关系。

第二，在各词类中，名词、动词、形容词是主体，占比超过 95%。

第三，汉语以单义词为主，占比高达 79.19%，在多义词中又以两个义项的多义词为主，占多义词的 78.96%。在统计时无论是否区分词性，所有词类的单义词词量均多于多义词词量，各词类的平均词义数在 1.06 ~ 1.55 区间内。此外，汉语词的多义性经常以不同词类呈现，兼类对词的多义性贡献是非常大的。

第四，词长和平均词义数大致呈反比关系。其中，形容词、副词、拟声词的词长和平均词义数没有呈现严格的反比关系，其余词类均呈严格的反比关系。词义数和平均词长呈反比关系，其中拟声词、代词、介词、数词、叹词的词义数和平均词长没有呈现严格的反比关系，其余词类均呈严格的反比关系。

第五，无论是否区分词性，词义数和词量整体上均呈反比关系。其中，形容词、副词大致呈反比关系，仅存在个别例外情形，其余词类的词义数和词量均呈严格的反比关系。

表中数据较为详细，但限于文章篇幅，本文仅对表中数据进行了概要分析，分析得还不够细致。此外，本文未结合真实语料库中统计的词频进行分析，也未对兼类词的数量特征进行分析。词频也是词汇的重要特征之一，关于词频与词性、词长、词义之间的数量关系特征以及兼类词的数量特征将另文讨论。

（责任编辑：陈华积）