

# 基于数据库定量分析的汉语学术词汇研究

刘鑫民, 刘畅

**摘要** 不同的汉语学习者需要学习的词汇的范围是有区别的, 因此为不同目的把汉语作为第二语言的学习者制定不同的词汇表是汉语词汇教学研究的一个有重要价值的课题。汉语学术词汇表是学术汉语教材编写和学术汉语教学的基础, 同时学术汉语词汇表也可以帮助以学术为目的的汉语学习者进行针对性的词汇学习。文章分析了汉语学术词汇的特点, 制定了汉语学术词汇的筛选原则, 以自建的 1500 万字的汉语学术文献数据库的词频统计为基础, 根据学术词汇在学术性、常用性、通用性方面的具体标准, 选定了 669 个汉语学术词汇, 制定了汉语学术词汇表。按照其使用频率的高低, 文章把汉语学术词汇分为最常用学术词汇、常用学术词汇、次常用学术词汇三类。同时, 根据通用性强弱, 两个学科门类以上的文献中高频出现的词归为通用词汇, 只在一个学科门类中使用的学术词汇作为次通用词汇。

**关键词** 学术汉语; 学术词汇; 词频统计; 语料库

## 一、引言

词汇学习是汉语学习的一个重要方面, 不同目的的汉语学习者, 对于词汇的学习具有不同的选择性。因此, 为不同目的的汉语学习者制定不同的词汇表, 是汉语教学研究中一个非常有意义的领域。在过去的二十年中, 随着学术英语教学和研究的迅速发展, 英语学术词汇研究取得了许多令人瞩目的研究成果, 但是由于学术汉语的教学需求不足, 因此学术汉语的教学与研究尚未起步, 汉语学术词汇的研究也还是一个空白。近些年来, 在大学进行专业学习, 以攻读学位、进行学术研究为目的的汉语学习者数量越来越多, 学术汉语的研究、学术汉语教材的编写已经提到了汉语教学研究的议事日程上。而要编写学术汉语教材, 制定汉语学术词汇表是一个基础研究。因此, 本文的目的就是对汉语的学术词汇进行筛选, 提出一个汉语学术词汇表, 为学术汉语教材的编写和学术汉语的教学提供参考。

## 二、研究背景和理论基础

学术词汇的研究肇始于专门用途词汇的研究。最早的学术词汇的研究者是从事学术英语研究的一些学者。学术英语是专门用途英语的一个分支。在早期的专门用途英语教学中, 研究者和教师比较偏重技术词汇(专业词汇)的教学, 但是在后来的研究和教学实践中发现, 准专业词汇的教学更为重要。<sup>1</sup> Nation (2001) 从词频和词汇覆盖率的角度将特殊用途英语学习者的词汇需求分为四类: 高频词 (high-frequency words)、学术性词汇 (academic words)、技术性词语 (technical words) 及低频词 (low-frequency words)。相关研究表明, 在学术阅读和学术写作中, 给以学术为目的的学习者造成困难的词汇既不是在各种学科中使用频率非常高的通用词汇, 也不是与专业学习内容高度相关、词义比较精确的专业词汇, 而是使用频率和专业程度介于通用

1 Smoak R. *What is English for Specific Purposes?* English Teaching Forum, 2003, 41(2):23.

的高频词汇和在使用范围上比较狭窄的专业词汇之间的学术词汇。<sup>2</sup>同时, Coxhead (2000) 的研究表明, 学术词汇在学术文献中的覆盖率很高, 其学术词汇表中的 570 个词族覆盖了约 10% 的学术文本, 而这些词汇在普通文献中的覆盖率远低于学术文献, 比如在新闻文献中这些词汇的覆盖率为 4.5%, 而在小说中的覆盖率更低至 1.4%。这说明这些学术词汇在通用语言学习中难以完全通过随机方式习得。正因为这样, 编制学术词汇表, 让以学术为目的的语言学习者进行针对性的词汇学习, 显得特别重要。同时, 有了这样的学术词汇表, 我们才能够明确, 在对以学术为目的的语言学习者进行词汇教学时, 哪些词汇最值得占用宝贵的学习者自主学习时间和教师的课堂教学时间。

Xue & Nation (1984) 发布了《大学词汇表》(The University Word List), 简称为 UWL, 筛选了 808 个学术文献中常用的词族, 这是英语学术词汇表研究的雏形。Coxhead (2000) 通过自建的 350 万个词标的学术英语语料库, 筛选了 570 个学术词族, 提出了在英语学术词汇研究中非常有影响的学术词汇表 (Academic Word List), 简称为 AWL。AWL 词汇筛选的标准主要有三个:

标准一: 入选词族是在学术文献中高频出现但又不是英语普通文献中的 2000 常用词汇;

标准二: 每个词族里的单词应该在其自建语料库的所有 4 个学科的文里出现,<sup>3</sup> 而且在每个学科中至少出现 10 次, 同时这些词应该至少出现在 15 个不同专业的文献中;

标准三: 每个词族里的单词在其自建的语料库中出现的频次不少于 100 次。

AWL 强调了学术词汇的通用性, 但是一些研究者也认为, 这种通用性的学术词汇在不同学科中的出现频率并不均衡, 在某一学科中出现频率非常高的词汇在另一学科中则可能很少使用, 因此, 既需要研制涵盖多学科的通用学术词汇表, 也需要针对不同学科的特点, 研制在某一范围内针对某一学科的学术词汇表。

### 三、学术词汇的特点

二语习得中的阅读研究表明, 要理解阅读材料, 认识阅读材料中 95% 的词是最低临界值<sup>4</sup>, 有学者指出, 学术词汇在词汇中所占的比例大约是 16%,<sup>5</sup> 因此, 如果不掌握学术词汇要想读懂学术文献是不可能的。学术词汇是汉语词汇中一个重要的组成部分, 这些词汇对于汉语学习已经达到了比较高的水平、将要运用汉语进行专业学习、需要阅读汉语学术著作和进行汉语学术写作的学习者尤为重要。学术词汇是进行学术阅读和学术写作时非常重要的一部分词, 如果不掌握这一部分词, 要读懂学术文献就不太可能, 也不可能写出地道的学术文章, 因此要进入专业学习和进行学术写作, 学习者就必须掌握这部分词汇。

学术词汇是指那些在学术性著作中通用、在各个专业学科中普遍使用的词汇, 这些词汇也可称作“半技术性词汇”或者“半专业词汇”。学术词汇不是纯粹的专业词汇, 专业词汇是只在某个

---

2 参见 Lam J, *A study of semi-technical vocabulary in computer science texts, with special reference to ESP teaching and lexicography*, Research reports, Vol. 3 Language Center, Hong Kong University of Science & Technology, 2001.

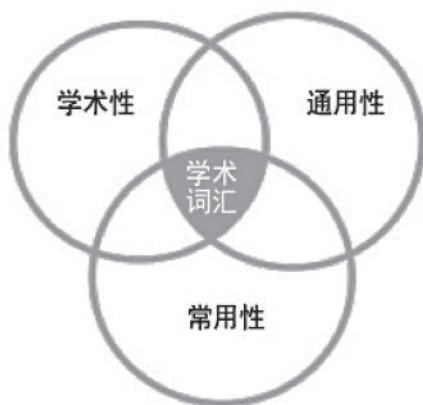
3 Coxhead 的语料库中没有包括工科的文献, 这一点也是其学术词汇表受到质疑的一个方面。如果其语料库中包括了工科的文献, 那么是不是这些词在所有学科中都能达到其要求的出现频率恐怕就值得怀疑了。

4 Laufer, B.1988. *What percentage of lexis is necessary for comprehension?* In C. Lauren & M.Norman (eds.). *Special Language: From Humans to Thinking Machines*. Clevedon: Multilingual Matters. 316-323.

5 Santos.M.2000. *Analyzing academic vocabulary and contextual cue support in community college textbooks*. Unpublished qualifying paper. Harvard: Harvard Graduate School of Education. <http://www.ncsall.net> (accessed 12/03/2006).

学科中使用的术语，它的使用范围较窄，其内涵和外延通常有清晰的界定，而且多为单义词。专业词汇（术语）的学习需要结合专业知识的学习来完成，而学术词汇是在各个专业文献中具有通用性和常用性的词汇。我们认为学术性、通用性、常用性是学术词汇同时具有的三个特点。学术词汇的学术性和常用性的表现是：这一类词语在学术文献中使用频率大大高于在普通文献中的使用频率。学术词汇应该是这样一类词：它们在学术文献中常用，使用频率很高，但在普通文献中的使用频率并不是特别高，远不及在学术文献中的使用频率，这一特点使得学术词汇不同于一般的常用词。学术词汇的通用性是指学术词汇在学术文献中具有跨学科性，在不同学科中都具有高频率使用的特点，学术词汇的通用性使得这一类词不同于使用范围比较狭窄的专业词汇。<sup>6</sup>

下图反映了学术词汇具有的特点：



#### 四、研究方法和步骤

本文主要采用定量分析的方法，对大型汉语学术文献语料库进行词频统计，将得到的统计结果和普通文本语料库的词频统计结果进行比较，并辅以人工干预手段，筛选出学术文献中高频使用并具有跨学科通用性质的学术词汇，从而制定汉语的学术词汇表。

基本的研究步骤如下：

##### （一）学术语料库的构建

学术语料库是研究筛选学术词汇、制定学术词汇表的基础。鉴于目前没有现成的大型汉语学术文献语料库，我们自建了1500万字的汉语学术文献语料库。选入语料库的文本主要是各个专业的教材以及讲义，也包括一定量的学术专著。这些文本涵盖21个学科，我们按学科的近似度把它们分为哲学法律、管理经济、人文社科、理工农医四个门类，学科门类的划分主要是为了在词频统计的基础上分析学术词汇通用性的强弱。

下表为我们自建的学术文献语料库涵盖的学科以及每个学科收入的文本的字数。<sup>7</sup>

6 当然根据学术词汇表的研究目的和适用对象，对通用性范围可以做不同的界定，比如通用范围可以界定为人文学科或者理工科等，从而制定出针对人文学科的学术词汇表、针对理工科的学术词汇表等。

7 收录的各个学科的文本数量，我们尽量做到了在各个学科门类之间的平衡，但是由于理工农医类的学术文本中包含了大量公式和运算，这些内容作为符号没有统计在文本字数内（之所以统计字数时没有包括符号，是

门类	学科	字数	分门类字数合计
哲学法律	政治学	717898	4621282
	国际关系	804390	
	哲学	2057802	
	宗教	141356	
	法学	539836	
管理经济	管理学	701976	3500279
	会计学	114759	
	经济学	2683544	
人文社科	语言文学	1004039	4283605
	心理学	1260806	
	教育学	317019	
	新闻学	248132	
	历史学	993932	
	文化学	356325	
	逻辑学	103352	
理工农医	物理学	340284	2696571
	化学	498465	
	数学	76230	
	计算机技术	118379	
	科技史	1158666	
	中医学	304919	
	工学	199628	
字数合计			15101737

## (二) 学术词汇筛选原则的制定

根据学术词汇的特点，制定学术词汇表时，学术词汇筛选应该把握的基本原则是：进入学术词汇表的词应该是在学术文献中具有学术性、常用性、通用性的词。具体来讲，入选学术词汇表的词应该是：(1) 具有学术性的词。这些词在学术文献中专门出现，其在学术文献中的使用频率要远高于在普通文献中的使用频率。因为在学术文献中专门出现，决定了这些词不包括同样也在一般文献中使用频率很高的那部分通用词汇，换句话说，在语言教学中，那部分高频通用词汇的教学是基础通用汉语教学的任务，而不是学术汉语的教学任务。因为学术词汇不是在一般文献中高频使用的通用词，因此，这些词汇对学习来说通常都是在通用汉语学习中不太容易习得、有一定难度的词。(2) 在学术文献中具有常用性的词。学术词汇是在学术文献中常用、使用频率非常高的一类词。(3) 在不同学科中具有通用性的词。学术词汇在不同学科及不同学科的不同专业中具有很强的通用性。当然，根据学术词汇的筛选范围，学术词汇通用性的最高标准是学术词汇应该在各个学科中都具有较高的使用频率（这部分词是通用学术词汇）；最低标准是在一个学科大类中各个专业的文献中通用的词（针对某个学科的学术词汇）。根据其通用性的强弱，我们可以制定适应不同学科范围的学术词汇表。

---

因为学术词汇的筛选目标是词，不是符号)。如果包括符号，理工农医类的学术文本和其它三个门类的文本数量没有大的区别。

### (三) 词频统计软件的选择及其效度、信度分析

本研究使用中国国家语委开发的词频统计软件 CorpusWordFrequencyApp 作为统计工具。

汉语书面语由于词和词不分写，在用计算机软件进行汉语词频统计时，词的切分是一个难点。因此，在汉语学术词汇筛选时，要对语料库中的词出现的频率进行统计分析，汉语面临一个英语等使用拼音文字的语言所没有的问题，那就是词频分析软件对词的切分正确与否直接关系到词频统计的准确性。大型语料库不可能完全手工切分词语，通常需要借助计算机自动分词软件进行词的切分，如果词的切分不正确，必然会影响词频统计结果的准确性。为了了解本研究使用的词频统计软件对汉语词的切分的正确率，我们在数据库中做了 1 万字符的语料抽样，对抽样语料中的词先进行人工切分及频率统计，再进行计算机自动切分及频率统计，并将人工切分统计结果和计算机切分统计结果进行对比。对比的结果发现词频统计软件在词的切分上有一部分会出现错误。通过分析，我们发现切分错误主要有三种情况：1. 多切。主要出现在一些生僻词语（包括专业性非常强的专业术语）或者以非词典中标准形态出现的词语（如重叠词）上，这些词语会出现一个词拆分为两个或者几个词的情况。比如人名“李嘉图”切分为“李嘉”、“图”，“狭窄性”切分为“狭窄”和“性”，“种种”切分为“种”、“种”等。2. 漏切。主要出现在数量短语和一些相对固定的组合上，统计软件会将它们作为一个词统计。比如“一个、第二、表面上、不会、不论是、不能、自我牺牲”切分时都作为一个词。3. 误切。比如将“对/穷人/比/对/富人/为/大”切分为“对/穷人/比/对/富人/为/大”，“负/责任”切分为“负责/任”，“一/个/人”切分为“一/个人”。

统计软件对抽样文本一共切分出 1370 个词，由于以上三个原因造成其中有 70 个词是错误切分出来的，错误切分的比例是 5%，其中多切的占 1.5%，漏切的占 2.7%，误切的占 0.8%；抽样文本中 1370 个词出现的总次数是 5603 次，其中错误切分的词语出现的总次数是 147 次，因此造成错误统计的频次占 2.6%，其中因为多切造成的错误统计的频次是 32 次，占 0.5%，因为漏切造成的错误统计的频次 106 次，占 1.9%，因为误切造成的错误统计的频次 19 次，占 0.3%。因为错误切分的词出现频次都很低，所以频次统计错误率低于词语切分错误率。<sup>8</sup> 统计软件的词语切分错误说明，因为汉语的特殊性，词汇统计软件对汉语词汇的切分还做不到百分之百正确，这就需要在学术词汇筛选的过程中对计算机切分的词汇进行人工干预。以上三类错误切分，漏切的都是可以在统计结果中进行人工干预予以纠正的，多切和误切的有一部分可以进行人工干预，抽样的文本经过人工干预以后可以将误切的比率降低到 2%，正确率达到 98% 以上，而词语频次统计错误率可以降低到 0.5% 左右，词频统计正确率达到 99.5% 左右。再加上错误切分的词汇多是一些生僻词和固定组合，很少会是我们筛选的目标词，因此本研究采用的词频统计软件在辅以人工干预的基础上可以保证学术词汇词频统计结果的误差在我们可以接受的范围，其效度和信度可以满足我们对统计结果的要求。

### (四) 汉语学术词汇的筛选

验证了统计软件的效度和信度以后，我们使用该词频统计软件对语料库进行了词频统计，并根据词频统计结果，按照学术词汇筛选原则制定了筛选学术词汇的具体标准，然后根据这个标准最终确定汉语的学术词汇。这个过程我们分以下几个阶段完成：

1. 对学术语料库中 1500 万字的文本进行词频统计。
2. 对计算机自动统计得出的使用频率比较高的词进行人工干预，重新切分统计软件误切的词语，并据此对词频数据进行调整。
3. 在初步统计结果的基础上，确定筛选学术词汇的具体标准。

---

8 这是因为错误切分的词在抽样文本中多是出现频率只有一次的词，特别是多切的词多是使用频率很低的词。



(1) 学术词汇常用性的具体标准：学术词汇应该属于学术文献语料库中使用频率最高的3000词的范围。

(2) 学术词汇学术性的具体标准：汉语学术词汇在学术文献中的使用频率应该高于在普通文献中使用频率的一倍以上，且不属于通用汉语语料库中使用频率最高的2000个常用词。

(3) 学术词汇通用性的具体标准：学术词汇应该至少在两个以上的学科门类中高频出现，且在每个学科门类中出现的频次应不少于40次。如果只在一个学科门类中高频出现的词则另列为次通用词汇。

4. 按确定的标准进行筛选，最后确定汉语的学术词汇。

按照以上标准(1)软件自动统计后得出前3000个使用频率最高的词，这些词在学术语料库中每个词的出现频率都超过了340次。<sup>9</sup>

经过标准(2)的筛选，我们得到685个词，在这些词中，我们重点检查了有“多切”可能的单音节词，检查结果发现，其中有16个单音节词，存在误切可能。对这16个单音节词在语料库中使用的情况进一步分析发现，它们基本上是由于软件“多切”把部分构词语素作为独立的词统计，因而抬高了其使用频次，因此我们剔除了这16个单音节词(语素)。对余下的669个词语的难度进行分析发现，这些词都是属于HSK水平词汇丙级词以上的词。<sup>10</sup>这说明标准(1)和(2)在考虑词语出现频率的同时实际上起到了有效控制词语难度的作用。

经过对选定词语在普通文本和学术文本中出现频次和覆盖率的统计结果进行分析，<sup>11</sup>我们发现，这些词充分体现了学术词汇学术性和常用性的特点，反映了这些词在使用范围上的特殊性。669个学术词汇在1500万字的学术文献语料库中每个词出现的平均频次是763次，在2000万字符的现代汉语通用平衡语料库中每个词平均出现的频次是261次(折合为1500万字出现的频次是约190次)，前者每个词出现的平均频次是后者的4倍，这些词在学术文献中的出现频次远远高于一般文献中出现的频次。669个学术词汇在普通文献中的文本覆盖率为1.8%，在学术文献中的文本覆盖率约为7.1%，同样远高于普通文献中这些词语的文本覆盖率。虽然这个覆盖率不及Coxhead(2000)AWL的10%，但考虑到AWL选出的英语学术词汇是570个词族，其包含的词语的个数实际上要大于我们给出的学术词表的669个词，因此7.1%已经是一个相当高的覆盖率了。

和通用词汇、专业词汇的筛选比较起来，学术词汇的筛选不是一件简单的容易完成的工作，在选择的过程中要尽量做到词语学术性和通用性的平衡、习得难度和使用频率的平衡，这是一个非常困难的任务。按照以上三个具体标准进行筛选后，我们认为最后确定的669个学术词汇，比较好地体现了学术词汇的学术性、常用性和通用性。这样的筛选既保证了学术词汇在学术文献中的常用性和一定的难度，又体现了学术词汇和一般通用词汇的差异。由于学术文献是一种专业性的文献，因此同一学科、同一学科门类、不同学科门类之间在使用词语的相似性上有着显著差异。因为专业性的差异，要在不同学科门类之间做到平衡，选择学术词汇时要求所有的学术词汇都要在所有学科门类之间通用其实是不太现实的。正因为这样，我们在筛选学术词汇时保留了一部分在一个学科门类中出现但使用频率极高的学术词语，但是因为其通用性有别于其它的学术词汇，因此，我们把这些词都归为单独的一类：次通用学术词汇。

9 以这个标准确定的词出现的频率在总词频中所占的比例不少于0.004%。

10 词汇水平分级依据中国国家对外汉语教学领导小组办公室1992年制定的《汉语水平词汇和等级汉字大纲》。在669个词中只有“作业”一个词属于HSK水平词汇中的乙级词，但是考虑到这个词在学术文献中使用最多的是该词“从事某种生产活动”这个义项(如“工程作业”)，不是其作为乙级词的意义和用法，因此我们仍然保留了这个词。

11 普通文本的词频统计数据以中国国家语委现代汉语通用平衡语料库的词频统计数据为依据，该语料库收录了约2000万字符的语料。

## 五、汉语学术词汇表的制定

在选定 669 个学术词汇后，我们按照其使用频率的高低，把它们分为最常用学术词汇、常用学术词汇、次常用学术词汇三类，并按照使用频率的高低排序。同时根据通用性强弱，两个学科门类以上的文献中高频出现的词归为通用词汇，只在一个学科门类中使用的词语单列为次通用词汇，并在后面注明其出现的学科门类，其中（1）表示这个词高频出现于“政治哲学”学科门类中，（2）表示这个词高频出现于“管理经济”学科门类中，（3）表示这个词高频出现于“人文社科”学科门类中，（4）表示这个词高频出现于“理工农医”学科门类中。

### （一）汉语 200 个最常用学术词汇

下面是在对汉语学术文献数据库进行词频统计的基础上筛选出来的汉语 200 个最常用的学术词汇，他们在数据库中出现的次数从 4359 到 852 次，绝大部分词语出现在三到四个学科门类中，只有 13 个词出现在一个学科门类中。在汉语的学术词汇中，这一部分词不但是使用频率非常高的词，而且也是通用性非常强的词。在这部分词中，只出现在一个学科门类中的词所占比重很少，这些次通用词汇之所以在语料词频统计中能够进入 200 个最常用的学术词汇，是因为这些词在该学科门类中超高的使用频率。

#### 通用词汇

1. 资本	39. 近代	77. 通货膨胀	115. 意味（着）	153. 墨子
2. 需求	40. 均衡	78. 解	116. 处罚	154. 公元前
3. 德	41. 购买	79. 拥有	117. 攻	155. 致
4. 罪	42. 定义	80. 最终	118. 传媒	156. 行业
5. 证	43. 阴	81. 预算	119. 就业	157. 看作
6. 美元	44. 消费者	82. 协议	120. 甲	158. 区分
7. 总统	45. 理性	83. 沟通	121. 故意	159. 当事人
8. 利	46. 规范	84. 涉及	122. 运行	160. 恶
9. 供给	47. 动机	85. 预期	123. 操作	161. 兼
10. 曲线	48. 财富	86. 分类	124. 作业	162. 公平
11. 假设	49. 模型	87. 损害	125. 福利	163. 禁止
12. 权	50. 税收	88. 会计	126. 足以	164. 特性
13. 州	51. 选举	89. 儒家	127. 物理学	165. 开支
14. 国会	52. 公共	90. 图形	128. 模式	166. 额
15. 经济学	53. 义	91. 网络	129. 利率	167. 愈
16. 命题	54. 定律	92. 他人	130. 乃	168. 学派
17. 物品	55. 学术	93. 官	131. 监督	169. 题
18. 联邦	56. 是以	94. 散	132. 装置	170. 参与
19. 法院	57. 支出	95. 效应	133. 投入	171. 凭证
20. 决策	58. 版	96. 列	134. 技能	172. 依赖
21. 支付	59. 程序	97. 方案	135. 失业	173. 舆论
22. 亿	60. 描述	98. 士	136. 产业	174. 得以
23. 贸易	61. 特定	99. 史记	137. 激励	175. 出售
24. 阳	62. 报酬	100. 汉	138. 情节	176. 牛顿
25. 天下	63. 邪	101. 外部	139. 契约	177. 记录
26. 脉	64. 取决于	102. 律	140. 官员	178. 私人
27. 公	65. 舌	103. 创新	141. 宏观	179. 载
28. 税	66. 立法	104. 儒学	142. 粒子	180. 数字

29. 函数	67. 出版	105. 正如	143. 承担	181. 出于
30. 虚	68. 苏 (苏联)	106. 储蓄	144. 案件	182. 春秋
31. 例子	69. 交易	107. 案	145. 推理	183. 哲学家
32. 垄断	70. 实	108. 相关	146. 有助于	184. 谈判
33. 收益	71. 孔子	109. 公式	147. 有限	185. 可能性
34. 刑法	72. 数据	110. 湿	148. 诉讼	186. 所得
35. 犯	73. 波	111. 知觉	149. 政党	187. 存款
36. 变动	74. 职业	112. 复	150. 管制	
37. 君	75. 职能	113. 客体	151. 背景	
38. 假定	76. 风险	114. 善	152. 计算机	

### 次通用词汇

1. 民法 (1)	4. 条约 (1)	7. 痕迹 (3)	10. 外交 (1)	13. 胃 (4)
2. 边际 (2)	5. 乾隆 (3)	8. 肺 (4)	11. 有期徒刑 (1)	
3. 政体 (1)	6. 肝 (4)	9. 肾 (4)	12. 赋 (1)	

### (二) 汉语 200 个常用学术词汇

下面是在汉语学术文献数据库词频统计的基础上筛选出的 200 个汉语常用的学术词汇, 这些词汇在语料库中出现的次数从 845 次到 548 次。其中通用性很强和比较强的学术词汇有 174 个, 只出现在一个学科门类的次通用学术词汇 26 个。这 26 个次通用词汇中, 应用于“理工农医”这个学科大类的词语占了相当大的比例, 其次是用于“政治哲学”和“管理经济”门类的一些学术词语, 只使用于“人文社科”学科门类中的次通用词语数量很少, 这从一个方面反映出了不同学科门类在使用学术词汇上的特点。

### 通用词汇

1. 持	36. 方程	71. 论证	106. 进展	141. 臣
2. 末	37. 危害	72. 侵犯	107. 以致	142. 信用
3. 以便	38. 刑罚	73. 设置	108. 协定	143. 足够
4. 痛	39. 甚	74. 事务	109. 英格兰	144. 限度
5. 间接	40. 黑人	75. 法定	110. 数额	145. 债务
6. 放弃	41. 降	76. 欲望	111. 呈现	146. 平民
7. 变革	42. 违反	77. 当代	112. 边界	147. 寻求
8. 著	43. 高于	78. 输入	113. 判决	148. 用以
9. 衡量	44. 伦理	79. 见解	114. 正当	149. 除非
10. 非法	45. 柏拉图	80. 低于	115. 罗马	150. 纯粹
11. 资产	46. 过失	81. 耳	116. 略	151. 修正案
12. 事故	47. 刑	82. 法官	117. 存	152. 乃是
13. 商人	48. 注	83. 含义	118. 仁	153. 减
14. 赔偿	49. 之上	84. 编	119. 合法	154. 论点
15. 叙述	50. 后果	85. 候选人	120. 视为	155. 品德
16. 汗	51. 礼	86. 奇	121. 大臣	156. 通信
17. 核心	52. 假如	87. 西欧	122. 术	157. 签订
18. 司法	53. 联合国	88. 议员	123. 候	158. 法人
19. 公众	54. 吉	89. 一方	124. 码	159. 行使
20. 辞	55. 论文	90. 证据	125. 改良	160. 儒
21. 较为	56. 计	91. 丁	126. 希腊	161. 未能



- |         |        |           |         |           |
|---------|--------|-----------|---------|-----------|
| 22. 弹性  | 57. 所能 | 92. 确立    | 127. 孟子 | 162. 公正   |
| 23. 产出  | 58. 二者 | 93. 物价    | 128. 干涉 | 163. 基于   |
| 24. 大于  | 59. 负担 | 94. 场合    | 129. 假说 | 164. 雇主   |
| 25. 危机  | 60. 瓦  | 95. 亚里士多德 | 130. 补偿 | 165. 仪    |
| 26. 失   | 61. 威胁 | 96. 正义    | 131. 王朝 | 166. 出版社  |
| 27. 生产者 | 62. 教会 | 97. 达成    | 132. 金融 | 167. 居    |
| 28. 剩余  | 63. 现金 | 98. 变量    | 133. 暴力 | 168. 名义   |
| 29. 人格  | 64. 团体 | 99. 代理人   | 134. 试图 | 169. 干预   |
| 30. 利息  | 65. 持续 | 100. 国民   | 135. 证实 | 170. 逆    |
| 31. 遂   | 66. 圣  | 101. 潜在   | 136. 短期 | 171. 世界大战 |
| 32. 所有者 | 67. 予以 | 102. 代理   | 137. 法则 | 172. 重要性  |
| 33. 领导者 | 68. 货物 | 103. 享有   | 138. 联盟 | 173. 取代   |
| 34. 论述  | 69. 属性 | 104. 意识形态 | 139. 认定 | 174. 镑    |
| 35. 拒绝  | 70. 积  | 105. 尚未   | 140. 征收 |           |

### 次通用词汇

- |             |            |             |            |            |
|-------------|------------|-------------|------------|------------|
| 1. 痰 (4)    | 7. 竞选 (1)  | 13. 共和党 (1) | 19. 外长 (1) | 25. 传输 (4) |
| 2. 国民收入 (2) | 8. 管理者 (2) | 14. 科目 (2)  | 20. 援助 (1) | 26. 列传 (3) |
| 3. 经济学家 (2) | 9. 城邦 (2)  | 15. 化学家 (4) | 21. 燥 (4)  |            |
| 4. 民主党 (1)  | 10. 诗歌 (3) | 16. 研制 (4)  | 22. 记账 (2) |            |
| 5. 地租 (2)   | 11. 脾 (4)  | 17. 选民 (1)  | 23. 上涨 (2) |            |
| 6. 起重机 (4)  | 12. 苔 (4)  | 18. 民事 (1)  | 24. 引力 (4) |            |

### (三) 汉语 269 个次常用学术词汇

下面是在汉语学术文献数据库词频统计的基础上筛选出的 269 个汉语次常用学术词汇, 这些词汇在汉语学术文献语料库中出现的次数从 548 次到 342 次。在 269 个次常用学术词汇中, 出现于一个学科门类中的次通用词汇有 59 个, 数量比“最常用学术词汇”和“常用学术词汇”中的次通用词汇数量明显增加。这说明, 在学术文献中, 使用频率和学术词汇的专业性密切相关, 专业性越强, 在整个学术文献中的使用频率则自然降低。作为半专业词汇, 学术词汇的应用范围具有跨学科性质, 因此在各种学术文献中的综合使用频率会比较高, 而真正的属于某一个学科或者专业的专业词汇, 由于使用范围有限, 因此在整个学术文献中的使用频率大多并不会太高。

### 通用词汇

- |        |        |          |         |         |
|--------|--------|----------|---------|---------|
| 1. 序   | 43. 助  | 85. 天文学  | 127. 纳税 | 169. 关税 |
| 2. 伐   | 44. 主管 | 86. 战后   | 128. 圣人 | 170. 雇用 |
| 3. 后期  | 45. 总额 | 87. 贫困   | 129. 税率 | 171. 便利 |
| 4. 款   | 46. 单一 | 88. 实体   | 130. 外在 | 172. 卒  |
| 5. 金额  | 47. 玄  | 89. 财物   | 131. 无效 | 173. 惯性 |
| 6. 与其  | 48. 法治 | 90. 输出   | 132. 奠定 | 174. 为人 |
| 7. 保险  | 49. 参见 | 91. 爱因斯坦 | 133. 期限 | 175. 债券 |
| 8. 本能  | 50. 股票 | 92. 佛    | 134. 专制 | 176. 事例 |
| 9. 条款  | 51. 关注 | 93. 可知   | 135. 罚  | 177. 例证 |
| 10. 持有 | 52. 古典 | 94. 认知   | 136. 权威 | 178. 域  |
| 11. 编制 | 53. 分歧 | 95. 授权   | 137. 是非 | 179. 证券 |
| 12. 惩罚 | 54. 君子 | 96. 量子   | 138. 小于 | 180. 人和 |

- |         |        |           |          |           |
|---------|--------|-----------|----------|-----------|
| 13. 领导人 | 55. 诸侯 | 97. 种族    | 139. 律师  | 181. 误差   |
| 14. 对应  | 56. 理学 | 98. 无关    | 140. 疏   | 182. 犹    |
| 15. 诉   | 57. 罚金 | 99. 后世    | 141. 集合  | 183. 光学   |
| 16. 对外  | 58. 进而 | 100. 确认   | 142. 动态  | 184. 人力资源 |
| 17. 理智  | 59. 言论 | 101. 同年   | 143. 俄   | 185. 抒情   |
| 18. 履行  | 60. 期望 | 102. 股东   | 144. 工会  | 186. 被告   |
| 19. 数目  | 61. 术语 | 103. 赤    | 145. 几何  | 187. 侵害   |
| 20. 人为  | 62. 因果 | 104. 坐标   | 146. 迫使  | 188. 心理学家 |
| 21. 灵   | 63. 弟子 | 105. 必定   | 147. 意愿  | 189. 迁移   |
| 22. 效用  | 64. 情境 | 106. 家族   | 148. 歧视  | 190. 锡    |
| 23. 治理  | 65. 广义 | 107. 大国   | 149. 私   | 191. 气质   |
| 24. 效力  | 66. 济  | 108. 执政   | 150. 限于  | 192. 违法   |
| 25. 淡   | 67. 俗  | 109. 邑    | 151. 职位  | 193. 破产   |
| 26. 替代  | 68. 滞  | 110. 蒙古   | 152. 稳定性 | 194. 经学   |
| 27. 争论  | 69. 信念 | 111. 强制   | 153. 职责  | 195. 雇佣   |
| 28. 议会  | 70. 阐述 | 112. 万物   | 154. 荷兰  | 196. 价值观  |
| 29. 答案  | 71. 世家 | 113. 恒常   | 155. 引入  | 197. 徒    |
| 30. 起源  | 72. 百姓 | 114. 权益   | 156. 雅   | 198. 对抗   |
| 31. 置   | 73. 扩展 | 115. 微观   | 157. 全球  | 199. 而后   |
| 32. 西洋  | 74. 耕作 | 116. 增长率  | 158. 侵权  | 200. 关联   |
| 33. 系列  | 75. 症  | 117. 武    | 159. 判处  | 201. 媒体   |
| 34. 兴起  | 76. 表述 | 118. 法规   | 160. 数学家 | 202. 演化   |
| 35. 等级  | 77. 人权 | 119. 假使   | 161. 投票  | 203. 吏    |
| 36. 被试  | 78. 亡  | 120. 联结   | 162. 预言  | 204. 辨证   |
| 37. 准则  | 79. 有权 | 121. 首次   | 163. 趋于  | 205. 次年   |
| 38. 利于  | 80. 配置 | 122. 推论   | 164. 智   | 206. 纳税人  |
| 39. 附   | 81. 征税 | 123. 物理学家 | 165. 国君  | 207. 宣告   |
| 40. 培训  | 82. 有用 | 124. 伤害   | 166. 当选  | 208. 耕种   |
| 41. 阻止  | 83. 寡头 | 125. 所有权  | 167. 理念  | 209. 栏    |
| 42. 经典  | 84. 中期 | 126. 自愿   | 168. 协会  | 210. 模    |

### 次通用词汇

- |             |              |             |             |             |
|-------------|--------------|-------------|-------------|-------------|
| 1. 磁 (4)    | 13. 越南 (1)   | 25. 音节 (3)  | 37. 要件 (1)  | 49. 阴虚 (4)  |
| 2. 康熙帝 (3)  | 14. 所致 (4)   | 26. 友好 (1)  | 38. 准备金 (2) | 50. 砂轮 (4)  |
| 3. 主权 (1)   | 15. 苏格拉底 (1) | 27. 韵 (3)   | 39. 电磁 (4)  | 51. 古巴 (1)  |
| 4. 比率 (2)   | 16. 神经病 (3)  | 28. 乘数 (2)  | 40. 词语 (3)  | 52. 气血 (4)  |
| 5. 清朝 (3)   | 17. 会谈 (1)   | 29. 下属 (2)  | 41. 雍正 (3)  | 53. 散文 (3)  |
| 6. 滑 (4)    | 18. 康熙 (3)   | 30. 凯恩斯 (2) | 42. 账户 (2)  | 54. 症候 (4)  |
| 7. 教皇 (1)   | 19. 总需求 (2)  | 31. 甘草 (4)  | 43. 道教 (1)  | 55. 视网膜 (3) |
| 8. 君主 (3)   | 20. 签署 (1)   | 32. 核算 (2)  | 44. 谷物 (2)  | 56. 疼痛 (4)  |
| 9. 泻 (4)    | 21. 声明 (1)   | 33. 清热 (4)  | 45. 花费 (2)  | 57. 尚书 (3)  |
| 10. 余额 (2)  | 22. 劳务 (2)   | 34. 厂商 (2)  | 46. 激光 (4)  | 58. 译者 (1)  |
| 11. 利息率 (2) | 23. 首脑 (1)   | 35. 载荷 (4)  | 47. 失业率 (2) | 59. 丸 (4)   |
| 12. 相对论 (4) | 24. 拘役 (1)   | 36. 法典 (1)  | 48. 帧 (4)   |             |

## 参考文献

- 中国国家对外汉语教学领导小组办公室 (1992) 《汉语水平词汇和等级汉字大纲》，北京语言大学出版社。
- 中国国家语委《现代汉语语料库词语频率表》，<http://www.cncorpus.org/resources.aspx>。
- Coxhead, A. (1998) *An academic word list. LALS Occasional Publication*. Victoria University of Wellington, Wellington.
- Coxhead, A. (2000) *A new academic word list*. TESOL Quarterly, 34, 213-238.
- Coxhead, A. (2002) *The academic word list: A corpus-based word list for academic purposes*. In B. Ketterman & G. Marks (Eds.), Teaching and language corpora (TALC) conference proceedings (pp. 73-89). Atlanta, GA: Rodopl.
- Coxhead, A, & Nation, I.S.P. (2001) *The specialized vocabulary of English for academic purposes*. In J. Flowerdew & M. Peacock (Eds.), Research perspectives on English for academic purposes (pp. 252-267). Cambridge: Cambridge University Press.
- David Hirsh. (2010) *Academic Vocabulary in Context*. Bern: Peter Lang AG International Academic Publishers.
- Ken Hyland & Polly Tse. (2007) *Is There an “Academic Vocabulary”?* TESOL Quarterly, Vol. 41. No. 2 (Jun., 2007), pp. 235-253, Published by Teachers of English to Speakers of Other Languages, Inc. (TESOL)
- Nation, P. (2001) *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

