

中高级科技汉语课文词汇使用特点 及计量分析

An Econometric Analysis of the Features of Vocabulary Use in Intermediate and Advanced Scientific Chinese Texts

姜永超¹

Yongchao JIANG

中国燕山大学文法学院

College of humanities and Law, Yanshan University, China

38425043@qq.com

周露²

Lu ZHOU

中国燕山大学文法学院

College of humanities and Law, Yanshan University, China

luzhou202210@163.com

郝佳昕³

Jiixin HAO

中国中央民族大学国际教育学院

College of International Education, Minzu University of China

22400277@muc.edu.cn

摘要 目前,“中文+专门用途”教材资源建设成效显著,其中,科技用途中文教材出版量位列前三。对其课文词汇使用和编排特点研究,可以更好地发挥其作为通用语言教材与专业技能教材之间的桥梁作用,服务专门用途汉语教与学。本研究将《科技汉语中级阅读教程》《科技汉语高级阅读教程》中的课文作为研究材料,对其的词汇丰富度、文本难易度、词汇复现度三个要素进行计量考察,并与《国际中文教育中文水平等级标准》的词汇进行比较。研究发现,词汇多样性与难度层级不成正相关;平均的词长和句长下,《中级》课文之间难度系数变化较大;两部教材的专业词汇复现率和复现间隔占比低;两部教材中非《标准》词汇占比约五成,六次及以上《标准》外的词汇仅占 20%左右。基于科技汉语教材词汇使用特点,我们从教材编写、教师教学、学生学习三个方面给出了建议,基于与《标准》相比词汇覆盖率特点,我们认为应建立科技汉语领域专属大纲,并提出了三个参考原则。

关键词 词汇;教材分析;科技汉语;计量;专门用途汉语

Abstract At present, the construction of “Chinese plus special-purpose” teaching materials has achieved

收稿日期: 2023-10-14

作者简介: ¹ 姜永超, 燕山大学文法学院教授, 硕士生导师。

² 周露, 燕山大学文法学院硕士。

³ 郝佳昕, 中央民族大学国际教育学院博士。

remarkable results, among which the publication volume of Chinese teaching materials for scientific and technological purposes ranks among the top three. The study on the characteristics of vocabulary uses and arrangement in the text can better play its role as a bridge between general language teaching materials and professional skills teaching materials, and serve the special purpose Chinese teaching and learning. In this study, the text in *Science and Technology Chinese Intermediate Reading Course* and *Science and Technology Chinese Advanced Reading Course* is taken as research materials, and the three elements of vocabulary richness, text difficulty and vocabulary repetition are measured, and compared with the words in the “International Chinese Language Education Chinese Proficiency Level Standards.” The results show that there is no positive correlation between lexical diversity and difficulty level. The difficulty coefficient of Intermediate text varies greatly under the average word length and sentence length; The repetition rate and repetition interval of the two textbooks are low; About 50% of the words in the two textbooks are non-standard words, and only about 20% are words that are not standard words for six or more times. Based on the characteristics of vocabulary use in sci-tech Chinese textbooks, we give suggestions from three aspects: textbook compilation, student learning and teacher teaching. Based on the characteristics of vocabulary coverage compared with the Standards, we believe that we should establish a special outline for sci-tech Chinese, and put forward three reference principles.

Key words Vocabulary; Textbook Analysis; *Ke Ji* Chinese; Measurement; Professional Chinese

一、引言

教材是建立国际中文教育学科知识体系和价值理念的载体，是实现学科教学目标的基础材料。截止 2020 年底，全球已累计出版 564 种专门用途的中文教材，教材类型涉及科技、经贸、旅游、医学等十余个专业和职业方向，“中文+职业技能”教材资源建设成效显著。其中，科技用途中文教材数量占 14.23%，位列第三，该系列教材作为学历留学生（包括本科、硕士和博士）从汉语语言学习到专业知识学习的桥梁，能使学历留学生更快地适应各自的专业学习。

杜厚文（1986）率先提出了科技汉语的教学体制和教材问题是当前中国理工类留学生教学存在的主要问题。随后，单韵鸣（2008）、韩志刚/董杰（2010）、马宏（2011）等学者从不同角度提出了教材的编写原则和词汇选择、文体、句式等方面的改进建议。张明月（2011）针对科技汉语文体的衔接方式、语相特征、隐喻现象及词汇特征进行了深入研究，并对教材的发展提出了具体建议。此后，张莹（2014）对近 30 年来的科技汉语教材编写进行了梳理，发现其发展相比通用型汉语教材较慢，强调了编写教材时需要注重系统化的重要性。李若岚（2020）对七类已出版的科技汉语教材在课文、词汇、练习、语法等方面进行了综合分析，总结了它们的优缺点，并提出了相关建议。代睿（2021）、胡越（2021）分别从编排、选材、语言形式以及编写模式的角度对科技汉语教材进行了深入考察，并提出了以功能语言形式和分层课程设计为核心的编写建议。目前在对外汉语领域有关科技汉语教材的研究，整体上遵循从教材编写、教材定位等宏观角度转向词汇、课文、练习等微观角度的趋势。但是至今还没有学者从计量的

角度对科技汉语教材进行综合考察。

词汇是语言学习的基础,研究词汇可以了解语言的基本构成。因此,本文以词汇为切入点,采用计量方法,以安然主编的《科技汉语中级阅读教程》(以下简称《中级》)《科技汉语高级阅读教程》(以下简称《高级》)两本教材中的课文词汇为研究材料,展开以下两个方面的研究:

(1)从词汇丰富度、词汇难易度、词汇复现度三个方面探究科技汉语教材课文词汇的使用特征;(2)将《标准》中的词汇和科技汉语教材课文词汇进行对比,思考《标准》是否能作为科技汉语教材编写的参考大纲。

二、研究设计

(一) 研究材料

经初步统计,自上世纪80年代始,已出版的《科技汉语》类教材约25部,这些教材在不同的时期或院校曾作为学历留学生的(特别是理工类在华学历留学生)必读教材。其中,安然主编的《中级》《高级》(北京大学出版社)两部教材体例结构类似,自成体系,每课均由一篇重点阅读课文、三篇快速阅读文章和一篇“科技写作规范”组成,其中,《中级》共18课90篇文章,《高级》共16课80篇文章。此外,两部教材的内容具有一致性,均是以理工类本科培养计划中的公共课程为纲,将高等数学、计算机、物理化学、统计学等学科的基础知识作为主题纳入教材。值得注意的是,这两部教材在同质内容的基础上,又区分为中级和高级,以满足不同学期的需要,适应不同水平的学历留学生。此种设计有益于考察科技汉语课程设计的等级划分和目标水平。因此,我们以《中级》《高级》教材中的34篇课文为研究材料。

(二) 研究方法

第一,材料收集方法:纸质或电子资源的数字化。采用OCR图文识别将两本教材的170篇文章转化为一个约8万字的语料库,并辅之以人工校对。

第二,材料处理方法:数据清洗、Jieba分词标注和人工校对相结合。首先,进行数据清洗,去除链接等特殊符号。其次,调用Jieba的精确模式进行分词,按照名词、时间词、处所词、方位词、动词、形容词、区别词、状态词、代词、数词、量词、副词、介词、连词、助词、叹词、语气词、拟声词、前缀、后缀、字符串、标点符号22大类对语料库词性进行标注。再次,与《标准》中的词汇比较。

第三,数据生成方法:文本数据量化与统计。量化对象包括字、词的类符、词的形符、词性等形式参数。量化方法包括统计检验、U值、词汇密度、平均词长、平均句长、覆盖率、复现率和复现分布。同时,运用SPSSAU、Tableau、Excel等软件将量化结果转化成图或表格。

第四,研究内容和材料分析方法:量化对比分析法。运用第三步的量化结果,其一,对比分析《中级》《高级》课文词汇的丰富度、难易度和复现度特征;其二,与《标准》中的词汇

对比，揭示两部教材词汇使用特点。

三、结果与分析

（一）科技汉语教材课文词汇使用特点

以表 1 中《中级》《高级》教材的字数、词长、形符和类符四项数据为基础，通过词汇丰富度、词汇难易度和词汇复现度的相关计算，说明两部教材词汇使用特点。

表 1 《中级》《高级》量化分析基础数据表

教材	字数	词长	形符	类符
《中级》	29061	1.7624	16614	4266
《高级》	47981	1.7622	27064	5432

其中，类符（Type）是指语料库中的词汇总数，但是不包括重复的词汇。形符（Token）是指语料库中的所有词汇，包括重复的词汇。例如“微分学的主要内容包括极限理论、导数、微分等，积分学的主要内容包括定积分、不定积分等。”（《高级》第二课），在这句话中“主要”作为类符记一次，作为形符记两次。

1. 词汇丰富度

有关词汇丰富度（或称词汇丰富性）的研究，多通过文本词汇指标来衡量第二语言学习者语言水平。该指标可以用于测量教材文本词汇的丰富度，揭示教材文本词汇的使用特征，能为教材编写或修订提供参考。Laufer, Batia & Paul Nation（1995）认为可以从多样性、密度等四个维度分析词汇丰富性。参照其观点，尝试从词汇多样性和词汇密度两个维度量化分析《中级》《高级》教材。词汇多样性量化有计算 TTR 值、Uber Index 值等方法，其中，U 值计算能消除文本长度影响，因此，本研究主要采用 U 值法测量每课词汇，然后对比分析总体的 U 值平均数和标准差。

1.1 基于 U 值的词汇多样性

与其它多数词汇丰富性计算方法相同，U 值是以词形统计为基础，不涉及词义、词性和语境信息以及汉字的书写难度等，Vermeer（2000）、Jarvis（2002）等均认为 Uber Index 在测量词汇多样性这一指标上，具有较高准确性。其计算公式为：

$$\text{Uber index: } U = \frac{(\log \text{Tokens})^2}{(\log \text{Tokens} - \log \text{Types})} \quad (\text{Jarvis, 2002:59})$$

（注：Types 为类符，Tokens 为形符。）

对课文词汇 U 值进行 S-W 正态性检验，结果显示《中级》《高级》都没有呈现显著性（ $p=0.538$ ， $p=0.985>0.05$ ），所以均具备正态性特征。因此，可进行独立样本 T 检验，结果显示两部教材的 U 值均未表现出显著性（ $t=1.917$ ， $p=0.064>0.05$ ）。这意味着《中级》《高级》教材

的 U 值均表现出一致性，没有差异性¹。两部教材统计意义上的一致性应该与同为科技汉语主题相关，从而在词汇使用与行文表述上具有同类特征。否则，可能成为两类不同性质的教材。尽管，两部教材词汇多样性没有显著性区别，但并不意味着两者完全相同。如下表 2。

表 2.《中级》《高级》课文词汇 U 值的平均数和标准差

教材	U 值平均数	U 值标准差
《中级》	68.60	7.63
《高级》	63.84	6.72

由表 2 可知，两部教材整体上存在如下区别：第一，《中级》课文词汇多样性高于《高级》（68.60>63.84）²，反映《中级》课文所使用的非重复词汇数量更多，学习者记忆负担也相应增加。第二，《中级》部分课文词汇多样性数量更偏离均值，《中级》相对偏离值大于《高级》（7.63>6.72），反映《中级》课文之间词汇多样性差异较大，相比较没有《高级》均衡。可见，两部教材所标难易度与词汇多样性指标不成正相关。究其原因，可能与《中级》教材课文数量和主题数量均多于《高级》，文本长度更长，形符较多，其词汇所涉及范围或领域更广等等相关。这种差异在课文间的具体表现如下表 3。

表 3 《中级》《高级》课文词汇 U 值表

教材	第 1 课	第 2 课	第 3 课	第 4 课	第 5 课	第 6 课	第 7 课	第 8 课	第 9 课
《中级》	76.54	66.18	68.75	50.96	55.18	79.44	64.28	72.28	82.34
《高级》	55.38	59.87	70.45	65.75	68.34	64.53	52.88	58.22	67.65
教材	第 10 课	第 11 课	第 12 课	第 13 课	第 14 课	第 15 课	第 16 课	第 17 课	第 18 课
《中级》	69.82	67.49	70.88	68.54	64.22	72.61	62.74	71.21	71.37
《高级》	63.92	56.87	60.46	77.51	62.53	74.17	62.99		

首先，在两部教材中，每课词汇使用的丰富度均不小于 50.96（《中级》第 4 课），词汇最为丰富者达 82.34（《中级》第 9 课），值得注意的是，词汇丰富度的最大值和最小值均出现在《中级》，课文间词汇丰富性差异最高达三成以上（与表 2《中级》标准差 7.63 相符）。如果按词汇多样性加强学习难度，《中级》课文间难度设置不均衡，而《高级》优于《中级》。其次，除了《中级》多两课内容外，《中级》丰富度大于《高级》的有 10 课，低于《高级》的仅第 3、4、5、13、15、16 课。再次，《中级》最大值比《高级》最大值多 4.83。因此，无

¹ 注：在 TTR 试验中，S-W 正态性检验 $P=0.478, p=0.325>0.05$ ，具备正态性特质。但是，在独立样本 t 检验时，其显著性 $p=0.00^{**}<0.01$ ，其值说明《中级》《高级》教材词汇多样性存在差异。这与 U 值检验结果不同，也许是因为 TTR 值受文本长度不同造成的，因为《中级》18 课 90 篇文章，《高级》16 课 80 篇文章。
² 在基于 TTR 试验中，独立样本 t 检验的显著性 $p=0.00^{**}<0.01$ 、 $T=5.870$ 时，《中级》均值为 0.51，《高级》均值为 0.42，也说明《中级》教材词汇使用更加多样。

论从平均数反映的集中程度、标准差反映的多样性差异以及每课具体的 U 值量化对比, 均揭示《中级》课文丰富度大于《高级》。换言之,《中级》记忆负担多于《高级》。

基于 U 值分析结果和 TTR 值辅助, 无论从词汇多样性的总体特征值、课文之间多样性差异,《中级》都高于《高级》。因此, 课程所标难易度与词汇多样性不一致。但是, 这种差异不足以从量化统计上认为两部教材在词汇丰富度上有本质差异。

1.2 词汇密度和词类

词汇密度即实词在分析要素中所占的比例。实词指既有词汇意义, 又有语法意义还能充当句法成分的词, 虚词指仅有语法意义的词。参考黄伯荣、廖序东《现代汉语》, 在标注时将名词、动词、形容词、副词、区别词、状态词、数词、量词、代词、拟声词、叹词划为实词。将介词、连词、助词、语气词作为虚词。将课文中涉及的阿拉伯数字、英文字母等归入其它词。

根据词汇密度计算公式, 词汇密度=实词数/总词数(Ure, 1971)或词汇密度=实词数/小句总数(Halliday, 1985), 分析要素可以是总词数、实词+虚词或小句总数。限于篇幅该研究采取Ure观点。为考察两部教材所有词类特征, 同时量化虚词和其它词, 结果如下图1。

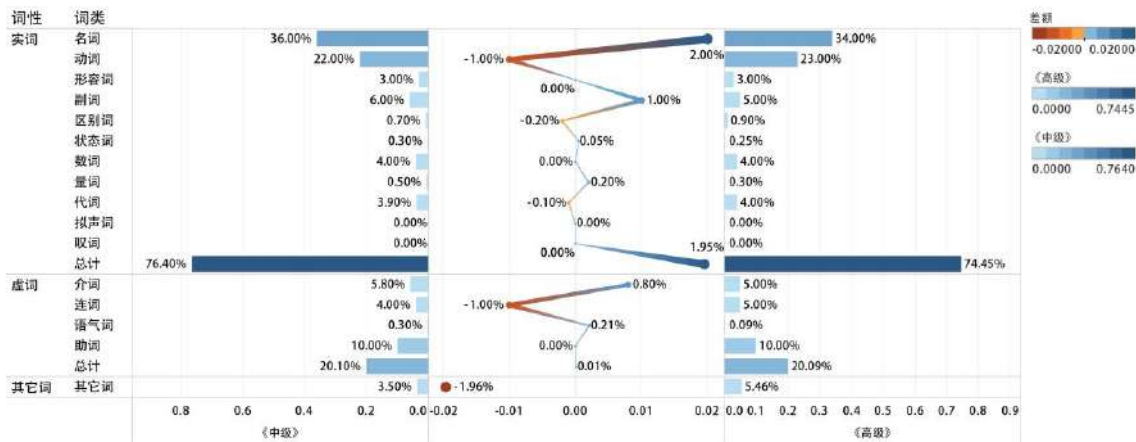


图1 《中级》《高级》教材词汇密度对比图

首先, 两部教材词汇密度总体相当,《中级》词汇密度大于《高级》(76.4%>74.45%)。两者相同点: 两书名词比例最高, 动词次之, 副词、代词和形容词又次之, 区别词、状态词和量词比例较低, 拟声词和叹词为零。可见, 在实词内部, 两书各类词分布趋势整体一致。两书名词较多的情况应该是教材中存在大量专业术语或学科名词的反映, 与这些名词或术语相匹配的动词也相应增加, 这种现象体现了从语言学习向专业学习的过程中, 需要掌握大量新词, 这是教材衔接语言学习与专业学习的体现, 而《中级》所占比例相对略高。不同点:《中级》名词、副词、状态词、量词比值分别略高于《高级》, 动词、区别词、代词分别略低于《高级》。

其次,《高级》(5.46%)其它词高于《中级》(3.5%) 1.96 个百分点, 是 16 种词类中除名词外差别最大的一组。这是由于所采取的分类将黄廖版《现代汉语》未提及的单位符号、网址

URL等归为其它词。由于《高级》中有近三分之一的课文内容涉及到数学公式、字母、符号。如第一课一元函数的拓展——多元函数、第二课“人类精神的最高胜利”——微积分的创立及应用、第十课《几何原本》与生活中的几何、第十一课谈谈空间直角坐标系、第十六课线性代数简介等。所以《高级》其它词比例比《中级》高。

以上词语分布状况,说明实词,特别是名词及与名词相搭配共现的动词在专业学习中的重要性。而《高级》公式或符号标记的其它词超过《中级》,反映《高级》词汇符号的抽象性数量更加突出。

(二) 以词为基础的文本难易度

Dale(1949)提出文本可读性(Readability)的概念,可读性指读者对阅读材料理解的难易程度。可读性公式是针对某种阅读文本,将影响阅读理解的可量化因素(主要是文本因素)综合起来进而形成的评价文本难易程度的公式。它是一种评估文本难度的客观方法。在文本可读性分析中,平均词长、以词为单位的平均句长是衡量文本可读性的重要指标。

词长的计算公式为总字数除以词数,此指标能反映教材课文的平均词汇长度。统计显示:《中级》教材课文的平均词长约 1.7624 字/词,《高级》教材课文的平均词长约 1.7622 字/词,两者仅有 0.0002 的差距,从平均词长这一指标上看,整体上两部教材课文难易度区分不大,说明同类专门教材的平均词长具有一致性特征。具体到每课平均词长如下图 2。

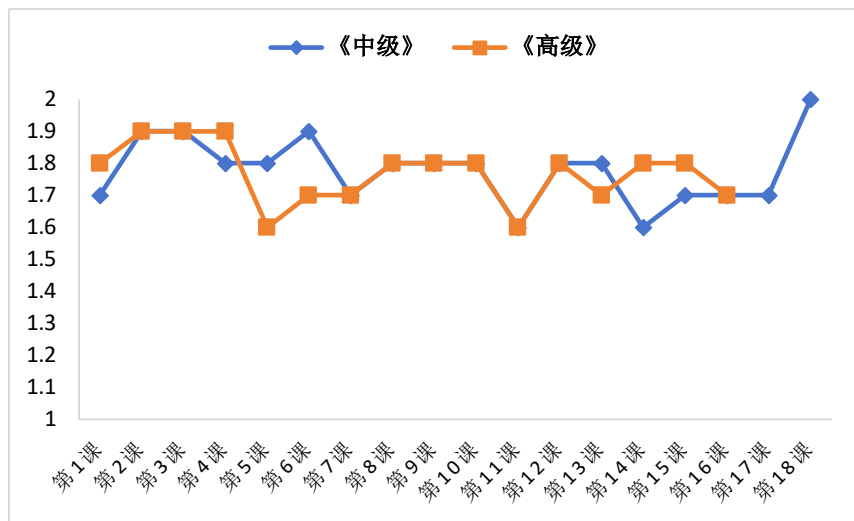


图2 《中级》《高级》课文平均词长

从平均词长来看,总体上,《高级》第1课至第16课课文难度高于或等于《中级》课文;但是,第5课、第6课和第13课课文难度低于《中级》。另外,《中级》第18课课文难度最高。可见,《中级》课文难度差异性大于《高级》。

句长计算公式为词数除以句子数,通过平均句长衡量文本难度。《中级》平均句长为 13.8,

《高级》平均句长为 14.8，整体上，《高级》课文难度略高于《中级》。每课平均句长如下图 3。

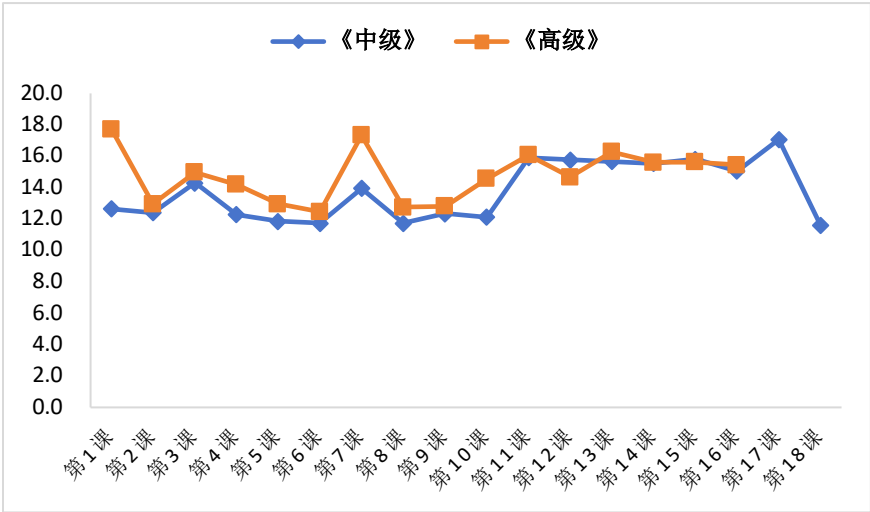


图 3 《中级》《高级》以词为单位的平均句长

从平均句长看，除了第 12 课和第 17 课略高于《高级》外，《中级》其它 16 课的课文难度均低于或等于《高级》。此外，通过以字为单位的平均句长分析来看，《高级》平均句长为 30.32 字/句，《中级》平均句长为 25.37 字/句，难易结果与以词为单位的平均句长相当。

(三) 词汇复现度

根据艾宾浩斯记忆法，在短时记忆转变为长时记忆的过程中，词汇复现对于词汇学习或习得至关重要，其中，教材中的词汇复现是影响留学生词汇学习的重要因素之一。赵金铭（1998）等学者认为复现是学习汉语词汇的重要方法。但是，部分研究者在复现次数与学习效果的认识上存在差异，Saragi（1978）等人实验发现：在语境中复现 6 次以内的生词会被一半被试者成功掌握，而重复出现大于 6 次的生词会被 93% 的被试者成功习得。柳艳梅（2002）认为生词最少复现 3 次才有明显的习得效果。江新（2005）提出复现 5 次后对生词习得具有重要作用。我们以最大复现次数 6 次为参照，作为第一分界线，得出两部教材的词汇复现次数如下表 4。

表 4. 《中级》《高级》词汇复现对比表

出现次数(次)	《中级》		《高级》	
	词语数量（个）	占教材课文词汇比例	词语数量（个）	占教材课文词汇比例
1-5	3726	87%	4571	84%
6-10	283	7%	407	7%
11-15	108	3%	149	3%
16-20	48	1%	97	2%

21 以上	101	2%	208	4%
合计	4266		5432	

《中级》《高级》课文词汇出现频次在 5 次及以下占比较高, 分别为 87%和 84%, 其中, 《中级》出现一次的词语占 56%, 《高级》出现一次的词语占 54%。因此, 两部教材课文词汇低频词语占比高, 出现一次的词均超过半数, 这必然导致课文之间词汇复现率低。究其原因, 其一, 与两部教材课文涉及主题较多有关, 不同主题其专业词汇必然不同; 其二, 与有些课文虽属同一课下, 但主题内容毫不相干有关。

《中级》《高级》6 次及以上词汇的占比较低, 分别为 13%和 16%, 其中, 复现次数在 21 次及以上的词汇分别占 2%和 4%。可见, 两部教材高频词占比均不高。根据学界关于复现次数与习得效果之间关系的研究, 两部教材能被大多数留学生习得的比例不到两成, 从而难以实现教材提出的扩大留学生多主题专业术语和词汇量的目标。出现频次在 6 次及以上且属于《标准》“三级九等”内的词汇, 《中级》有 453 个 (占比 83.9%)、《高级》有 686 个 (占比 79.7%), 很多都是基本语言交际或表达中使用的词语, 用于构成课文的一般性表达, 而这些词汇在留学生语言学习的中高级阶段基本上都已经接触, 例子在《中级》和《高级》中的出现频率如: 的 (1391/2377)、在 (243/334)、是 (242/484)、和 (170/395)、了 (159/227)、有 (115/173)、就 (102/88)、可以 (62/125)、我们 (47/99) 等。这些词的高复现有利于巩固和强化汉语中基础词汇的学习效果, 而非《标准》的专业词汇仅占两成左右, 考虑到科技汉语等专门用途教材的编排目的, 这类词汇的较高复现和专业词汇的低复现是否需要调整, 有待商榷。

根据艾宾浩斯遗忘规律, 适时的复习可以使短时记忆转化为长时记忆。对于教材而言, 除了复现次数, 词汇在课文之间的复现分布情况也会影响留学生习得效果。如: 出现 20 次的词仅分布在某一篇文章中, 而出现 6 次的词分布在某五篇或六篇文章中, 从而将有益于促进有目的学习和伴随性学习结合, 那么后者更有益于留学生词汇习得。基于此, 我们考察频次在 6 次及以上的词汇在课文之间的分布特点, 《中级》540 个词和《高级》861 个词所在课文分布情况如下图 4。

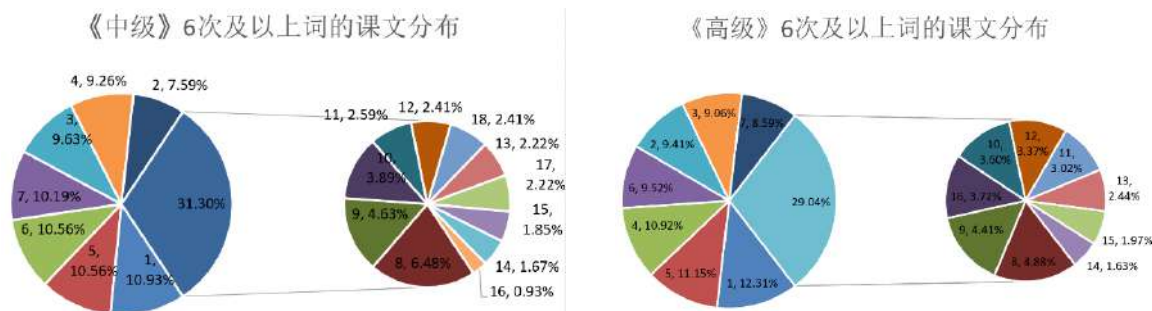


图 4 《中级》《高级》频次 6 次及以上词语的课文分布

(注：图中逗号前为词分布在不同课的数量、逗号后为该分布所占百分比。)

两书分布在7课及以下的词汇较多,《中级》《高级》分别占了七成左右;而分布在10课及以上的词汇数量明显减少,这反映了科技汉语等专门用途教材的特点。其中,仅出现在1课的词语占比最高,分别为10.93%和12.31%;这些词虽然出现频次在6次及以上,但是仅出现在整本教材的一课中,根据艾宾浩斯遗忘规律,如果没有其它因素的介入,仅就教材而言,很难实现理想习得效果。这类词很多属于专业词汇或高级汉语必备词汇,如《中级》频次“机械表(6)、静电(6)、肾脏(6)、化石(7)、变量(8)、病菌(8)、电流(8)、对称轴(8)、抛物线(8)、搜索引擎(8)、引力(8)、微生物(9)、胡夫(10)”等59个词;《高级》“胃蛋白酶(6)、源程序(7)、涡流(7)、微重力(7)、分子(7)、酸雨(8)、欧拉(8)、碱基(8)”等106个词。虽然这些具有专业特征的词汇或术语复现频次在6次及以上,甚至像《中级》中的纳米、压强等出现12次,《高级》中的投影面、多普勒、电场等出现17次,但是,这些词仅出现在一课中,在有目的的教学和引导学习后,虽然有利于留学生形成短时记忆,但是由于缺少课文之间的复现,加之课时限制和教学进度要求,会影响留学生的习得效果。

目前,鲜有涉及词汇课文分布与词汇长时记忆关系的研究。借鉴频次复现研究,频次和课文分布均在6次及以上,一定有益于留学生词汇习得,统计频次和课文分布均在6次及以上的词汇发现:《中级》《高级》分别占52.04%和47.15%。如下图,我们选取《中级》“出现、发展、过程、科学家、利用、美国、生活、速度、制造、重要”10个分布课数在6次及以上的词语。

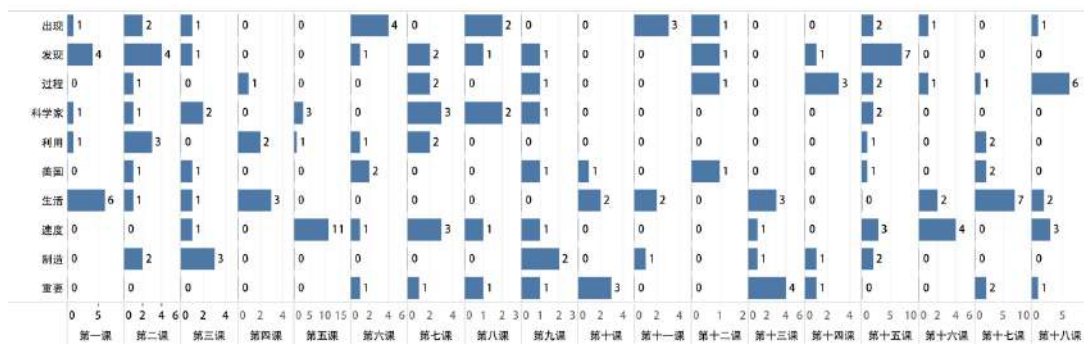


图5 《中级》部分频次和出现课数在6次及以上的词语分布

这些词不仅出现频率高,而且间隔科学,更符合学习者的记忆规律,能达到较好的学习效果。如“速度”在第3课首次出现后,在第5课出现11次,然后在第6、7、8、9、13、15、16、18课再次出现,再如“制造”在第2、3、9、11、13、14和15课分别出现,这样在不同课中的复现,实现目的性词汇学习和伴随性词汇学习的结合,从而形成长时记忆的词汇习得。但是,作为专门用途的教材而言,其复现和分布最高的词汇,并不一定体现编写者目的。如:《中级》分布在18课的词语占2.41%,分别为“不(54)、据(60)、可以(62)、也(73)、阅

读(86)、有(115)”等13个词;《高级》分布在16课的词语占3.72%,分别为“科技(28)、写作(29)、及(49)、不同(55)、一(56)、从(61)、用(65)、到(71)”等32个词。这些语言基础类词汇不仅频次高而且分布在每课中,是构成基本句法表达的基础元素,但是,它们不是专门汉语类教材词汇习得和语言表达的目标,很多词在初中级阶段均已经习得。

另外,教材之间的词汇复现影响着词汇习得效果。经统计,《高级》共复现《中级》词汇1850个,占《中级》类符词汇的43.37%。其中,《标准》之外的词有454个,占《中级》《高级》共现词的24.5%,如:“压强、函数、晶体、数轴、变量、电流、电子、对称轴”等等,两部教材之间的共现分布有益于这些专业词语的习得。

(三)与《标准》相比词汇覆盖率特点

《标准》是2021年由中华人民共和国教育部和国家语言文字工作委员会发布的,具有权威性、规范性,对教材编写起着指导性作用。该标准将词汇等量化指标分为“三等九级”(1-3级为初等,4-6级为中等,7-9级为高等)作为衡量学习者汉语水平的重要指标,成为国际中文教育的学习、教学、测试和评估提供参考。我们将《标准》涵盖的11092个词与教材课文词汇进行对比,从《标准》看科技汉语课文词汇的使用特点。结果如下图6。

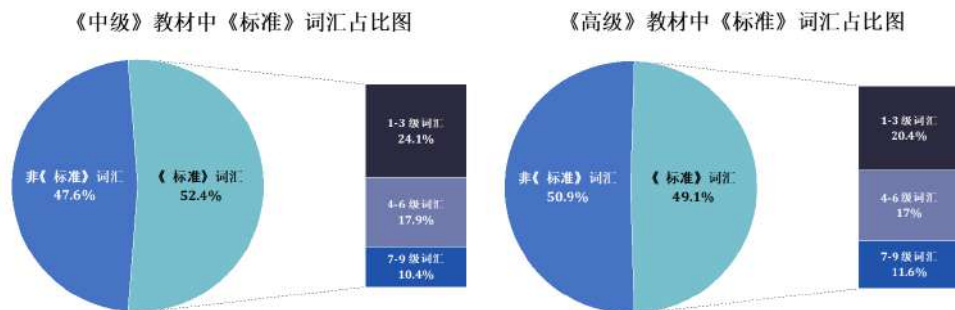


图6《中级》《高级》与《标准》词汇对比结果

相同特点:第一,两部教材非《标准》内词汇均在半数左右,《高级》非《标准》所占比例(50.9%)高于《中级》(47.6%)。第二,属于《标准》内的词汇,两部教材占比最多的是初等,其次为中等,最后为高等,各等级所占比例相当。《中级》(24.1%)和《高级》(20.4%)初级基础词汇是为了完成基本文本书写或交际需求,与现代汉语常用词相当,如“在、子、中、为、就、它、和”等,留学生对这些词没有认知和习得难度。而《中级》(10.4%)和《高级》(11.6%)高等词汇涉及到多种主题和复杂的交际任务,这部分主题达到中高级学习者的学习需求,且接触较少,如文艺、法律事务、学术研究、政策制度、国际事务等;加上非《标准》词汇达六成以上,可见科技汉语教材大量新主题及其词汇,是学习者普遍认为难学的重要原因之一。因此,从《标准》角度来看,两部教材课文词汇的编排、选取和使用等同质特征突出。

此外,我们对比教材生词表词汇与课文词汇类符发现:《中级》生词表词汇仅占课文词汇

类符的 13%，《高级》占比为 8%，生词表生词对课文词汇的覆盖率并不高。因此，课文词汇中还存在大量非生词表和非《标准》词汇内的理工类专业词汇和术语，以及人名、地名、机构名等实体类的词。采用哈工大 LTP 库进行命名实体识别发现，《中级》教材中存在约 3% 的命名实体类名词，人名如“达尔文、牛顿”等，机构名如“福特公司、通用公司、丰田公司、英国皇家学会”等，地名如“周口店、龙骨山”等。《高级》教材中存在约 4% 的命名实体类名词，人名如“多普勒、莱布尼茨、牛顿”等，机构名如“加利福尼亚大学、美国明尼苏达大学、同济大学、浙江大学”等，地名如“太和殿、故宫”等。

四、启示

（一）对“三教”的建议

从词汇丰富度、以词为单位的文本难易度、词汇复现度三个方面探讨《中级》《高级》具有专门用途的科技汉语教材词汇使用特点。研究发现：（1）《中级》词汇多样性反映的难度和课文多样性差异均高于《高级》，但此差异没有显著性区别，仍为同类型的教材；（2）《中级》词汇密度大于《高级》，两书实词词类量化排列顺序相同，名词及其搭配动词是两套教材的重点，而《高级》使用更多单位符号、网址 URL 等，符号抽象性更为突出；（3）平均词长和平均句长的说明，两部教材课文难易度相当，值得注意的是，《中级》课文之间难度系数不一致，《高级》课文难度略高于《中级》；（4）两部教材出现次数在六次以下的低频词均占 80% 以上，其中仅出现一次的词占 50% 以上；（5）出现次数在六次及以上的词中，《标准》之外的专业词汇占 20% 左右，分布在七课及以下的词汇占 70% 左右，仅出现在一课中的专业词汇约占 10%，出现六课及以上的词汇约占一半；（6）《高级》复现了 24.5% 的《中级》词汇。

1. 教材编排建议

在分析科技汉语教材的难度系数时，我们发现《中级》与《高级》教材在词汇多样性和词汇密度上存在显著差异。《中级》教材中个别课文的难度超过了整体平均水平，暴露出教材编排中难度系数的不均衡。确定教材主题后，应采用渐进式编排，逐步增加词语的数量和类型，并合理控制词长和句长，以适应学习者认知发展的规律，并确保教材难度在系列之间和课文之间的逐渐提升。

此外，词汇复现的频次和间隔是编排科技汉语教材课文的关键，据认知心理学研究，分散的复习间隔（即分布式练习效应）和在不同时间点对知识进行测试（即测试效应）有助于提高学习效率和记忆的持久性。因此，教材设计应巧妙地安排词汇复现，让学生在不同的章节和活动中反复接触关键词汇，深化学生对词汇的理解和记忆。特别是对于《中级》教材课文中的科技汉语专业词汇，建议在《高级》教材课文中以不同的形式和语境再次引入，加强学生对这些词汇的掌握，同时促进了他们对专业知识的理解。

2. 学生学习建议

学生在学习专门用途的科技汉语教材课文时,要有目的的学习和复习专业词汇,自主安排复习间隔,将短期记忆转化为长期记忆。为此,学生可采用主动学习策略,比如构建词汇网络和应用联想记忆法,加强对专业词汇的记忆。同时,鼓励学生探索多样化的自学资源,包括在线课程、多媒体材料和参与实际语言实践活动等,在提高学生语言能力的同时,增强学生的自学能力和独立解决问题的能力。

3. 教师教学建议

在科技汉语的课文教学中,教师作为引导者和促进者,应设计多样化的教学活动和真实的交流场景,让学生在实践中应用所学的专业词汇。通过案例分析、角色扮演、小组讨论等互动教学方法,帮助学生深入理解和记忆专业词汇,培养他们的批判性思维和解决问题的能力。在此过程中,教师要有意识地增加专业词汇教学和复习的次数。

(二) 对制定科技汉语领域大纲的建议

《高级》非《标准》词汇所占比例(50.9%)高于《中级》(47.6%),两书均有近 50%的词汇不在《标准》中。刘英林(1995)研究认为编写教材时应纳入词汇大纲中 75%-85%的词。可见,科技汉语教材与此观点差异较大,教材课文词汇使用特点突出。造成差异的原因:其一,与编写目的有关,两部教材主要是为了提高学习者理工类专业词汇、术语或句子结构的阅读能力,帮助学习者实现从汉语语言知识学习向高精专业知识学习能力的过渡,课程目的是为了辅助学习者理工类课程和知识的学习,因此,并不把语言交际完成水平作为最终目标,《标准》的尺度作用不明显。其二,与选择的主题有关,科技汉语主题来自理工科院校公共基础课程,教材生词以理工科专业词为主,教材课文词汇不可避免使用大量专业词或术语,以及与之相关的实体类词,而《标准》基本上不涉及专业领域的词汇或术语。其三,与建议适用的对象有关。刘文主要针对体现国际中文教育汉语水平的语言类教材而言,而科技汉语教材作为专业汉语教材,其教材性质与前者不同,其比例要求并不一定适用于科技汉语等专门专业用途的汉语教材。

根据以上研究结果可知,为了更好地满足留学生对科技领域特有的语言学习需求,科技汉语教材迫切需要一套专属大纲。我们提出了以下三个建议以供参考:

1. 增加具有中国科技特色的实体类名词

我们认为科技汉语教材大纲应加入大量反映中国科技特色的实体类名词:(1)中国著名科学家,如“屠呦呦”“钱学森”等;(2)与中国科技发展紧密相关的地点、机构,如“中国科学院”“国家重点实验室”“清华大学”等;(3)在全球产生过较大影响的科技项目名称或成果,如“嫦娥工程”“北斗导航系统”“五百米口径球面射电望远镜(FAST)”等;(4)与高新技术相关的名词,如“5G 通信”“智能制造”“绿色能源”等。这类实体名词反映了中国当前科技发展的前沿趋势,有利于学习者全面地了解中国在现代科技领域的成就和地位。

除了现代科技成就外,大纲还应加入与中国古代重大科技发明相关的名词,如“造纸术”“印刷术”“火药”“指南针”“九章算术”等,这类名词代表着中国古代社会的科技创新与文明进步,在全世界范围内产生过深远的影响。

2. 坚持实用导向, 结合“一带一路”倡议

我们认为科技汉语教材大纲应注重实用性和时效性,满足学习者在科技领域的实际需求。截止2023年8月,中国已与152个国家、32个国际组织签署共建“一带一路”合作文件,考虑到合作国家与国际组织的具体需求,我们建议将高铁技术、信息技术、环保技术等领域的专业词汇纳入教材的重点内容。将这些具有时代意义和国际视角的专业词汇纳入科技汉语教材大纲,有利于促进中国与“一带一路”沿线国家在科技领域的交流。

3. 注重专业词汇与通用词汇均衡融合

教材大纲应平衡专业术语与通用语汇的结合,确保学习者既能掌握科技领域所需的专业词汇,也能够有效地进行日常交流。除了专业领域的术语和概念外,还应包含日常生活中频繁使用的动词、名词、形容词等,促进学习者在不同语境下的灵活运用汉语的能力,提升其语言适应性。

参考文献:

- Laufer, B. & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics* 16 (3): 307-322. <https://doi.org/10.1093/applin/16.3.307>
- Saragi, T., Nation, I.S.P & Meister, G.F. (1978). Vocabulary learning and reading. *System* 6 (2): 72-78. [https://doi.org/10.1016/0346-251X\(78\)90027-1](https://doi.org/10.1016/0346-251X(78)90027-1)
- Ure, J. (1974). Lexical density and register differentiation. In Perren, G. & Trim, J. (Eds.), *Applications of linguistics: Selected papers of the second international congress of applied linguistics* (pp. 443-452). Cambridge University Press. <https://doi.org/10.1177/026009357202300310>
- 单韵鸣.(2008).专门用途汉语教材编写的问题——以《科技汉语阅读教程》系列教材为例.《暨南大学华文学院学报》(02),31-37. DOI:10.16131/j.cnki.cn44-1669/g4.2008.02.006
- 胡越,王宇晨.(2021).科技汉语教材编写模式探析——兼论科技汉语分层次课程设计.《高教学刊》(06):80-83. DOI:10.19980/j.cn23-1593/g4.2021.06.019
- 江新.(2005).词的复现率和字的复现率对非汉字圈学生双字词学习的影响.《世界汉语教学》(04). 31-38+2-3
- 教育部中外语言交流合作中心.(2021).《国际中文教育教学资源发展报告》.北京语言大学出版社.
- 刘珣.(1994).新一代对外汉语教材的展望——再谈汉语教材的编写原则.《世界汉语教学》(01). 58-67
- 柳燕梅.(2002).生词重现率对欧美学生汉语词汇学习的影响.《语言教学与研究》(05). 59-63
- 张莹.(2014).近30年科技汉语教材编写情况的回顾与思考.《出版发行研究》(11):66-68. DOI: 10.19393/j.cnki.cn11-1537/g2.2014.11.022.
- 赵金铭.(1998).论对外汉语教材评估.《语言教学与研究》(03). 4-19

(责任编辑: 陈湘云、刘洁)