

教育部人文社会科学研究“基于语料库的科技英语词汇
多维研究(12YJA740084)”项目资助



基于语料库的科技 英语词汇多维研究



 Corpus-based
Multidimensional Study on
English Vocabulary for
Science and Technology

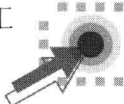
许明武 著



华中科技大学出版社

<http://www.hustp.com>

教育部人文社会科学研究“基于语料库的科技英语词汇
多维研究(12YJA740084)”项目资助



基于语料库的科技 英语词汇多维研究

Corpus-based
Multidimensional Study on
English Vocabulary for
Science and Technology

许明武 著



华中科技大学出版社

<http://www.hustp.com>

中国·武汉

图书在版编目(CIP)数据

基于语料库的科技英语词汇多维研究/许明武著. —武汉: 华中科技大学出版社, 2014. 12
ISBN 978-7-5609-8289-2

I. ①基… II. ①许… III. ①科学技术-英语-词汇-研究 IV. ①H313

中国版本图书馆 CIP 数据核字(2014)第 289852 号

基于语料库的科技英语词汇多维研究

许明武 著



策划编辑: 周清涛

责任编辑: 梅欣君

封面设计: 刘 卉

责任校对: 曾 婷

责任监印: 周治超

出版发行: 华中科技大学出版社(中国·武汉)

武昌喻家山 邮编: 430074 电话: (027)81321913

录 排: 武汉正风天下文化发展有限公司

印 刷: 湖北新华印务有限公司

开 本: 710mm×1000mm 1/16

印 张: 9 插页: 1

字 数: 180 千字

版 次: 2015 年 5 月第 1 版第 1 次印刷

定 价: 28.00 元



本书若有印装质量问题, 请向出版社营销中心调换
全国免费服务热线: 400-6679-118 竭诚为您服务
版权所有 侵权必究

内容简介

本书通过自建小型语料库,基于具体数据尽力做到客观、全面地描述和阐释科技英语词汇,重点研究科技英语词汇的信息意义、语义功能、语用功能等,探索科技英语词汇的英汉翻译规律。本书从科技文体演变谈起,尝试讨论在文体演变的影响下各类词汇的本质意义及附加意义的关联。具体探讨了科技英语名词化信息功能、科技英语中间词素的转义、科技英语中动词语义句法特征、科技英语新词的文化特征、科技英语形动兼类词形动义项间的映射关系、科技英语转述动词语用功能、科技英语半科技词汇研究、汉语科技报道中字母词信息功能、汉语科技报道中英源外来词的汉化特点等十个方面。这些研究的内容基本全面覆盖科技英语词汇领域之相关问题。

前言

随着科技的迅速发展,对于各类科技信息的捕捉变得日益重要。科技英语在信息传递的过程中形成独有的特点。这些特点体现在词汇层面、句法层面以及篇章层面。科技英语研究在上个世纪 80 年代炙手可热。研究覆盖科技文体、信息传递、英汉汉英翻译技巧等。对于科技英语词汇研究仅在部分著作中零星涉及,缺乏较为全面的关于科技英语词汇研究。现有研究多侧重语言意义层面的现象研究,缺乏较深层次的语言信息意义等方面的研究,且研究多采用定性方法,少有从定量的角度来研究此类语言现象。因此,本书较为全面地研究了科技英语词汇的特点、功能、翻译等,旨在较为明晰地梳理各类词汇与信息传递之间的关系。本书或基于一些现存语言学理论如信息理论、词素理论、言语功能等,或通过语言现象归纳总结出一些一般规律。

本书研究重点在科技文体演变、科技英语名词化信息功能、科技英语中间词素的转义、科技英语中动词语义句法特征、科技英语新词的文化特征、科技英语形动兼类词形动义项间的映射关系、科技英语转述动词语用功能、科技英语半科技词汇研究、汉语科技报道中字母词信息功能、汉语科技报道中英源外来词的汉化特点等 10 个方面。这些研究内容基本覆盖科技英语词汇领域之相关问题。

从研究方法角度来看,主要是自建语料库,基于具体数据较为客观、全面地描述和阐释科技英语词汇,突出研究科技英语词汇的信息意义、语义功能、语用功能等,探索科技英语词汇的英汉翻译规律。本书是对大量语料的研究,而非对个案进行研究。在此基础上得出的结论将具有较高的可信性和实际指导意义。研究结果将对相关语言理论在科技英语中的应用及科技英语本身(专门用途英语理论)的研究有较大的丰富作用。从实践上讲对于科技英语文献的理解和翻译有较高的实际指导意义。

本书在语料收集、数据处理、章节撰写等过程中得到了多位研究生鼎力相助,他们为本书的出版作出了巨大贡献,在此表示感谢!

由于水平有限,本书浅露之处在所难免,恳请广大读者、专家批评指正!

作者

2014 年 10 月于喻家山

目录

第一章

科技英语文体演变

1

第一节	引言	1
第二节	文体的定义及文体分析	2
第三节	科技英语的文体特征及其变化研究	2
第四节	科技英语句子层面的可量化标志性构件	4
第五节	五大可量化标志性构件的分析	5
第六节	小结	24

第二章

科技英语新闻语篇中转述动词语用功能研究 ——以《时代周刊》为例

25

第一节	引言	25
第二节	转述动词及语用功能	26
第三节	转述动词的语用功能	27
第四节	小结	36

第三章

科技英语名词单复数的转义

37

第一节	引言	37
第二节	中间类型词素“s”	37
第三节	中间类型词素“s”引起的科技英语名词单复数转义类型	39
第四节	小结	43

第四章

科技报道中中动词语义句法特征研究

44

第一节	引言	44
第二节	文献综述	44
第三节	研究设计	46
第四节	中动词语义句法特征	50

第五节	研究结果	58
第六节	小结	60

第五章

科技英语形动兼类词形动义项间的映射关系研究

61

第一节	引言	61
第二节	格式塔结构、兼类词与语料收集	61
第三节	形动兼类词形容词义项和动词义项之映射关系	64
第四节	小结	71

第六章

科技英语名词化信息功能研究

72

第一节	名词化的研究背景	72
第二节	语言信息研究	75
第三节	名词化在科技英语中的信息功能研究	78
第四节	小结	82

第七章

义位理论下的半科技词汇研究

83

第一节	引言	83
第二节	义位	83
第三节	半科技词汇	85
第四节	研究设计	88
第五节	半科技词汇词义转换	89
第六节	小结	93

第八章

科技新词文化特征分析及关联翻译策略研究

94

第一节	引言	94
第二节	研究设计	95
第三节	构词方式	95
第四节	科技新词所折射的文化	97



第五节	科技新词的翻译策略	100
第六节	小结	104

第九章

汉语科技报道中字母词信息功能探究

105

第一节	引言	105
第二节	科技报道中的字母词及信息传递	105
第三节	研究过程及数据分析	107
第四节	字母词的信息功能	108
第五节	小结	117

第十章

汉语科技报道中英源外来词的汉化特点

——以《环球科学》为例

118

第一节	引言	118
第二节	汉语外来词的历史研究	119
第三节	研究设计	121
第四节	结果与讨论	121
第五节	小结	127

参考文献	128
------------	-----

第一章

科技英语文体演变

第一节 引言

科技英语是以普通英语为基础,最终发展成为具有自身特殊文体和功能的一种全新的语言体裁。作为语言形式的一种,科技英语一直经历着变化、更新与发展。科技英语的文体特征也一直遵循变化发展的趋势。最初的科技英语风格简明,通俗易懂。然而,随着这一语言形式的深入发展,在进行科技英语文章写作时,人们竭力追求一种客观、精确及严肃的学术基调,并且制定出大量的规则来限定科技英语的文体范围,从而催生一种为大多数人所参照的文字艰涩难懂、语句冗长、句法枯燥的表达模式。

科技英语文体如今面临着至关重要的革新,其文体再度发生变化。人们开始倾向于站在与先前观点截然相反的立场上,接受与其雏形相仿的不受拘泥和易于理解的特征。

尽管前人已经就科技英语的文体特征进行过不少研究,然而目前透彻、系统地阐明这种文体特征变化的研究还甚少,对这一方面演变趋向的支持论证也极其有限。在这样的前提条件下,本书选取 1955 年与 2005 年这两年间发行的《应用物理学期刊》中的 60 篇(各 30 篇)文章作为研究对象,将研究范围缩小到其句子层面,运用包括统计频率和计算概率在内的量化与计算手段来对反映句子层面文体特征的五大标志性构件进行分析,试图通过对这两年登载的文章的比较来挖掘科技英语文体在这 50 年时间里发生的变化,并总结出其变化发展的趋势。

第二节 文体的定义及文体分析

自从文体学形成以来,其定义就存在着各种各样的说法。即便在如今,学术界仍然无法对这一概念作出统一的定义。当前最流行的文体定义如下:布封(1985)认为“风格即人”;Swift (1997)提出,“在适当的地方使用适当的词,这就是风格的真实意义”;Mukarovsky (1964)和 Widdowson (1975)的看法为“文体就是偏离”;Halliday (1971,1985)提出“文体即意义潜势”及“文体即前景化”的说法;Enkvist (1964)的观点是“文体是以最有效的方式讲恰当的事情”。

迄今为止,人们已在文体分析领域进行了许多研究。徐有志(2005)提出了分析文体时应遵循的3个步骤:①对通篇进行系统的研究,并在不同的标题下分别写下认为具有一定文体重要性的要点;②可用准确的统计术语描述一种语言特征出现的频率,如“动词缩约式出现30次”,“平均句长为10个词”,或“以粗略统计的频率来看”,用像“很少”、“通常”、“往往”、“经常”这样的量词等;③评估文体特征所具有的重要性。当一种特征比其他变体出现的频率更高时,它就显得更为重要。换言之,第一步是确定文章的体裁,总结出某一特定结构的使用频率,然后定义这种文体的特征。对文体特征,从语法层面来看,“文体特征的第二层有2个成分语法和词汇层。在语法层面下,主要讨论句型、分句句型、词组类型及词型。句型探讨的对象是传统概念上的各种句子,即陈述句、疑问句、祈使句或感叹句;简单句、复合句、复杂或复合复杂句、完整句或省略句”(徐有志,2005)。Leech (1981)对文体分析作出了如下诠释:“在确定某一特征或类别的频率时,将计算出它的出现次数用X表示,这里X代表的是某个标准量度”。“在某些情况下,以一种类别对另一种类别的出现次数的比率来表示频率同样可行”。“一些文体特征本身就是富于变化的,诸如句子的复杂性。计量句子复杂性的一种粗略简便的办法是统计各句中的平均词数,即先通过词数来确定每句的复杂程度,再由此作出推论。例如,可通过计量句长频率分布的离中趋势来决定句与句之间的复杂性差别之大”。

第三节 科技英语的文体特征及其变化研究

科技英语历经语言无止境的发展过程,已逐渐形成了自己特有的文体,包括

词汇和语言规则。前人已就科技英语文体及其文体特征进行了大量的研究,取得了以下认识:大部分文章就其语言风格作出了“鉴于科技英语很大程度上以论断的准确性和逻辑性为衡量标准,其语言务必要讲求平衡、理性客观与庄重严谨”的界定(姚旦墅、张明秋,2003)。为了达到这一目标,科技英语文体应当在词汇和语法的使用上严格把关,特别要注意句子层面的语法。科技英语对句子层面的语法有很多要求,以吴洁(2003)的观点为例,广泛使用被动语句;大量使用名词化结构、非谓语动词、长句、后置定语以及一些特殊句型,如强调句型 *it is / was that*、*as* 结构句型等;通常选用一般现在时等。何小阳(2004)进行了类似的阐释:科技英语偏重于使用长句和大段的陈述,几乎不用问句说明问题;同时要求尽可能少地使用描述性词汇,富有抒情意味的感叹句、疑问句和祈使句(2004)。王丽荣和徐艳英(2003)认为:大约 1/3 的动词用的是被动语态,固定结构 *it should be made clear that/it can be assured that* 颇为常用。据统计,句子的平均长度可达 22 个词。现在分词与过去分词都很常见等。以周春梅(1999)之见,科技英语的文体特征应包括:高频率使用动词的一般现在时;因科技英语描述的对象通常为事实、现象、过程,人称显得较为次要,大多数句子以被动语态构成;句子倾向于采用完整形式,而较少使用省略、口语形式、不完整句及缩写形式;受到科技英语以自然现象和技术原则为主题的局限,此类文章从不使用感叹句和反意句,且甚少使用拟人、夸张、隐喻及明喻等修辞手法。

还有许多研究用以说明科技英语文体特征的变化。例如,周春梅(1999)认为,任何一种文体的语言特征都不会是一成不变或绝对固定的。为了满足科学技术普及的需要,许多科技人员都赋予他们的著作一种平实的风格,使其具有尽量少用应用范围狭窄的专门术语而使用结构简单的句子以及少用 *it is that* 句型的特征。新文体提倡更多地使用如“你”和“我们”这样的人称代词来拉近读者与作者间的距离,也主张穿插使用一些修辞手法,如明喻、拟人等,从而使语篇读起来更生动,更容易被读者接受。范瑜与李国国(2004)探讨过科技英语的新文体。他们主张用主动语态取代被动语态,大量引用古今中外的权威论证,并最终让人对科技英语的演变趋势和科技工作者应对这种改变的策略产生明确印象。他们在文章中讲到:“语言文体永远不可能保持不变,科技英语文体亦是如此。”随着科技的发展及文化的繁衍,科学技术领域正尝试着摒弃机械保守、一板一眼的写作风格,取而代之形成一种更自然、简洁、朴实的风格。

随着世界政治、经济、文化和科技的迅猛发展以及不同语言间相互影响的渗透,英语在诸多方面都发生了不同程度的变化,引领着一种全新的潮流。这一趋势在科技英语领域体现得尤为明显:①新词汇大规模涌现;②短语与其他结构间的搭配更为灵活;③句子结构变得更加简明准确,省略增多,有时会关注准确性而不是死扣句子结构;④标点符号的使用更为自由,如大量状语从句不加逗号地

出现在主句前后等(北京工业学院外语教研室,1981)。

张中元与周吉红(2005)提出:“为了使科技英语写作风格更自然达意,建议作者少用被动语态,更多地使用短句和不断变换句中结构以吸引读者。”

科技英语文章中新词的涌现、外来词的借入以及各种修辞手法的应用,都是科技英语文体特征发生变化的最好证明。特别是在近代,随着科技的发展,科技英语的文体特征产生了显著的改变。

第四节 科技英语句子层面的 可量化标志性构件

侯维瑞(1989)认为:“与其他类别的文体一样,对科技英语语法特征的研究应包括以下方面:对句长的研究,对并列与主从关系的研究,对从句类型的研究,对从句次序的研究,对句子中主句与从句的相对出现频率的研究,对主语和谓语之间位置关系的研究,对包括句首、句中及句尾在内的各部分位置和句子中具有副词性质的修饰语包括副词短语和介词短语数量的研究,以及对诸如被动句和‘it is that...’这样的特殊句型的出现频率的研究。”北京工业学院外语教研室编写的《科技英语基础语法》一书也有对这些结构的阐释:“句法主要是研究与词的用法、句的构成、句子中词与词组间的关系以及句子的种类和结构相关的一些法则。”(1981)通过以上的叙述,辅之以与文体及文体分析相关的知识,人们可以明白无误地获知科技英语文章句子层面主要的可量化的简单结构,即句子长度,句子中主句与从句的相对出现频率,包括主语、谓语、副词修饰语等由插入成分代表的句子结构间的相对位置关系,省略结构的出现频率,虚拟语气的出现频率及疑问句的出现频率。这些结构在许多相关著作中都有论述。以《科技英语基础语法》为例,科技英语文体作为英语的一种特殊文体,在句子层面具有这样几个方面的特征:①经常使用陈述。为了说明自然现象及进化规律,科学学者在写科技英语文章时应给出确切的定义和公式,故而使用陈述最多。换言之,科技英语文章中的疑问句和虚拟句应当少之又少。因此,本书将统计此类文章中疑问句和虚拟句的出现频率,用以说明科技英语文体在句子层面发生的变化。②经常使用被动语态。通常人们关注的对象或主题是一个物体、一个过程,而不是某个人,因此,在淡化人称而强调物体时,被动语态更能派上用场。范瑜与李国国(2004)已对此进行了详尽的阐述。他们批评了大量使用被动语态的做法,并以充分的例证证明当今的科技工作者已开始更多地使用主动句,从而说服作者较多地选用主动语态。此外,Leech(1981)以从航空领域的一份权威杂志中选取的

大量科技文章作为语料研究了主动语态的使用问题,得到了主、被动语态的出现频率,以此向读者清晰地说明了近代科技工作者对这两种语态的使用偏好以及科技英语文体特征的发展趋势。所以,为避免重复研究将不再讨论主、被动语态间的平衡问题。③经常使用词组和从句。为了阐明句子内部及句与句之间的各种关系,如时间关系、条件关系、因果关系、目的关系和对比关系等,科技英语文章采用大量的介词短语、非限定性动词短语及独立短语等,使简单句复杂化,从而催生出更多的复杂句,甚至出现更多的并列从句或彼此间相互限定的从句。鉴于科技英语文章中句子具有的复杂性,本书旨在进一步统计得出从句出现的频率,发掘从句的变化,进而为科技工作者找到一种更精准易行的写作模式。④经常使用省略。科技研究的对象之间的关系错综复杂,因此科技英语语言的准确性至关重要。通常情况下,具有准确性的省略可以通过名词词组替代句子,分词替代从句及省略结构替代完整句来实现。然而,根据很多语法教材的观点,省略作为一种特殊的语法现象,被认为不适用于科技英语文章。因为较多地使用省略不仅会导致句子结构与整体意思的不完整,而且会造成对文章的曲解。那么,使用省略的实际情况究竟是怎样的呢?本书将对科技英语文章中出现的省略进行具体分析,希望能发现或总结出一个明确的事实,来解决这一问题。

由此便得到了量化科技英语文体特征变化的五大构件:句子长度,句子中主句与从句的相对出现频率,省略结构的出现频率,倒装结构的出现频率以及虚拟语气的出现频率。

第五节 五大可量化标志性构件的分析

本书以句子长度,句子中主句与从句的相对出现频率、省略结构、倒装结构和虚拟语气的出现频率等5大构件为分类标准,通过比较从1955年与2005年的文章中得出的数据,试图说明科技英语的文体变化及总结出其发展趋势,归纳出科技英语文章的写作风格。

数据处理包括3个步骤,即把数据编码,进行数值计算及准备较为成熟的数据展示。每列数值反映的均为各个可量化标志性构件的普遍状况。随后,作者将这些数据按几个方面进行了分类。例如,通常将与某一构件有关的数值分入几个不同的区间。解读数据之后,再对这些数据进行几种数值计算。其中最常用的是概率计算。概率用各个数字对应的百分数表示,而百分数则是由此公式得出的: $\text{百分率} = 100 \times \text{与一个值相关的数字} / \text{构件中包含的总数}$ 。除了计算各

个值的概率以外,还需算出各构件的平均概率、最大概率以及最小概率。

一、句子长度▶▶▶

对句子长度进行研究的程序如下:

① 逐一计算出每篇文章里各个句子的长度,也就是组成各句的词数。对于难以量化的所有图表与公式则忽略不计。

② 将句子长度分为 6 个区间,分别是:词数小于 10 个的区间,词数介于 10~20 的区间,词数介于 20~30 的区间,词数介于 30~40 的区间,词数介于 40~50 的区间及词数超过 50 的区间。再逐篇统计出属于各区间的相关句子数目。

③ 计算出每篇文章中各区间内的句子数目在句子总数中的出现概率。

④ 最后计算出各区间内数值的平均概率。

这些概率是证明句子长度变化及变化的相关趋势的主要参数。

下列表格里列出的数值为统计结果。表 1-1 与表 1-2 分别表示的是 1955 年与 2005 年选用的文章中的句子数目在各区间内的分布情况。表 1-3 与表 1-4 则分别为 1955 年与 2005 年选用的文章中各区间内的句子概率统计表。表 1-5 则是各区间内平均句子概率的统计。

表 1-1 1955 年每篇文章中句子数目在 6 大区间内的分布情况

篇号/区间	<10	10~20	20~30	30~40	40~50	>50
1.	26	52	26	5	1	2
2.	4	5	9	7	1	0
3.	5	25	24	18	5	6
4.	2	28	31	15	6	5
5.	13	82	82	44	20	4
6.	4	24	16	13	5	1
7.	15	51	47	20	5	2
8.	8	42	40	17	10	3
9.	5	17	26	8	12	3
10.	9	52	55	18	10	3
11.	5	4	17	4	1	0
12.	3	18	9	0	0	0

续表

篇号/区间	<10	10~20	20~30	30~40	40~50	>50
13.	8	54	41	54	16	9
14.	8	51	59	40	18	7
15.	5	52	61	25	7	2
16.	0	3	3	0	0	2
17.	1	20	30	19	5	3
18.	19	91	56	11	8	2
19.	13	79	50	14	6	4
20.	8	48	43	15	7	1
21.	5	31	22	26	12	7
22.	2	27	48	27	15	7
23.	3	8	23	14	7	5
24.	0	11	44	24	13	1
25.	11	44	24	13	1	2
26.	5	23	27	7	3	0
27.	0	9	10	6	0	3
28.	13	50	29	13	4	3
29.	6	39	38	10	0	1
30.	4	49	51	10	1	4

表 1-2 2005 年每篇文章中句子数目在 6 大区间内的分布情况

篇号/区间	<10	10~20	20~30	30~40	40~50	>50
1.	5	39	49	35	15	11
2.	5	35	31	17	3	0
3.	9	74	63	33	9	8
4.	4	62	87	40	23	9
5.	7	40	31	10	4	2
6.	5	61	40	27	7	2
7.	11	52	31	8	0	0
8.	0	6	18	14	11	4

续表

篇号/区间	<10	10~20	20~30	30~40	40~50	>50
9.	6	53	63	39	19	8
10.	4	60	60	27	4	2
11.	3	46	26	7	2	1
12.	12	62	23	5	2	0
13.	3	29	52	7	7	2
14.	3	35	34	13	2	1
15.	1	33	38	25	11	8
16.	21	57	40	16	1	0
17.	3	28	48	15	6	5
18.	6	44	41	11	3	0
19.	20	80	33	10	3	0
20.	8	48	60	19	1	1
21.	14	76	71	23	12	1
22.	0	18	20	19	7	1
23.	7	66	51	16	3	3
24.	2	73	50	16	4	3
25.	40	128	49	13	1	0
26.	3	36	56	19	14	7
27.	7	67	50	20	2	0
28.	20	76	36	5	1	0
29.	15	66	66	23	2	1
30.	2	12	20	8	8	2

表 1-3 1955 年每篇文章中句子数目在 6 大区间内的概率分布情况

篇号/概率	<10	10~20	20~30	30~40	40~50	>50
1.	0.232	0.464	0.232	0.044	0.009	0.018
2.	0.154	0.192	0.346	0.269	0.038	0
3.	0.060	0.301	0.289	0.217	0.060	0.072
4.	0.023	0.322	0.356	0.172	0.069	0.057

续表

篇号/概率	<10	10~20	20~30	30~40	40~50	>50
5.	0.053	0.335	0.335	0.180	0.082	0.016
6.	0.063	0.381	0.254	0.206	0.079	0.016
7.	0.107	0.352	0.324	0.138	0.034	0.014
8.	0.067	0.350	0.333	0.142	0.083	0.025
9.	0.070	0.239	0.366	0.113	0.169	0.042
10.	0.061	0.354	0.374	0.122	0.068	0.020
11.	0.161	0.129	0.548	0.129	0.032	0
12.	0.1	0.6	0.3	0	0	0
13.	0.044	0.297	0.226	0.297	0.088	0.049
14.	0.027	0.279	0.322	0.219	0.098	0.038
15.	0.033	0.342	0.401	0.164	0.046	0.013
16.	0	0.375	0.375	0	0	0.25
17.	0.013	0.256	0.345	0.244	0.064	0.038
18.	0.102	0.487	0.299	0.059	0.043	0.011
19.	0.078	0.476	0.301	0.084	0.036	0.024
20.	0.066	0.393	0.352	0.123	0.057	0.008
21.	0.048	0.301	0.214	0.252	0.117	0.068
22.	0.016	0.214	0.381	0.214	0.119	0.056
23.	0.050	0.133	0.383	0.233	0.117	0.083
24.	0	0.118	0.473	0.258	0.140	0.011
25.	0.116	0.463	0.253	0.137	0.010	0.021
26.	0.077	0.354	0.415	0.108	0.046	0
27.	0	0.322	0.357	0.214	0	0.107
28.	0.116	0.446	0.259	0.116	0.036	0.027
29.	0.064	0.415	0.404	0.106	0	0.011
30.	0.034	0.412	0.428	0.084	0.008	0.034