

英语新闻标题词汇层面的计算文体学研究

翁玉秀

(集美大学 诚毅学院, 福建 厦门 361021)

[摘要] 词汇是语言结构的基本单位,是表达意义和实现交际目的的工具。新闻英语作为信息传播媒介,而新闻标题是对新闻信息的浓缩和概括,有其独特的用词特征。通过自建语料库,采用语料库驱动的计算文体学研究方法,量化分析新闻标题的词长、词汇密度和词频使用特征。研究发现新闻标题以短小词为主,平均词长 5.13;其 STTR 值为 63.07,表明标题词汇密度大、用词丰富多样;词频分布分析发现新闻标题中功能词常被省略、没有第一和第二人称代词、“say”词使用频繁、新闻标题词汇具有时代特征。同时本研究还分析了这些词汇特征产生的文体效果和成因。

[关键词] 英语新闻标题;计算文体学;词长;词汇密度;词频

[中图分类号] H 313.1

[文献标识码] A

[文章编号] 1008-889X (2018) 02-0122-08

专门用途英语 (English for Specific Purposes, 简称 ESP), 是英语语言学领域的一个分支学科。新闻英语是 ESP 的一个重要分支, 在外语教学与研究中, 有着不可忽视的重要地位。在语言学领域, 对新闻英语语言特征的研究已有很长的历史。而新闻标题作为新闻的“灵魂”和“眼睛”, 是对新闻报道的浓缩和概括, 具有提示、评价新闻内容和吸引、引导读者的作用。由于新闻标题在新闻报道中的独特地位, 历来受到专家学者的重视, 许多国内外的学者从不同的角度对新闻标题进行了深入研究。有些学者从语言学的角度概括其词汇^[1]、句法^[2]、修辞^[3-4]等层面的语言特征; 还有些学者主要研究新闻标题的功能, 例如 Iarovici 和 Amel 探讨了标题的双重功能^[5]; Rew 研究标题中标点符号的功能^[6]; 而 McIntyre 总结了标题的五要素功能^[7]。另外有些学者研究特定的某一类标题, 如陶雪城、张萍研究 NBA 新闻标题的特点^[8]; 或者标题的某一层面特征, 如刘改琳、李定顺研究标题中的词汇分布特征^[9]。

词汇是语言结构的基本单位之一, 是表达意义和实现交际目的的工具。Halliday 认为, 词汇特征是一个语体最突出的特征, 甚至可以单独作

为这一语体语言特征的标志符。^[10]新闻标题作为新闻的浓缩精华所在, 能帮助读者以最小的努力获得最佳的语境效果, 其措辞往往简洁而具体, 行文流畅而不佞屈聱牙。新闻标题的语言具有时效性 (timeliness)、简洁性 (succinctness)、醒目性 (conspicuousness) 等特点。^[11]对新闻标题的词汇特征进行量化分析, 有助于读者把握其用词特点和措辞风格, 提高新闻浏览效率。然而, 在 CNKI 中, 笔者以“英语新闻标题词汇 (特征)”为关键词进行检索, 发现相关的中文期刊文章不多, 这说明新闻标题的词汇特征问题并没有引起国内学者足够的关注。进一步分析这些文章发现: 只有极少数是基于语料库的研究, 其余皆是笼统的定性阐述, 可见对英语新闻标题的词汇特征的量化分析的研究很少。传统的定性研究根据经验对文本进行分析, 存在主观判断, 研究结果缺乏说服力。因此, 在新闻标题的词汇特征研究中引进量化计算的方法亟待进行。

一、理论背景

20 世纪 30 年代计算主义的发展, 使人们意识到整个自然界都是由算法构成的, 定理证明、

[收稿日期] 2017-02-14

[修回日期] 2017-03-10

[基金项目] 福建省中青年教育科研项目 (JAS160121); 集美大学诚毅学院教改项目 (C16075)

[作者简介] 翁玉秀 (1979—), 女, 福建漳浦人, 讲师, 硕士, 主要从事应用语言学和英语教学研究。

文字翻译等都是计算。信息技术的发展加上文体分析中新兴的统计理念,逐步催生了倾向于统计的分析方法。研究者希望通过机器全面、快速地找出文本标记,这种方法被定义为计算文体学(Computational Stylistics)或语料库文体学(Corpus Stylistics)。^[12]计算文体学作为计算语言学的分支,真正兴起于20世纪60年代末,它将计算机技术和计算主义思想运用于语篇分析中,为理论分析找出科学的数据来证明。通过科学的计算技术手段发现那些被传统方法所忽视的语言现象,在分析模式上加入了计算语言学、语料库语言学和统计学的因素,使得有了数据支持的研究结果更可靠、更具说服力。例如,Mosteller和Wallace利用对高频词汇的概率分析,测定出当时有预谋的12篇匿名的“联邦主义者”文件是美国第四任总统Madison所写,而非政治家Hamilton。^[13]Hori采用Dickens的小说语料库,研究Dickens独具特点的搭配形式。^[14]计算文体学通过统计方法为文本分析提供数据支持,成功鉴定作者的身份、鉴别作品文体,弥补传统定性分析在数据统计方面的不足,使研究结果更具可信度和说服力。

通过对已有研究的检索发现:计算文体学的研究对象基本为基于语料库的长篇小说作品,如小说、诗歌、散文、杂志评论文章、信件、科技文章等,鲜有以简短的新闻标题为研究对象的相关研究。究其原因,大致是由于标题短小,收集起来费时费力,目前还没有专门的英语新闻标题语料库。本研究通过自建英语新闻标题语料库,采用计算文体学的方法研究新闻标题词汇层面的语言特点;利用计算机检索从微观入手,获得大批单项或多项语言特征的实例,分析其出现的频率、分布规律和语境特征,由此归纳出新闻标题词汇层面的整体文体色彩和措辞风格。通过计算机技术发现英语新闻标题词汇中特有的语言现象,如功能词、“say”词、人称代词等的使用频率,为阐释新闻语体,特别是标题语域特有的用词特征提供量化依据。另外,实践教学经验告诉我们,目前许多外语能力测试(大学英语四/六级、雅思、托福等)都以报刊文章为素材,考生对报纸文章的理解效果却不尽如人意,究其原因很重要的一点在于词汇的理解存在障碍。本研

究通过结合定量统计和定性分析新闻标题词汇特征,以期减少英语新闻阅读者和英语学习者词汇理解方面的困难,同时为《英文报刊选读》课程教学提供理论和现实帮助。

二、研究设计

本研究采用语料库驱动的计算文体学分析方法,以真实语料为研究对象,对选取自纸质版的《纽约时报》和美国新闻网站YAHOO News(雅虎新闻网站)的2000个新闻标题的词汇特征进行实证研究。

1. 研究问题。英语新闻标题是对新闻内容的浓缩和概括,它在词长、词频、词汇密度方面有什么特征?为什么会产生这些词汇特征?这些词汇特征会产生怎样的文体效果?本研究利用语料库分析软件就这些问题进行量化统计得出结果并加以归纳分析。

2. 研究对象。本研究以纸质版《纽约时报》和美国新闻网站YAHOO News的新闻标题为研究对象。《纽约时报》作为美国严肃报刊的代表,在美国乃至全球都具有很高的发行量。由于其对国际、国内重要事件详细且忠实的报道,使得它在国内外都拥有极高的权威性和公信力。而雅虎新闻网站是美国门户最大的新闻网站,集结了全世界主流的新闻机构,包括美联社、路透社、法新社、华盛顿邮报等。本研究从上述两个媒体中各选取1000个新闻标题,建立新闻标题语料库。语料库的建立严格按照Allan Bell对新建语料库的原则要求:(1)语料收集要有一定的时间长度;(2)该时间长度内要有不定次数的收集行为;(3)对同一板块的内容要有持续时间段的收集。^[15]本研究所选取的标题时间段为2015年10月至2016年2月,在该时间段内不定时对同一板块的新闻标题进行随意选取,以保证语料的代表性。最终收集到2000个标题,共18103个字符数、5364个类符数。本研究选取的标题覆盖面广,涵盖了国际新闻、国内新闻、政治外交、财政经济、商业社交、教育文化、科学技术、体育健康、环境卫生等。

本研究选取Brown语料库作为参照语料库。该语料库收集了20世纪60年代有代表性的美国

英语语料,语料选自各种出版物,建库时考虑到了各种文体的平衡,严格按照随机原则取样,是一个标准语料库。本研究选取 Brown 语料库共 1 202 022 个字符数、44 403 个类符数进行对比分析。

3. 研究方法和工具。本研究结合定量和定性分析,利用语料库分析软件和定量统计方法,从词长、词汇密度、词频三个维度对新闻标题语料和 Brown 语料库进行实证对比分析。本研究中,主要使用 AntConc 和 WordSmith 两个语料库软件进行语料库基本统计、语料检索、词类考察等。其中, AntConc 用于统计分析两个语料文本的词频。WordSmith 用于统计分析语料的词长、类/形符比(TTR)以及标准化的类/形符比值(STTR)。

三、数据分析与讨论

(一) 词长

词长(word length)是词汇特征的一个重要衡量指标。词长在一定程度上决定了文本的文风和难易程度。新闻用语作为特殊的语域,其词长有自己的特点。同时由于受到新闻时效性和信息传递量的约束,编辑在标题的遣词造句过程中力求简明扼要、精炼通俗,从而节省篇幅以减少成本并节省受众的浏览时间。通常,测量词长有两种方法:计算单词的字母个数和计算单词的音节个数。本研究采用计算字母个数的方法测量单词的长度。通过运行 WordSmith 获得图 1。

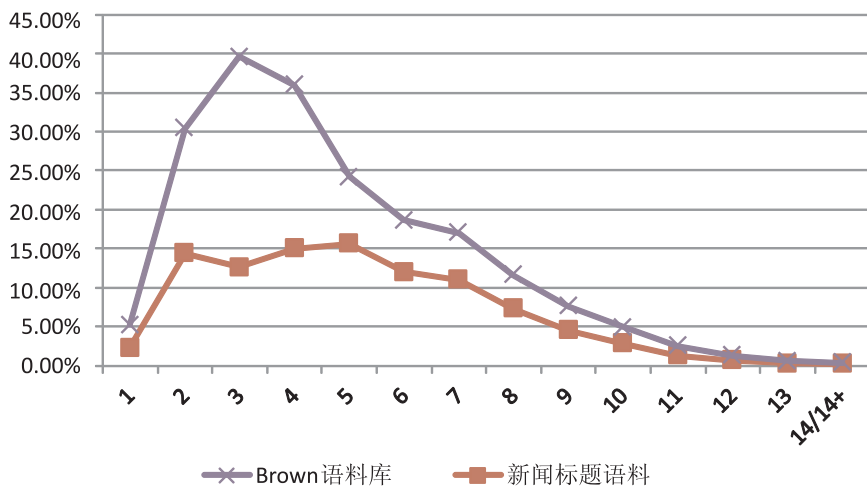


图1 新闻标题语料和 Brown 语料库的词长对比

图1显示,随着字母个数的递增,单词数量呈递减分布。新闻标题的单词集中在词长为2—5个字母的单词之间,占57%;Brown语料库的单词集中在2—4个字母的单词之间,占64%。可见,两个语料文本的单词长度都集中在2—5个字母之间。曾从事过新闻工作的美国著名作家马克·吐温说过:“普通英文单词的平均长度为四个半字母。”^[16]语言学界普遍认为2—5个字母的单词为短小词(midget word)。因此,新闻标题的用词以小词为主。通过再次运行 WordSmith,获得两个语料文本的平均词长分别为5.13和4.48。新闻标题的平均词长稍长于

Brown 语料库。说明比起普通英语,新闻标题的措辞相对更正式、专业,难度也相应地更高。但整体上看新闻标题的词汇以短小词为主。其原因有三:(1)从文体学的角度分析,这与新闻标题的写作理论和实践相关,早在1972年,Evans就指出新闻编辑应该首选词形简单短小,词义具体单一,有盎格鲁-撒克逊起源的词而非字形复杂抽象的外来词。^[17](2)从新闻语体和标题功能的角度分析,通俗易懂是新闻语言用词的首要原则。新闻受众是一般读者,短小词利于读者阅读,单词越短读者越容易抓住信息。新闻报道五要素中,动词描述新闻主体的发展(how),特

别是短小动词和动词短语的高频使用，搭配揭示新闻事件主体（what）的名词，如此一动一静结合，完美传递新闻事件的要义。（3）从新闻出版业的角度考虑，由于排版需要，报纸版面空间有限，为节约版面空间，编辑们自然首选短小精悍、铿锵有力、字形短小的单音节名词、动词等。从以下例子可窥一斑：

例1 Two European Drug Makers in Talks about Asset Swaps

（《纽约时报》，2015年12月16日）

例2 British PM Defends Google Tax Deal

（雅虎新闻网站，2016年1月26日）

例1中，talks和swaps都是5个字母的小词，分别意为negotiation和exchange。例2中，缩略词PM的完整形式为Prime Minister，deal意为exchange。显然两个标题例子中所选用的词语更为简洁易懂，阻碍读者阅读的可能性小。同时小词的使用使整个标题视觉上简洁大方，协调对称，易于排版。信息时代背景下，读者每天接触大量信息，费事耗时的新闻标题自然不会得到青睐。因此，新闻标题中多使用生动、灵活、语义固定的短小词，少用或不用意义空泛、概念抽象的词。美国语言学家H. L. Mencken把这类经常使用于新闻标题的简短词称为“万能同义词（synonyms of all work）”。^[18]以上两例中，这类简短词常用于替换与其同义的较长词汇。如：deal替代negotiation；hit/harm/hurt替代damage；blast/crash/ram替代explode；aim常用于替代purpose、intention、object；cop替代policeman/policewoman等。

（二）词汇密度

词汇密度（lexical density）是词汇丰富性的四个观测变量之一，是语言应用特点的重要表征。在语料库语言学中，词汇密度用于分析或衡量语料中的词类范畴和种类。一般来说，语料文本的类/形符比越高，说明其用词变化性越大，词类越丰富。关于词汇密度的测量，学界存在两种方法：一种是利用类/形符比（TTR）来测量。即语料库中所有类符（type）与所有形符（token）的数量比。由于存在这样的事实：一定时期内，语言的类符不会发生大的变化，所以语料库越大，其类/形符比就越小。因此，Scott提出

了标准化的类/形符比（STTR）的测量方法，即将语料文本分为若干个特定的基数单位（Word-Smith的默认基数为1 000）来计算类/形符比值，最后取其平均值。^[19]用STTR来计算词汇密度，有效地避免了因语料库的大小而产生的对TTR的影响。另一种测量词汇密度的方法是由Stubbs提出的，是一种根据语料文本中的实词数与总词数的比值来测量语料库的词汇密度的方法。^[20]限于篇幅，此种方法不作赘述。本研究中，由于两个研究对象语料文本大小存在较大差异，故采用Scott的STTR测量法。根据Word-Smith的分析结果，新闻标题语料和Brown语料库的词汇密度如表1所示。

表1 新闻标题语料和Brown语料库的词汇密度

考查项	新闻标题	Brown语料库
形符数	18 103	1 202 022
类符数	5 364	44 403
类/形符比值（TTR）	29.63	4.24
标准化类/形符比值（STTR）	63.07	39.01

从表1的数据可见，新闻标题和Brown语料库的类/形符比值分别为29.63和4.24，而两者的标准化类/形符比值（STTR）分别为63.07和39.01，两个语料在词汇密度上显示出较大的差距。如前所述，标准化类/形符比值越高，语料文本的词汇密度越大，可见新闻标题的词汇密度要远高于Brown语料库。说明比起普通英语，新闻标题的用词更丰富多样。其成因主要有三方面：（1）本研究所用语料选材范围大、覆盖面广，涉及国际政治、经济、商业、文化、科技、体育、卫生等。（2）新闻标题作为新闻报道的浓缩，为保证新闻报道信息传递的最大化，措辞上倾向多样化。（3）为使标题立意新颖、吸引眼球，编辑在遣词造句上力求词汇多样性和丰富性。而Brown语料库作为普通英语的代表，其用词重复性高，因此STTR较小。

（三）词频

词频（word frequency）是指特定单词在文本中出现的次数，某种程度上反映了一个文本的语体特征，是发现文体或语体特征的重要参考指

标。Johansson 对比分析了学术文本语料库和小说文本语料库中词频最高的 50 个词。发现: ox-ides、constants 和 equations 这些词汇属于学术文本语料库,而 mister、cheek、sofa 和 wallet 属于小说文本语料库。^[21]这说明语体、语境不同,其

高频词也不同。新闻语体中高频词实际承担一定的语篇功能。通过 AntConc 的单词列表功能,检出两个语料文本的前 50 个高频词,因篇幅所限这里仅列出前 5 个以及一些具有典型特征的词项(见表 2)。

表 2 新闻标题语料和 Brown 语料库的前 50 个高频词列举

新闻标题			Brown 语料库			新闻标题			Brown 语料库		
排序	词项	频率	排序	词项	频率	排序	词项	频率	排序	词项	频率
1	In	561	1	The	70 002	24	Trump	45	24	Are	5 348
2	To	523	2	Of	36 472	25	It	44	25	But	5 147
3	Of	397	3	A	32 003	29	Zika	41	29	They	3 762
4	A	356	4	And	28 935	31	Police	39	31	You	3 634
5	The	300	5	To	26 239	36	Deal	34	36	Her	3 037
16	Says	69	16	On	6 766	48	Talks	30	48	More	2 226
21	China	55	21	This	6 387	49	Clinton	29	49	No	2 219
23	Court	48	23	Not	5 382	50	President	29	50	If	2 199

表 2 显示,新闻标题文本的前 50 个高频词中 31 个为功能词,占 62%; Brown 语料库的前 50 个高频词皆为功能词。标题文本的前 5 个功能词为 in、to、of、a 和 the。In 作为经常搭配时间和地点状语的介词,在新闻用语中出现频率很高,位列第一。To 具有作为介词和不定式的符号的双重功能,在语言中使用广泛,排在第二

位。Brown 语料库的前 5 个功能词为 the、of、a、and 和 to。分析两个语料的 50 个高频词,总结英语新闻标题高频词的使用特点如下。

1. 功能词省略现象明显。为进一步研究新闻标题中功能词的使用情况,将两个语料文本的 50 个高频词中的功能词按词性细分(见表 3)。

表 3 新闻标题语料和 Brown 语料库 50 个高频词中的功能词分布

语料来源		新闻标题		Brown 语料库		
词性	单词	个数	比例	单词	个数	比例
介词	in/to/of/for/on/with/ at/ by/ over/from/ up/into/out/against	14	28	of/to/in/for/with/ on/at/by/ from	9	18
连词	and/as/ after	3	6	and/ as/or/if/ but	5	10
冠词	a/ the/ an	3	6	the/ a /an	3	6
否定词	no/not	2	4	no/not	2	4
be 动词	is/be/are	3	6	is/was/be/were/been/ are	6	12
情态动词	will/may	2	4	would/will	2	4
人称代词	his	1	4	he/his/I/they/you/her/she/ we/ their/him	10	22
疑问代词	who/ how	2	4	which/ when/ who	3	6
指示代词	it	1		it/that /this	3	4
助动词				have/has/had	3	6
其他				one/ all/ there/more	4	8
总计		31	62		50	100

从表3可见,两个语料文本的功能词中存在诸多重叠,从比例对比可见,新闻标题语料中的功能词在50个高频词中的比例远低于Brown语料库。因为新闻标题中功能词省略现象明显。首先,从文体学的角度分析,新闻语体要求简洁易懂,为使标题语言简练、短小,一些不妨碍读者理解的功能词常被省略;其次,从新闻出版的角度看,报刊杂志的版面空间极其宝贵,从节约空间以及方便排版的角度考虑,编辑在写作技巧上倾向于选择实义词而略去诸如冠词、助词、连词等功能虚词。

例3 Glen Sonmor, 86, a Brawling N. H. L. Coach

(《纽约时报》, 2015年11月16日)

例4 Teen Shot by Chicago Cop Called 911 Three Times Asking for Help

(雅虎新闻网站, 2016年1月26日)

例3中,书面完整的句子应为:Glen Sonmor **who is 86 years old, is** a Brawling N. H. L. Coach。而在实际出版时省略了定语从句的引导词(who)、谓语动词(is),以及主句的谓语动词(is)。例4中,完整句子为:**The Teen who was** Shot by Chicago Cop Called 911 Three Times

Asking for Help。在不妨碍读者理解的前提下,雅虎新闻网站的编辑发布该条新闻时省略了先行词的定冠词(the)、定语从句引导词(who)以及表示被动语态的be动词的第三人称单数(was)。从例3、例4可以看出,新闻标题中省略功能词的现象普遍,且省略了功能虚词的新闻标题语义简明扼要,句式精干对称而不繁冗拖沓,版面干净利落、节省空间且吸引眼球。

2. 第一和第二人称代词缺失。对比研究两个语料的前50个高频词可见,Brown语料库中有10个人称代词(he、his、I、they、you、her、she、we、their、him),覆盖了所有的人称;而新闻标题中只出现一个第三人称的人称代词(his),如表4所示。这个研究结果验证了1981年Simon-Vandenberg发现的事实:“新闻标题中没有第一和第二人称代词。”^[22] Quirk指出:“话语或者语篇越私人化,第一、第二人称出现的频率就越高。”^[23] 新闻标题是一门独特的语言,用客观的语言传递信息。标题中编辑主要陈述事实而不对新闻本身品头论足,以示新闻的客观性和媒体的中立立场。因此,为保证报道的客观性,在表达形式上避免使用富有感情色彩和主观判断的描述性词汇,包括人称代词。

表4 新闻标题语料和Brown语料库的人称代词对比

词项	he	his	I	they	you	her	she	we	their	him
Brown 语料库	9 797	7 000	5 935	3 762	3 634	3 037	2 984	2 846	2 671	2 619
新闻标题		32								

3. 新闻词汇具有时代特征。新闻标题语料的前50个高频词中有19个实义词,占50个高频词的38%。其中16个名词(16个名词中有些与政

治相关,例如,president、police、court、case;有些与商业经济相关,如year、deal、talk),1个动词和2个形容词(见表5)。

表5 新闻标题语料50个高频词中的实义词

词性	词项	数量	比例
名词	US/China/court/Trump/man/Zika/police/year/city/deal/case/state/Iran/talks/Clinton/president	16	32
动词	says	1	2
形容词	New/south	2	4
总计		19	38

新闻词汇体现了新闻语言的时代感和实效性特点。例如 Clinton、Trump、President 和 Zika 这几个词还原了标题选取时间段内国际上发生的大事件。前三个词展现了 2016 年美国社会的大事件——总统大选,作为世界超级大国总统候选人的希拉里·克林顿和唐纳德·特朗普自然成为媒体关注的焦点。而当大选慢慢退出人们的视线,这些词汇可能被人们慢慢遗忘。Zika 是波及了几乎整个美洲的致命性病毒,关乎民生,自然也是媒体关注的焦点所在。另外,在 16 个实义词中位列第一的是 China,在前 50 个高频词中位列第 21 位,出现 55 次,这说明随着全球经济一体化的发展以及中国国际影响力的不断提升,与中国相关的新闻成为美国媒体和国际媒体关注的焦点,因此许多与中国相关的新闻见诸报端。这些词汇特点体现了《纽约时报》和雅虎新闻网站的时代性和严肃性,也体现了新闻标题词汇的时代性特征。

4. “say”词的高频使用。表 5 中的数据显示,实义词中只有一个动词 says, says 是 say 的第三人称单数形式。在新闻语言中编辑常需要直接引用新闻当事人或者说话者的言论,以示新闻的客观性和可靠性,因此新闻中常有…says/say/said…句式。通过运行 Microsoft Excel 的查找功能,获得含有 say 的词条搜索结果 87。也就是说,除了 50 个高频词表中显示的 69 例的 says 之外,还有 18 例的 say;进一步搜索 say 的一般过去式 said,没有得到相关的词条。

新闻标题中高频使用的 say 词还原了新闻事件,让标题更显客观、具有说服力。标题中还有一些词语与 say 有异曲同工之妙处,如 testify、warn、imply、find、argue 等,但是相比之下,后者更长、意义更丰富,不如 say 简短且意义单一稳定,易于理解和阅读。

例 5 WHO Says Zika Likely in all but 2 Countries in Americas

(雅虎新闻网站,2016 年 2 月 19 日)

例 6 Leading a Nation May Mean a Shorter Life, A Study Says

(《纽约时报》,2015 年 12 月 15 日)

例 7 Vaccine for Zika Virus May Be Years Away, Disease Experts Warn

(《纽约时报》,2015 年 12 月 10 日)

例 8 Some Afghans View Taliban as the Key to Aid, Study Finds

(《纽约时报》,2015 年 12 月 16 日)

例 5、例 6 中编辑用 says 来直接引用新闻当事人的话语,站在第三者的角度凸显新闻叙事性报道的本质特征,使报道更具客观真实性和权威性。例 7、例 8 中的 warn 和 finds 在功能上等同于 say,凸显记者站在第三者的角度报道新闻事件的事实,不加入个人评论和感情色彩,显示媒体的中立态度和报道的可靠性和严肃性。

四、结 语

本研究基于真实的语料,采用语料库驱动的计算文体学研究方法,量化统计分析英语新闻标题中词汇的词长、词汇密度和词频分布特征。发现英语新闻标题中的词汇多为平均词长 5 个字母左右的小词;标题中的词汇密度远高于普通英语,体现了新闻标题中措辞的多样性和丰富性;对词频的统计分析表明:在新闻标题中省略了那些在句子中起语法作用的功能词和虚词,使标题在结构上更简洁、对称、易于排版;标题中没有第一和第二人称代词,这是由新闻的本质特征决定的;新闻词汇具有时代特征;此外,新闻标题中高频使用“say”词,这是新闻标题特有的语言现象。通过量化分析以上词汇特征及其文体效果和成因,帮助读者提高标题浏览速度,更好地理解新闻标题。

[参考文献]

- [1] MARDH I. Headlines on the Grammar of English front-page headlines [M]. Lund: Gleerup, 1980.
- [2] BRISAU A. Complex sentence structure in headlines [J]. English Studies, 1969 (50): 31-38.
- [3] 端木义万. 英语报刊标题的功能及语言特色 [J]. 外语研究, 2001 (2): 46-50.
- [4] 刘路. 新闻标题论 [M]. 北京: 中国社会科学出版社, 2002.
- [5] IAROVICI E, RODICA A. The strategy of the headline [J]. Semiotica, 1989, 77 (4): 441-459.
- [6] REW L J. Editing for writers [M]. New Jersey: Prince Hall, 1998.

- [7] MCINTYRE B T. English news writing: a guide for journalists who use as a second language [M]. Hong Kong: Chinese University Press, 1996.
- [8] 陶雪城, 张萍. 基于自建语料库对 NBA 英语新闻标题的分析 [J]. 东南大学学报 (哲学社会科学版), 2010 (12): 165-172.
- [9] 刘政琳, 李定顺. 中国英语新闻标题中词汇分布特征研究 [J]. 现代妇女, 2014 (5): 278-279.
- [10] HALLIDAY M. The users and uses of language in J. Fishman [M]. The Hague: Mouton, 1968.
- [11] 张健. 新闻英语文体与范文评析 [M]. 上海: 上海外语教育出版社, 2004: 164-168.
- [12] 孙爱珍, 叶向平. 计算文体学模式构建 [J]. 外语电化教学, 2010 (6): 32-38.
- [13] MOSTELLER F, WALLACE D L. Applied bayesian and inference: the case of the federalist papers [J]. Berlin: Springer, 1963.
- [14] HORI, M. Investigating dickens' style: a collocational analysis [M]. New York: Palgrave Macmillan, 2004.
- [15] BELL A. The language of news media [M]. Oxford: Blackwell Publishers, 1991.
- [16] 罗建国. 英语新闻阅读与写作 [M]. 上海: 上海译文出版社, 1989: 93.
- [17] EVANS H. Newsman English [M]. London: Heinemann, 1972.
- [18] HOBSON FRED. Mencken: a life [M]. New York: Random House, 1994.
- [19] SCOTT M. Wordsmith tools version 4 [M]. Oxford: OUP, 2004.
- [20] STUBBS M. Lexical density: a technique and some findings [M] // M Coulthard. Talking about text, discourse analysis, monograph No. 13, English Language Research. Birmingham: University of Birmingham, 1986: 27-48.
- [21] JOHANSSON S. Some aspects of the vocabulary of learned and scientific English [M]. Gothenburg: Acta Universitatis Gothoburgensis, 1981.
- [22] SIMON-VANDENBERGEN A M. The grammar of headlines in the times 1870-1970 [M]. Bruselas: Paleis der academien-hertogsstraat, 1981.
- [23] QUIRK R, GREENBAUM S, LEECH G. A comprehensive grammar of the English language [M]. London: Longman World Publishing Corp, 1985: 347-349.

A Computational Stylistic Study of English News Headlines at Lexical Level

WENG Yu-xiu

(Chengyi College, Jimei University, Xiamen 361021, China)

Abstract: Words are the basic units of languages and the tool of expressing meaning and realizing communicative purposes. News English works as the medium of information dissemination, while news headline, the condensed and generalized news information, has its unique lexical feature. Based on self-built corpus and corpus-driven approach of computational stylistic analysis, the present study reports a quantitative analysis of the lexical features of news headlines on the level of word length, lexical density and word frequency, and arrives at the following conclusions: there are mainly midget words in news headlines, which has a mean word length of 5.13; the news headline has a STTR of 63.07, which indicates its great word diversity and vocabulary richness; the study of lexical frequency finds out that function words are often omitted in news headlines, there are no first and second personal nouns in news headlines, there is a high frequency of "say" word, and words in news headlines have the characteristics of the times. The study also analyzes the stylistic effects and causes of these lexical features.

Key words: English news headline; Computational Stylistics; word length; lexical density; word frequency

(责任编辑 陈蒙腰)