

口语能力描述语的语体分析: 基于中国英语能力等级量表的研究*

揭 薇¹ 金 艳²

摘要: 制定语言力量表需要使用描述语来描述语言能力水平,而描述语的语体分析研究十分鲜见。本研究采用语料库方法对中国英语能力等级量表的口语能力描述语进行语体分析。研究建立了口语能力描述语库,编程序对描述语进行查重筛选,考察了筛选所得描述语的词汇特征、高频词、词类分布和名词化等方面,并分析了 3 类口语描述语的高频词分布和 N 元序列特征。研究发现,英语口语能力描述语在词汇运用、词语搭配、语法结构选择等方面具有典型特征,这能为中国英语能力等级量表的口语描述语撰写、修订和翻译提供一定指导。

关键词: 中国英语能力等级量表;英语口语能力;口语描述语;语体特征

Abstract: One of the major tasks in developing language proficiency scales is to profile language users' level of proficiency with descriptors. Few empirical studies, however, have been conducted to investigate the register features of the descriptors frequently used in language proficiency scales. In this study, a corpus-based approach was adopted to analyze the register features of oral proficiency descriptors in the China's Standards of English. A descriptor pool was constructed, and statistical programs were used for duplicate screening. The remaining descriptors were examined in terms of lexical features, high frequency words, distribution of word types by part-of-speech, and nominalization. The high frequency words and the N-gram structure of descriptors from three different categories were further analyzed. The findings of the study revealed that the descriptors have some typical features in use of words, collocation patterns and grammatical structures, which will provide useful guidance for the writing, editing and translation of the speaking scale descriptors in the China's Standards of English.

Key words: China's Standards of English; oral English proficiency; descriptor of oral proficiency; register feature

中图分类号: H319 文献标识码: A 文章编号: 1004-5112(2017)02-0020-09

1. 研究背景

中国英语能力等级量表(China's Standards of English, CSE)的制定以交际语言能力模型为基础,采用面向运用的方法(use-oriented)对语言能力进行科学的描述或评价,通过明确、直观的描述语对语言能力等级进行“能做”(can-do)描述。量表先对语言能力进行综合性描述,然后根据语言学习者和使用者能力水平高低与社会需求,详略得当地分技能描述各种能力,即听、说、读、写、译及其语言行为(刘建达 2015)。根据教育部的工作计划,CSE 制定须使用采样法、文献法和专家法广泛收集描述语,并对收录的描述语进行整理、修改和审核,以符合描述语规范(教育部考试中心 2014)。描述语是量表的主体,因此开发最具典型性和代表性的描述语、确保每一条描述语的质量,是量表研制工作之本。

* 本文得到 2015 年教育部哲学社会科学研究重大课题攻关项目“中国英语能力等级量表建设研究”(编号 15JZD049)的资助,口语量表建设为该项目的子项目。作者感谢项目组的其他成员:邹绍艳、张晓艺、王隽、王华、张琳、刘黎岗、高森、程欣、周燕和严华。

1.1 语言能力描述语

描述语是提取、确认、描述信息或事物典型特征的语汇,在课程要求、课程标准、评分量表和能力量表中指以句子或段落形式出现的语言能力的具体表述。我国学者对语言能力描述语的理解和应用多基于交际语言能力的相关理论和实践研究(Hymes 1972; Canale & Swain 1980; Bachman & Palmer 1996; North 2000)。杨惠中等(2011, 2012)指出,语言能力描述语是语言测试和语言教学工作中对语言能力进行描述的基本操作工具,是语言测试和语言教学过程中用来描述语言能力的基本规范用语。王信旻(2013)认为,描述语用于描述现实生活中不同水平等级学习者能够完成的语言任务。目前,国内外许多语言力量表都采用“能做”描述语(can-do descriptor)描述语言能力,“能做”描述语是描述各水平等级语言能力的各方面所达到水平的语句(方绪军等 2011)。比如,《欧洲语言共同参考框架》(以下简称《欧框》)(Council of Europe 2001)基于交际语言能力理论,采用“能做”描述语对交际语言能力进行分层次、分技能、分等级描述,为评估和认证语言能力提供了指导性框架,对于课程设计、教学实践和评估具有重要的现实意义,从而成为现代欧洲乃至世界其他地区语言教育和语言政策的重要参考框架。

基于 CSE 项目的总体目标,口语量表的构成要素包括分类描述框架、级别和“能做”描述语。North (2000: 29)将量表的分类描述框架比作坐标的 X 轴,级别比作 Y 轴,两者将“能做”描述语组织起来形成多个分量表。构建量表时,分类描述框架需要根据语言能力理论确定,但 North(2000: 32)也指出描述语言能力的理论模型有别于实际使用的操作模型。力量表的分类描述框架应源于理论但简明且易于操作,框架中每个类别的特征都应有助于使用者更准确地理解其中“能做”描述语的含义。基于交际语言能力理论中的口语能力定义,口语量表的操作性分类框架将描述语分为口头交际活动、口头交际策略和口语文本特征 3 类(教育部考试中心 2014)。

1.2 语体特征

语体(register)是修辞学、语篇分析、语料库语言学等领域的主要研究对象。对于语体的界定,学界尚存在争议(Lee 2001; 刘大为 2013)。唐松波(1961)认为语体是在社会发展过程中,人们在不同活动领域运用语言的特点形成的体系。王德春(1987)认为由于社会生活的复杂性,人类在不同社会活动领域的交际因不同交际环境形成了一系列语言运用特点,即为语体。研究语体必然会涉及语体特征,即一种语体区别于其他语体的显著特征,如词汇、语法结构、语音特征等。语体研究较多关注不同语体之间的差异,如书面语体和口语语体的差异。文秋芳(2009)指出,学习者的英语语体特征是指口语语体与书面语体中出现频次存在显著差异的语言现象。

一种语体的特征取决于文本对某些表达方式的偏重程度。语法结构、词汇或词汇的不同义项在各语体中的分布都有差异(雷秀云、杨惠中 2001)。在语料库语言学领域,语体特征分析涉及语言结构的使用程度,因此语体特征研究涵盖词汇频率、词汇分布、语法结构、句法结构等多个层面(Biber & Conrad 2009: 56)。语体由语言项目在语篇中的使用频率不一引发,因此语料库方法可用于对某一语体大量文本的众多语言特征进行提取和统计分析,并加以量化描述。Biber(1988, 1996)提出多特征/多维度分析法(Multi-Feature/Dimensional Analysis, MF/MD),使用 67 个语言特征,运用统计分析法对英语口语和笔语的语体特征进行对比分析,系统揭示了口语和笔语之间的语体差异。雷秀云、杨惠中(2001)应用 MF/MD 方法探究了学术英语语体。van Rooy(2008)揭示了南非英语学习者的书面语体特征。文秋芳(2009)选择了 14 个典型语体特征,归纳口语语体与笔语语体特征群,分析中国英语专业学生口语语体特征的变化,考察其差异变化方向与英语本族语者是否一致。潘璠(2012)对中国非英语专业本科生和研究生书面语体进行了研究,证实了中国学习者的写作存在口语化倾向。

到目前为止,从语体特征角度分析语言能力描述语的研究尚未出现。不同于传统的围绕修辞和语法结构的语体分析,或者语体对比和区分研究,描述语的语体分析更注重讨论描述语的表述特点,其目的之一是为语言教师、测试工作者或量表开发者提供描述语撰写指导,目的之二是为描述语修改、翻译等实践活动提供依据。本研究基于中国英语能力等级量表的口语能力描述语库,运用语料库分析技术呈现描述语的语体特征,尝试探寻描述语这一特定语体与各语言要素之间的对应关系。

2. 研究设计

2.1 研究问题

本研究主要回答以下问题:

- (1) 英语口语能力描述语总体上具有哪些语体特征?
- (2) 英语口语能力的各分类描述语具有哪些语体特征?

2.2 研究工具

本研究使用的工具包括下列软件:Twins File Merger,合并文件生成大文件以便分析;Text Converter,将文本文件转成纯文本文件;Stata 13.1,编制程序对描述语进行查重处理;AntConc 3.4.3w,研究描述语的词频、索引(concordance)和搭配;Stanford POS Tagger,对描述语进行词性标注。

2.3 研究步骤

首先,建立用于研究的描述语库。建库要求描述语达到一定数量规模和质量,即有足够较好描述语言交际活动典型特征的描述语。描述语库的用途有其专门性,因此建立描述语库需要遵循一定原则。

在收集描述语阶段,需要考虑描述语库的代表性原则,即收集的描述语需最大限度地涵盖相关能力构念的各个方面,最大可能地代表描述语库所要反映的总体能力。本研究选取国内外 118 份文献资料,包括 44 份能力量表、20 份评分量表、42 份课程标准与教学大纲、12 份考试大纲。项目组成员将文献资料录入汇总,使用 Twins File Merger 软件合成描述语,并用 Text Converter 进行文本净化处理,将文件转成纯文本文件以便进一步分析。

在处理描述语语料阶段,需要考虑描述语库的结构合理性原则,即根据描述语库的建库目的,合理配置库容。项目组成员录入 118 份文献资料时,不可避免地会碰到来源相似、内容重复的描述语。如果不剔除这些描述语,语料统计分析结果可能会出现误差,这就需要对汇总的全样本语料进行必要的查重和整理。然而,人工查重费时又耗力,因此准确、便捷的描述语查重方法极为重要。项目组应用 Stata 13.1 编制程序检索剔除重复、相似的描述语,以实现非概率抽样。Stata 软件能够满足多种研究需求,如指示描述语储存位置、数据来源等,并能通过正则表达式循环匹配描述语,方便研究者快速、准确地找到完全或部分重复的描述语。

语体特征指标的选择主要依据 3 方面标准:(1) 已有语体特征研究中的相关指标,如类符/形符比(type token ratio, TTR)、平均词长(average word length)、平均句长(average sentence length)等反映语料基本特征的指标;(2) 能够通过语料库、词性标注等成熟的分析工具快速准确地提取和统计,使文本特征得以清晰显现;(3) 特征必须具有典型性。语体特征分析具体涵盖 4 个方面:(1) 词汇特征分析,涉及语料的基本倾向特征指标,如 TTR、词汇密度(lexical density)、平均词长、平均句长、文本可读性(readability)等;(2) 高频词分析,统计描述语中出现频率明显偏高的词并归纳其语言特征;(3) 词类分析,分析描述语的词类分布;(4) 名词化分析,考察描述语的名词化程度。

根据口语量表操作性分类框架,描述语库包含口头交际活动(1 805 条)、口头交际策略(1 221 条)和口语文本特征(402 条)3 个子库,便于进一步了解各分类描述语的语体特征。

3. 研究结果与分析

3.1 描述语查重和筛选

描述语 3 要素模型认为每条描述语应包含 3 个要素:(1) 表现(performance),即语言行为本身;(2) 标准(criteria),即语言表现的内在质量;(3) 条件(conditions),即可以支撑语言表现的其他外部条件(Pearson Standards and Quality Office 2014)。

描述语的筛选主要依据上述 3 个要素。以描述语言行为的“express”的筛选为例,在初始描述语库中检索(一级筛选)显示,含有动词“express”的描述语有 163 条。然后,检索描述语标准要素“fluent/fluently”(二级筛选),筛选出描述语 18 条。为了提高筛选精确度,再增加一个条件要素“spontaneous/spontaneously”(三级

筛选) ,筛选出描述语共 10 条(见图 1) 。经过核查 ,可以初步判定语料编号 143、145、158 等 10 条描述语为相似语料。

A	
143.	Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meani..
145.	"Can express him/herself fluently and spontaneously
158.	I can express myself fluently and spontaneously without much obvious searching for expressions.
177.	Can express him/herself fluently and spontaneously, almost effortlessly. Only a conceptually difficult subj..
205.	Can express him/herself fluently and spontaneously, almost effortlessly.
349.	Can express him/herself fluently and spontaneously, almost effortlessly. Only a conceptually difficult su..
389.	Can express him/herself fluently and spontaneously without much obvious searching for expressions.
397.	Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meani..
911.	By the end of the advanced stage, the candidate can:express themselves spontaneously, very fluently and pr..
1258.	By the end of the advanced stage, the candidate can:express themselves spontaneously, very fluently and pr..

图 1 基于描述语要素的查重筛选结果示例

经过查重和筛选处理 ,描述语库的容量由 4 263 条描述语减至 3 428 条 ,共 19 542 词次。这种基于抽样理论的描述语样本处理方法能够达到覆盖全面、反映整体、降低样本权重偏差等建库目的 ,尤其能够避免后续语料库分析过程中由于部分样本描述语语篇冗余、重复造成的词频分析和权重误差。

3.2 描述语总体语体特征

本研究对口语能力描述语的总体语体特征主要从描述语词汇特征、高频词、词类分布、名词化程度 4 个方面开展分析。

3.2.1 词汇特征

词汇特征通常可用以下标准测量并判断: (1) 文本的 TTR , (2) 平均词长 , (3) 10 个字母以上长词的使用比例 , (4) 词汇密度。为了解口语描述语的显著词汇特征 ,我们将其与其他类型语料的词汇特征进行比较 ,具体选择了张萍(2007) 对英语本族语者大学程度议论文体语料库(Written Samples from British National Corpus , WBNC) 的研究结果进行词汇特征比较。这样选择的主要原因是张萍(2007) 的研究语料提取自英语国家语料库(British National Corpus ,BNC) ,其中笔语约 9 千万词 ,是较具代表性的英语母语语料库 ,适合作为参照语料库。另外 ,描述语的开发者 and 编写者多为具有一定教学经验的教师与研究者 ,所提取的笔语子库在文本水平上与本研究的描述语更为接近 ,而且语料已经过净化处理 ,在词汇特征指标的选择上也和本研究相似 ,便于对比。

TTR 是量化词汇密度、衡量文本难度和词汇复杂度的重要指标 ,能在一定程度上反映文本词语的丰富度和变化度。TTR 比值越大 ,文本用词越复杂 ,语体书面化程度越高。对比结果显示 ,描述语库的 TTR 为 10.48 ,WBNC 语料库的 TTR 为 11.21 ,两者接近 ,说明文本的书面化程度均较高。描述语库的平均词长为 6.00 ,WBNC 为 4.91 ,相差 1.09。词长越长 ,文本越复杂 ,说明描述语和本族语议论文体的复杂性有较明显差异。描述语库中 10 个字母以上的长词占比 39% ,WBNC 为 41% ,描述语的长词使用略少于本族语议论文 ,但两者的差异相对较小。

描述语的句长、词汇密度和可读性也得到了分析 ,这些特征在张萍(2007) 的研究中未得到解读。描述语的平均句长为 15.95 ,比一般简易文本的句子要长 ,表明描述语句子具有一定复杂度。描述语库的词汇密度为 16.5% ,在一定程度上反映了描述语的用词多样性。描述语的文本可读性指标(Flesch-Kincaid Grade Level) (邢富坤 程东元 2010) 为 11.7 ,说明描述语的文本相当于美国高三阅读水平。

3.2.2 高频词

频率是语料库分析方法中重要的研究手段。语料库中的高频词多数为功能词。Sinclair(1991: 31) 提出 ,语料库中高频词的分布较为稳定 ,因此频率排序的任何显著变化都有可能具有重要意义。从描述语前 20 位高频实词的统计结果(见表 1) 来看 ,高频词大部分与口语能力的评价标准有关 ,如“range appropriate ,vocabulary”;有些高频词和口语活动密切相关 ,如“discussion ,conversation ,speech”。

表 1 口语描述语高频词及其覆盖面(前 20 位实词)

排序	实词	频次	比例(%)	排序	实词	频次	比例(%)
(1)	language	210	1.08%	(11)	conversation	77	0.39%
(2)	range	149	0.77%		skill		
(3)	use	131	0.67%	(13)	speech	74	0.38%
(4)	vocabulary	104	0.53%	(14)	give	70	0.36%
(5)	idea	88	0.45%	(15)	communicative	69	0.35%
(6)	express	85	0.44%	(16)	discourse	67	0.34%
(7)	topic	83	0.43%	(17)	information	62	0.32%
	appropriate			(18)	examiner	61	0.31%
	discussion			(19)	make	58	0.30%
	complex				wide		

口语描述语库中“can-do”动词词组和“can-do”四元序列的频次排序(前 14 位) 结果显示, “can-do”动词词组和四元序列体现了口语活动中的产出性(productive)、互动性(interactive) 等表达和交际特点, 具体如表 2 所示。

表 2 口语描述语“can-do”动词词组和四元序列

排序	动词词组	频次	排序	四元序列	频次
(1)	can use	58	(1)	can express himself/herself	11
(2)	can express	41	(2)	can use the correct	6
(3)	can give	22	(3)	can initiate , maintain and	5
(4)	can make	13		can use language flexibly	
(5)	can ask	12		can use vocabulary specific	
(6)	can take	11	(6)	can express his/her	4
(7)	can follow	10		can interact with a	
	can keep			can select a suitable	
(9)	can describe	9	(9)	can contribute effectively to	3
	can initiate			can give a clear	
	can produce			can keep up a	
	can understand			can take an active	
(13)	can communicate	7		can use the language	
(14)	can explain	6	(14)	can adjust what he	2
	can participate			can ask questions and	
	can handle			can carry out a	

3.2.3 词类分布

本研究对描述语的9种词类(名词、动词、形容词、副词、连词、介词、代词、限定词和“to”)进行标记,并统计了各种词类的数量。表3的统计结果表明,描述语中名词、动词和形容词等实词的使用数量显著高于其他词类(除介词外)。对名词的进一步考察显示,名词中的动词派生名词较多,如“contribution(44次),presentation(40次),description(20次)”。动词派生名词是动词名词化的一种形式,在语义上与动词存在密切的联系,常常表示动作本身或动作的结果。在描述语的形容词词类中,表示程度的形容词较多,如“familiar(52次),simple(44次),general(17次)”;从语义功能来看,表示比较和评价的形容词较多。值得注意的是,描述语中使用频率最高的虚词是介词,所占比例达到13.47%,其中位居前列的高频介词有“of(757次),in(386次),with(240次)”等。在高频介词的搭配中,“名词+of”组成的短语使用频率很高,如“ways of,ideas of, effects of”等;与“in”搭配的高频短语包括“in discussion,in communication,in English”等;与“with”高频搭配而成的有“with very little hesitation,with some confidence,with a degree of”等表示程度的状语。此外,不论“to”用作介词或是不定式标记,Stanford POS Tagger均将其标记为TO。TO的出现频率也较高(566次),其中动词不定式“to+动词”的结构占大多数,如“to describe,to express,to cooperate with,to explain”。

表3 口语描述语词类分布

词类	标记	词数	比例(%)
名词	NN/NNS/NP/NPS	6059	31.00%
动词	VB/VBN/VBG/VBD/VBP/VBZ	3105	15.89%
形容词	JJ/JJR/JJS	2364	12.10%
副词	RB/RBR/RBS/WRB	781	4.02%
连词	CC	1340	6.86%
介词	IN	2633	13.47%
代词	PP/PP \$ /WP/WP \$	418	2.14%
限定词	DT/WDT/PDT	1591	8.14%
to	TO	566	2.90%

3.2.4 名词化程度

名词化与“定义”、概念的“复指”关联密切,用来实现句中各种事件逻辑关系的内在化。名词化结构在各类语篇中的不同分布说明名词化频率与语体存在密切的联系。杨信彰(2006)认为,名词化是正式语体的一大特征,因为这些语体涉及抽象的概念。朱永生(2006)提出,名词化具有浓缩信息、增加小句信息量的功能。因此,名词化在语类区分中扮演重要角色,语篇的技术性越强,名词化的出现频率越高。

口语描述语含有大量名词或名词词组。描述语中的常用抽象名词化后缀包括“-tion,-sion,-ity,-ment,-ness”等,如“specialization,discussion”。对这些后缀的检索结果表明,描述语中“-tion”(4.80%)显著多于其他名词化后缀(见表4),而且比例也高于WBNC语料库中的“-tion”后缀形式(张萍2007)。描述语中的动词名词化表示具体动作的抽象引申,这样的结构能使描述语用词简洁、表达确切、结构严密,并增加描述语词句承载的信息容量。

表4 口语描述语名词化程度比较

后缀	描述语库(总数)	描述语库(标准化后)	WBNC总数(张萍2007)	WBNC(标准化后)
-tion	938	4.80%	16001	1.59%
-sion	250	1.28%	16000	1.59%
-ment	187	0.96%	16000	1.59%
-ity	111	0.57%	16000	1.59%
-ness	20	0.10%	7871	0.07%

3.3 分类描述语的语体特征

3.3.1 分类描述语库

描述语分为口头交际活动、口头交际策略和口语文本特征 3 类,各类描述语中的高频词及其比例如表 5 所示。这些高频词与描述语分类高度相关,并非杂乱无序,即不同分类下的描述语高频词存在较为明显的类别差异。对描述语高频词的归类统计能够突显各个类别描述语的主要特征,构建描述语高频词表。口头交际活动描述语中的高频词以典型的口语表达动词为主;口语交际策略描述语中的高频词以动词、名词和形容词为主,是执行口语任务的常用动作,如“use, convey, deliver”;口语文本特征描述语中的高频词则以形容词和名词为主,描述口语语言表现的质量和程度。

表 5 口语分类描述语中的高频词分布

口头交际活动			口头交际策略			口语文本特征		
实词	频次	比例(%)	实词	频次	比例(%)	实词	频次	比例(%)
deliver	55	0.79%	communication	52	1.14%	language	164	2.04%
make	51	0.74%	use	52	1.14%	range	127	1.58%
use	50	0.72%	audience	48	1.05%	vocabulary	96	1.19%
language	50	0.72%	convey	43	0.95%	appropriate	66	0.82%
express	47	0.68%	deliver	40	0.88%	topic	63	0.78%
describe	40	0.58%	conversation	37	0.81%	idea	61	0.76%
explain	37	0.53%	appropriate	36	0.79%	complex	57	0.71%
present	36	0.52%	background	34	0.75%	discourse	54	0.67%

3.3.2 分类描述语的 N 元序列特征

N 元序列(N-gram)是指含有一定数量单词的多词序列或词块,是语体中高频出现的词汇序列。N 元序列研究有助于探讨语体的语法结构、语篇功能和特定文体特征等(Biber & Conrad 1999; Carter & McCarthy 2006; Hyland 2008)。对描述语中 N 元序列的分析能够揭示描述语词语搭配、实义词块表达和特定语法结构选择等方面的典型特征。本研究使用语料库软件 AntConc 的 N 元序列提取功能,将最低频次设置为 5,分别对 3 类描述语中的高频多元词块(3—4 词)进行提取和标注,再对这些词块的结构、功能特点和语篇功能进行描述分析。限于篇幅,表 6 列举了与高频词搭配频率前两位的词块。

表 6 口语分类描述语中常用 N 元序列的频次

口头交际活动	频次	口头交际策略	频次	口语文本特征	频次
deliver well-organized	18	convey ideas clearly	15	range of language	11
deliver well-organized formal	18	convey ideas clearly and	15	stretches of language	11
make a presentation	5	of oral communication	15	range of language	11
make a presentation about	5	to oral communication	12	range of appropriate vocabulary	10
use a range	5	interests of the audience	7	appropriate vocabulary to	9
use a range of	5	of the audience	7	appropriate vocabulary to give	9

统计分析结果表明,不同分类描述语的词块具有各自的结构特点和语用功能。首先,各类描述语表达主

题意义的实义词块特点鲜明。比如,口头交际活动描述语中的“deliver well-organized ,make a presentation”等词块能够描述口语表达的具体动作和不同功能类型的任务;口头交际策略描述语中的“of oral communication , interests of the audience”强调交际条件和为改善交流效果采用的具体交际策略;口语文本特征描述语中的“range of language appropriate vocabulary to give”等具有明显的描述语言能力、表达立场的词块特点,表明说话人对对话内容的态度或评价。

其次,各类描述语在语法结构上的主要特点为词块种类较少,限于高频常用动词词块、“名词+介词短语”词块和含有同位语或定语从句的词块类型。词块的复现率较高,词块使用多样性不足,比如口语文本特征描述语频繁使用名词词块、介词词块来表达指示功能和组织文本。

最后,描述语中介词、连词、冠词等虚词和简单名词构成的功能性词块出现频率较低,如“at the same time ,one of the most”等常用书面词块使用较少,从而有别于其他书面语体。再者,描述语中的多词序列主要由表达实际意义的动词、名词、副词、形容词等构成,如“deliver well-organized ,range of appropriate vocabulary , make a presentation”。

综上所述,口语描述语的重点在于描述学习者最典型的口语能力与行为。词块结构和功能分析有助于进一步了解不同类别口语描述语的语体特征。口头交际活动描述语侧重描述学习者的语言活动或任务,强调学习者能做什么,做得怎样。口头交际策略描述语侧重描述学习者在口语活动中为达到交际目的采用的方法。口语文本特征描述语通常描述能够观察和测量的特征,在评价口语能力时具有较强的可操作性。

4. 结语

本研究基于中国英语能力等级量表的口语描述语库,运用语料库分析方法探讨描述语的语体特征。研究结果表明,描述语在词汇特征、高频词、词类分布、名词化程度等方面有其独特性,不同分类描述语的词块具有明显的结构特点和功能特点。在此基础上,研究提出编制描述语须做到以下几点:首先,描述语建库需要考虑描述语的代表性,选取合适的样本,并根据建库目的,确保结构合理。其次,描述语撰写、修订、审定和翻译需要遵循一定的语言规范。最后,描述语撰写和修改过程中应有针对性地增强研发者的语体意识,使其了解、注重描述语的语言特点和规范。

语言量表制定是一项艰巨的任务,须运用科学的方法来提高工作效率和质量。本研究对语言量表研发具有一定实践借鉴意义。以往的描述语研究虽已意识到描述语的复杂性,但是很少采用语料库分析方法考察描述语的语体特征。大部分研究只是通过文献法汇总整理描述语,再借助专家判断进行修订处理,这种方法在处理数量庞大的描述语时会出现偏差。描述语的语料库分析方法能够帮助研究者编制特点鲜明、认同度较高的描述语,提高描述语的质量,为量表研究提供充分的解释力和支撑力。

参 考 文 献

- [1] Bachman L F & Palmer A S. *Language Testing in Practice: Designing and Developing Useful Language Tests* [M]. Oxford: Oxford University Press , 1996.
- [2] Biber D. *Variation across Speech and Writing* [M]. Cambridge: Cambridge University Press , 1988.
- [3] Biber D. Investigating language use through corpus-based analyses of association patterns [J]. *International Journal of Corpus Linguistics* , 1996 , 1(2) : 171-197.
- [4] Biber D & Conrad S. The frequency and use of lexical bundles in conversation and academic prose [A]. In Hasselgard H & Oksefjell S (eds.) . *Out of Corpora: Studies in Honor of Stig Johansson* [C]. Amsterdam: Rodopi , 1999. 181-190.
- [5] Biber D & Conrad S. *Register , Genre , and Style* [M]. Cambridge: Cambridge University Press , 2009.
- [6] Canale M & Swain M. Theoretical bases of communicative approaches to second language teaching and testing [J]. *Applied Linguistics* , 1980 , 1(1) : 1-47.
- [7] Carter R A & McCarthy M J. *Cambridge Grammar of English* [M]. Cambridge: Cambridge University Press ,

2006.

- [8] Council of Europe. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* [M]. Cambridge: Cambridge University Press, 2001.
- [9] Hyland K. *Metadiscourse* [M]. Beijing: Foreign Language Teaching and Research Press, 2008.
- [10] Hymes D H. On communicative competence [A]. In Pride J B & Holmes J (eds.). *Sociolinguistics* [C]. Harmondsworth: Penguin, 1972. 269-293.
- [11] Lee D Y W. Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the BNC jungle [J]. *Language Learning & Technology*, 2001, 5(3): 37-72.
- [12] North B. The development of a common framework scale of language proficiency [J]. *The Modern Language Journal*, 2000, 86(3): 482-483.
- [13] Pearson Standards and Quality Office. *Writing Descriptors: Guidelines and Best Practice* [Z]. London: Pearson Publishing Ltd, 2014.
- [14] Sinclair J. *Corpus, Concordance, Collocation* [M]. Oxford: Oxford University Press, 1991.
- [15] van Rooy B. A multi-dimensional analysis of student writing in Black South African English [J]. *English Worldwide*, 2008, 29(3): 268-305.
- [16] 方绪军 杨惠中 朱正才.语言能力“能做”描述的原理与方案:以 CEFR 为例[J].世界汉语教学, 2011, (2): 246-257.
- [17] 教育部考试中心.中国英语能力等级量表口语组工作手册[Z].北京:教育部考试中心, 2014.
- [18] 雷秀云 杨惠中.基于语料库的研究方法及 MD/MF 模型与学术英语语体研究[J].当代语言学, 2001, (2): 143-151.
- [19] 刘大为.论语体与语体变量[J].当代修辞学, 2013, (3): 1-22.
- [20] 刘建达.我国英语能力等级量表研制的基本思路[J].中国考试, 2015, (1): 7-11, 15.
- [21] 潘璠.中国非英语专业本科生和研究生书面语体的多特征多维度调查[J].外语教学与研究, 2012, (2): 220-232.
- [22] 唐松波.谈现代双语的语体[J].中国语文, 1961, (5): 15-19.
- [23] 王德春.语体略论[M].福州:福建教育出版社, 1987.
- [24] 王佑旻.汉语能力标准的描述语任务难度研究——以中级口语能力量表为例[J].世界汉语教学, 2013, (3): 413-423.
- [25] 文秋芳.学习者英语语体特征变化的研究[J].外国语, 2009, (4): 2-10.
- [26] 邢富坤 程东元.基于统计语言模型的英语易读性研究[J].解放军外国语学院学报, 2010, (6): 19-24.
- [27] 杨惠中 朱正才 方绪军.英语口语能力描述语因子分析及能力等级划分——制定语言能力等级量表实证研究[J].现代外语, 2011, (2): 151-161.
- [28] 杨惠中 朱正才 方绪军.中国语言能力等级共同量表研究——理论、方法与实证研究[M].上海:上海外语教育出版社, 2012.
- [29] 杨信彰.名词化在语体中的作用——基于小型语料库的一项分析[J].外语电化教学, 2006, (3): 3-7.
- [30] 张萍.不同二语学习者词汇复杂度的语料库对比研究[J].中国外语, 2007, (3): 54-59.
- [31] 朱永生.名词化、动词化与语法隐喻[J].外语教学与研究, 2006, (2): 83-90.

作者单位: 1. 上海交通大学外国语学院, 上海 200240; 上海对外经贸大学国际商务外语学院, 上海 201620
2. 上海交通大学外国语学院, 上海 200240