

# 口述历史语篇的多维度分析

赵雪<sup>1</sup>, 鲁瑾芳<sup>2</sup>, 官春辉<sup>3</sup>, 张颜雪<sup>4</sup>

(1. 中国传媒大学人文学院, 北京 100024; 2. 绍兴市中等专业学校, 浙江 绍兴 312000; 3. 山东政法学院传媒学院, 山东 济南 250014; 4. 掌阅科技, 北京 100024)

**摘要:** 在与日常谈话、学术论文语篇对比的基础上, 对口述历史语篇的多维度分析表明, 可以提取的基本维度有四个, 即“人际互动与资讯传递”“个人情感的表达与对事实的客观陈述”“口语体与书面语体”“事件重构与论证阐释”。研究表明, 口述历史语篇倾向于人际互动; 它重在对事实的客观陈述; 它是准备程度较高、经过编辑加工的言语成品, 具有书面语体的倾向; 它重在重构历史事件, 具有叙事性。上述特征是由口述历史语篇的交际方式、交际目的以及生成环境决定的。

**关键词:** 口述历史语篇; 多维度分析; 人际互动性; 客观陈述; 书面语体倾向; 叙事性

**中图分类号:** H1      **文献标识码:** A      **文章编号:** 1001-5124(2019)04-0060-11

## 一、引言

现代意义上的“口述历史”(Oral History)发端于西方,“口述历史”这个术语诞生于20世纪40年代,指“以录音访谈(interview)的方式搜集口传记忆以及具有历史意义的个人观点”。<sup>[1]</sup>

20世纪50年代以来,具有跨学科性质的口述历史已广泛应用于社会学、人类学、民俗学、建筑学、文学等领域。西方语言学界在20世纪70年代,就开始了对口述历史语言的研究。我国关于口述历史语言的研究起步晚、成果少,目前仅有《口述史、民族志、新闻访谈话题的对比——基于词共现网络的分析》等为数不多的论文。<sup>[2]</sup>

语篇是指任何长度的、语义完整的口语或书面语的段落。<sup>[3]</sup> 本文将一次访谈、一场日常谈话以及一篇学术论文,均视为一个完整的语篇。

虽然,口述历史语篇和日常谈话语篇一样,都采用了面对面的交流方式,交际双方都共享时空,但是,它们在情景语境上还是有区别的。口述历史语篇是事先有准备、事后编辑整理过的语篇;话题为受访者亲历、亲闻的历史事件。而日常谈话语篇大多是无备的,话题

相对随意、不固定。那么,口述历史语篇在语体上有什么特点呢?这是一个值得探讨的问题。

## 二、语料来源、处理及研究方法

本文语料共计190篇,大约39万字,包括以下三个部分:一是抗战老兵、电影人、报人的口述历史语料;二是日常谈话语料;三是学术论文语料。其中,口述历史语料62篇,大约8万字,分别来自中国传媒大学崔永元口述历史研究中心、凤凰网等;<sup>①</sup>日常谈话语料64篇,大约8万字,由我们转写、标注;学术论文语料64篇,大约23万字,均为从“中国知网”随机下载的近几年的期刊论文,内容涉及语言学、文学、法律、大众传媒等领域。我们之所以将日常谈话、学术论文语篇作为对比语料,是因为日常谈话和学术论文分别是口语体和书面语体的典型代表。我们以这两类语体作为参照,来考察口述历史语篇的语体特征。

我们采用“中国传媒大学在线分词标注系统”,按照粗颗粒度对口述历史语料分词标注,并对标注的结果进行了人工校对,自建词类语料库。该语料库共有46731个词;参照哈尔滨工业大学语言技术平台的依存句法分析,对口

收稿日期:2019-05-03

基金项目:国家社科基金项目“口述历史语篇的计量语体研究”(17BY213)

第一作者简介:赵雪(1960-),女,四川成都人,教授,博士,主要研究方向:应用语言学。Email: zhaoxue1221@aliyun.com

述历史语料中的句子进行层次分析，并人工标注了短语的类型，<sup>②</sup>自建短语语料库，该语料库共有 5141 个短语；采用 Python 程序包<sup>③</sup>对口述历史语料进行分句，并人工标注了句型、句类和句式，自建句子语料库，该语料库共有 2579 个句子。<sup>④</sup>

我们以同样的方法对日常谈话和学术论文语料进行分词，标注短语和句子，自建日常谈话和学术论文的词类、短语和句子的对比语料库。其中，日常谈话语料共计 47002 个词、12389 个短语、6083 个句子；学术论文语料共计 124891 个词、14304 个短语、4132 个句子。

本文采用多维度分析 (Multidimensional Analysis) 的方法对口述历史语篇进行研究。多维度分析是在语料库的基础上，采用定量和统计的方法，根据不同的维度，来对比不同语体之间的差异。该方法是由美国学者 Biber 在前人研究的基础上提出来的。<sup>[3-4]</sup> 1988 年，Biber 曾在 *Variation across speech and writing* 一书中介绍过多维度分析法，并以此对英语口语体和书面语体进行了分析。他通过对口语体和书面语体的多维度分析，得出“交互性与信息性表达”“叙述性与非叙述性关切”“所指明确与所指有赖情景”“显性劝说型表达”“信息抽象与非抽象”“即席信息详述”六个维度。Biber 还对高校校园口语体和书面语体做过多维度分析，<sup>[5]</sup> 得出“口语体与书面语体”“程式化的话语与以内容为中心的话语”“对事件的重构”“以教师为中心的立场”四个维度。目前，国外关于多维度的语体研究，成果颇丰，例如 Conrad 等；<sup>[6]</sup> 国内的相关研究主要集中在英语语体的分析上，如桂诗春等。<sup>[7]</sup> 此外，中国台湾的学者张学谦也曾对闽南语口语及书面语体做过多维度分析。<sup>[8]</sup> 他通过对闽南语口语语体与书面语体的多维度分析，得出“人际关系与表达资讯”“表达个人情感”“逻辑说服与时间的联接”“叙述”“具亲和力的说明与详细的报道”五个维度。

Biber 等学者的研究，<sup>[4-8]</sup> 为我们提供了思路和方法。本文将根据语料中的语言参项，<sup>⑤</sup> 确定口述历史语篇的维度，并将其与日常谈话语篇、学术论文语篇进行对比。

### 三、口述历史语篇维度的确定

在多维度分析时，我们首先要确定的是所要统计的语言参项，然后才能对语言参项进行分析。

#### (一) 语言参项的提取

在以往的研究成果中，尚无有关汉语的语言参项提取框架供我们参考。因此，根据对语料的观察，我们参考张斌关于词类、短语和句子的理论，<sup>[9]</sup> 提取语言参项。具体步骤如下：

第一步：筛选语言参项。根据标注结果，我们从口述历史语篇、日常谈话语篇和学术论文语篇中，筛选出语法层面的 56 个语言参项，并对这 56 个语言参项进行统计。

第二步：确定语言参项。由于多维度分析中的因子分析 (Factor Analysis)<sup>⑥</sup> 规定了样本数必须是变量 (即语言参项) 的 5 至 7 倍，因此最多可以提取的语言参项只有 38 个 (本文共计 190 篇文本)。为使因子旋转<sup>⑦</sup> 后获得较为理想的结果，我们删除了贡献率 (解释力) 较低的、表示相同语言功能的语言参项，最后确定了 29 个语言参项，将其作为因子分析的变量。

在提取了这 29 个语言参项后，我们采用统计软件 SPSS19.0 对每个语言参项的分布情况进行了统计 (表 1)。需要说明的是，本文仅提取了语法层面的语言参项，未涉及词汇、语音等层面的语言参项。

#### (二) 维度的确定

在多维度分析中，维度是需要借助因子分析来确定的。因子分析的核心步骤是对语言参项进行降维，目的是将隐藏了重要相关性的语言参项进行分类，从而减少所要分析的语言参项的数量。我们用统计软件 SPSS19.0，采用主成分分析法，<sup>⑧</sup> 抽取特征值大于 1 的成分，进行因子旋转后，最后将 29 个语言参项简化为六组共现的参项 (因子) (表 2)。

六个因子涵盖了大约 75.794% 的信息，具体来说，因子 1-4 涵盖了大约 64.263% 的信息；而因子 5 和 6 仅涵盖了大约 6.22% 和 5.31% 的信息。虽然碎石图<sup>⑨</sup> 显示 (图 1)，因子 4 和因子 5 的显著性相差并不大；然而，按照数据挖掘理论“所提取的因子应能概括总体信息的

表 1 语言参项

语言参项		口述历史语篇		日常谈话语篇		学术论文语篇	
		均值 <sup>①</sup>	标准差	均值	标准差	均值	标准差
词	1. 名词	26.99	5.8	20.44	5.53	33.15	4.14
	2. 普通名词	18.04	3.91	15.11	5.1	28.4	4.33
	3. 地名	1.52	1.44	0.51	0.83	0.39	0.39
	4. 时间名词	1.84	0.93	1.17	0.68	0.58	0.34
	5. 代词	11.56	3.45	15.66	3.19	3.78	1.57
	6. 第一人称代词	3.38	1.81	4.88	1.89	0.32	0.46
	7. 第二人称代词	1.35	1.57	3.12	1.7	0.07	0.2
	8. 第三人称代词	2.02	1.25	1.65	1.39	0.63	0.78
	9. 指示代词	3.56	1.38	3.81	1.56	1.44	0.62
	10. 介词	3.14	1.01	2	0.93	4.58	0.93
	11. 连词	1.89	0.94	2.18	1	4.38	0.85
	12. 语气词	3.54	1.5	5.8	2.06	0.14	0.32
	13. 叹词	0.17	0.34	1.14	1.01	0	0.02
	14. 象声词	0.05	0.17	0.2	0.37	0	0.01
短语	15. 主谓短语	18.46	4.18	21.67	3.77	12.54	2.73
	16. 定中短语	17.26	4.91	13.59	5.02	35.53	7.08
	17. 状中短语	25.6	4.96	29.74	4.89	16.92	4.88
	18. 单句平均层次数 <sup>②</sup>	397.58	79.41	244.65	46.19	595.39	89.43
	19. 复句平均层次数	233.13	51.98	133.63	22.87	211.56	43.51
句型	20. 单句	38.36	11.15	72.06	8.65	31.54	10.37
	21. 主谓句	25.31	7.64	34.14	8.22	27.59	9.55
	22. 形容词性谓语句	1.64	2.71	3.27	2.63	0.75	1.4
	23. 非主谓句	4.21	4.94	23.68	8.76	2.83	3.73
	24. 形容词性非主谓句	0.82	2	2.9	2.71	0.02	0.14
	25. 复句	60.95	11.35	23.93	9.65	68.46	10.37
句类	26. 陈述句	77.19	10.54	66.58	13.1	98.27	3.4
	27. 感叹句	3.56	5.46	6.38	8.8	0.34	1.6
	28. 疑问句	16.81	7.95	19.49	8.49	2.59	2.98
句式	29. 省略句	16.02	8.34	18.75	7.58	33.15	4.14

60%以上”的观点,<sup>[10]</sup>结合各因子所包含的语言参项,我们决定提取前四个因子作为多维度分析的基本维度。

四个因子(维度)中的因子负荷量<sup>②</sup>绝对值在 0.25 以上的语言参项代表所对应的因子(维度)(表 3)。我们根据各因子(维度)中语言参项的分布情况,以及在每个维度中共现的语言参项的最主要的功能,对维度进行命名。

#### 四、口述历史语篇维度的功能性解释及命名

语言参项的共现是多维度分析中的一个核心概念。多维度分析将语言参项不同的共现模式,视为不同的基本变异维度。在每个维度中,共现的每组语言参项都是以量化的方法确定的。我们根据语言参项的分布情况,运用语

体理论,结合情景语境,对每个维度中共现的语言参项进行功能性解释,并对维度进行命名。

##### (一) 维度 1: 人际互动与资讯传递

在维度 1 中共现的正向语言参项有:代词、疑问句、第一和第二人称代词、主谓短语、语气词、指示代词、状中短语、单句、非主谓句、省略句、时间名词、形容词性非主谓句、感叹句、第三人称代词、叹词。上述语言参项的共现与人际互动有关。例如,代词、疑问句、第一和第二人称代词,多共现于互动性的会话中。<sup>[5]7-8, 94</sup>

在维度 1 中共现的负向语言参项有:陈述句、定中短语、单句平均层次数、普通名词、连词、名词、介词、复句。上述语言参项多共

表 2 解释的总方差<sup>⑬</sup>

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	13.59	46.849	46.849	13.586	46.849	46.849	9.169	31.618	31.618
2	2.763	9.527	56.375	2.763	9.527	56.375	3.936	13.573	45.191
3	1.74	5.999	62.375	1.74	5.999	62.375	3.396	11.709	56.901
4	1.53	5.275	67.65	1.53	5.275	67.65	2.135	7.362	64.263
5	1.33	4.586	72.235	1.33	4.586	72.235	1.805	6.224	70.487
6	1.032	3.558	75.794	1.032	3.558	75.794	1.539	5.307	75.794
7	0.834	2.875	78.668						
8	0.718	2.477	81.145						
9	0.621	2.14	83.285						
10	0.567	1.954	85.239						
11	0.521	1.796	87.036						
12	0.497	1.714	88.749						
13	0.478	1.649	90.399						
14	0.443	1.529	91.927						
15	0.376	1.297	93.224						
16	0.349	1.203	94.427						
17	0.299	1.032	95.46						
18	0.261	0.9	96.36						
19	0.237	0.816	97.176						
20	0.222	0.764	97.94						
21	0.182	0.627	98.567						
22	0.12	0.412	98.979						
23	0.096	0.333	99.312						
24	0.078	0.27	99.582						
25	0.046	0.16	99.742						
26	0.033	0.114	99.856						
27	0.025	0.085	99.941						
28	0.015	0.05	99.992						
29	0.002	0.008	100						

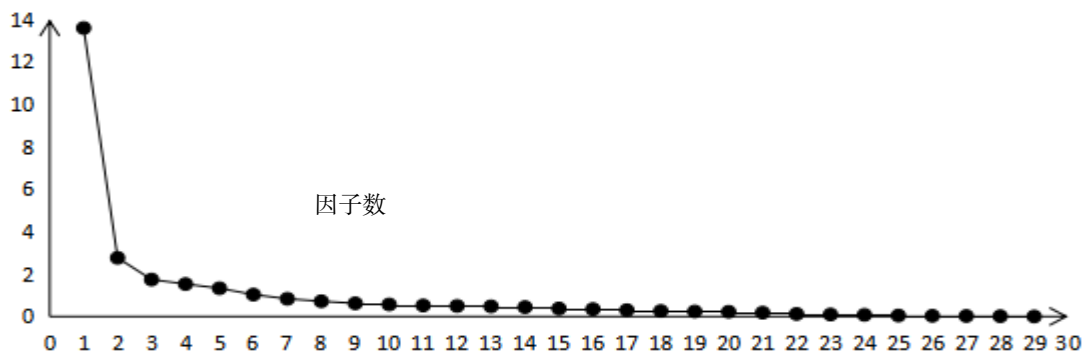


图 1 因子碎石图

现于信息型语篇中，用来传递较丰富的信息。  
例如，名词、陈述句、复句多共现于内容丰富

的资讯中。<sup>[5]116</sup>  
总之，从交际目的、参与者、生成环境、

交流渠道等情景因素来看, 维度 1 上正向与负向的语体差异可以概括为: 互动与非互动, 即时与有准备。换言之, 正向语言参项的共现, 表现出较强的人际互动性; 负向语言参项的共现, 则倾向于资讯的单向传递。因此, 我们将维度 1 命名为: 人际互动与资讯传递。

(二) 维度 2: 个人情感表达与对事实的客观陈述

在维度 2 中共现的正向语言参项有: 叹词、象声词、形容词性非主谓句、非主谓句、单句、语气词、省略句、状中短语。上述语言参项的共现与个人情感的表达有关。例如, 叹词、语

气词的共现, 多用于表达主观情感。<sup>[9][262]</sup> 在维度 2 中共现的负向语言参项有: 复句、单句平均层次数、名词、复句平均层次数、普通名词、陈述句。上述语言参项的共现与对复杂事件的陈述有关。例如, 陈述句、结构复杂的复句多用于叙述复杂事件。<sup>[11]</sup>

维度 2 上正向与负向的语体差异主要体现在是否能够表达发话者的情感。换言之, 正向语言参项的共现, 体现了对个人情感的表达; 负向语言参项的共现, 则用于对事实的客观陈述。因此, 我们将维度 2 命名为: 个人情感表达与对事实的客观陈述。

表 3 四个因子(维度)中的语言参项及其因子负荷量

	因子 1		因子 2		因子 3		因子 4	
	语言参项	因子负荷量	语言参项	因子负荷量	语言参项	因子负荷量	语言参项	因子负荷量
正向语言参项	代词	0.83						
	疑问句	0.82						
	第一人称代词	0.77						
	第二人称代词	0.77						
	主谓短语	0.77						
	语气词	0.73						
	指示代词	0.62						
	状中短语	0.52						
	单句	0.51*	叹词	0.81				
	非主谓句	0.47*	象声词	0.74				
	省略句	0.46*	形容词性非主谓句	0.68	主谓句	0.82		
	时间名词	0.33*	非主谓句	0.66	单句	0.69		
	形容词性非主谓句	0.29*	单句	0.44*	形容词性谓语句	0.58	地名	0.80
	感叹句	0.27	语气词	0.41*	非主谓句	0.35*	时间名词	0.63
第三人称代词	0.26	省略句	0.36*	主谓短语	0.28*	省略句	0.52	
叹词	0.25*	状中短语	0.35*	状中短语	0.25*	状中短语	0.29*	
负向语言参项	复句	-0.51*	陈述句	-0.27*	单句平均层次数	-0.25*	连词	-0.54*
	介词	-0.62	普通名词	-0.28*	时间名词	-0.33*		
	名词	-0.63	复句平均层次数	-0.35*	介词	-0.36*		
	连词	-0.64	名词	-0.38*	复句	-0.67		
	普通名词	-0.72	单句平均层次数	-0.44*	复句平均层次数	-0.72		
	单句平均层次数	-0.73	复句	-0.47*				
	定中短语	-0.76						
	陈述句	-0.77						

注: “\*” 为重复的语言参项, 这些语言参项往往与其他维度中的一些语言参项共享某些交际功能。由于其因子负荷量较低, 我们仅将其作为解释维度的参考特征, 而不计算其分值。

(三) 维度 3：口语体与书面语体

在维度 3 中共现的正向语言参项有：主谓句、单句、形容词性谓语句、非主谓句、主谓短语、状中短语。上述语言参项的共现与口语体有关。例如，单句、非主谓句等语言参项，多共现于口语体中。

维度 3 中共现的负向语言参项有复句平均层次数、复句、介词、时间名词、单句平均层次数。在与正向语言参项对比后，我们认为这些语言参项的共现，可能与书面语体有关。例如复句平均层次数、复句等语言参项，多共现于书面语体中。

维度 3 上正向与负向的语体差异主要表现为口语体和书面语体的差异。换言之，正向语言参项的共现具有较强的口语体倾向，负向语言参项的共现则具有较强的书面语体的倾向。因此，我们将维度 3 命名为：口语体与书面语体。

(四) 维度 4：事件重构与论证阐释

在维度 4 中共现的正向语言参项有：地名和时间名词、省略句、状中短语。上述语言参项的共现与事件重构有关。例如，地名、时间名词的共现，多用于交待与事件相关的时间、地点等。

在维度 4 中共现的负向语言参项只有连词。连词表示并列、承接、转折、因果、条件、假设、让步等逻辑关系，<sup>[9]197-203</sup> 多出现于书面语体中，用于连接、说明事件对象之间的关系。

在与正向的语言参项对比之后，我们认为该语言参项可能与论证有关。

维度 4 上正向与负向的语体差异可以概括为叙事性或论证性。换言之，正向语言参项的共现，表现出较强的叙事性；负向语言参项的共现，表现出较强的论证性。因此，我们将维度 4 命名为：事件重构与论证阐释。

五、口述历史语篇的语体分析

不同的语体在维度上存在着某些差异，研究者可以通过对维度分值的计算，有效地、客观地描述语体之间的差异，进而概括出某种语体的普遍性特征。<sup>[5]223</sup> 因此，我们根据上文所确定的四个维度，计算出口述历史语篇在每个维度中的平均维度分值；<sup>④</sup> 通过与日常谈话、学术论文语篇的对比，并结合语境因素，来分析口述历史语篇的语体特征。

根据语言参项解释力的大小，我们在口述历史、日常谈话、学术论文语篇中，选取因子负荷量绝对值为 0.4 以上的语言参项，对它们在这四个维度上的平均维度分值进行计算，并确定了口述历史语篇在每个维度上的位置（图 2）。

(一) 口述历史语篇具有人际互动性

在维度 1 上，口述历史语篇和日常谈话语篇的维度分值均分布于正向，分别为 3.5 和 12.0；而学术论文语篇的维度分值则分布于负向，为-15.3。这说明口述历史语篇在这个维度

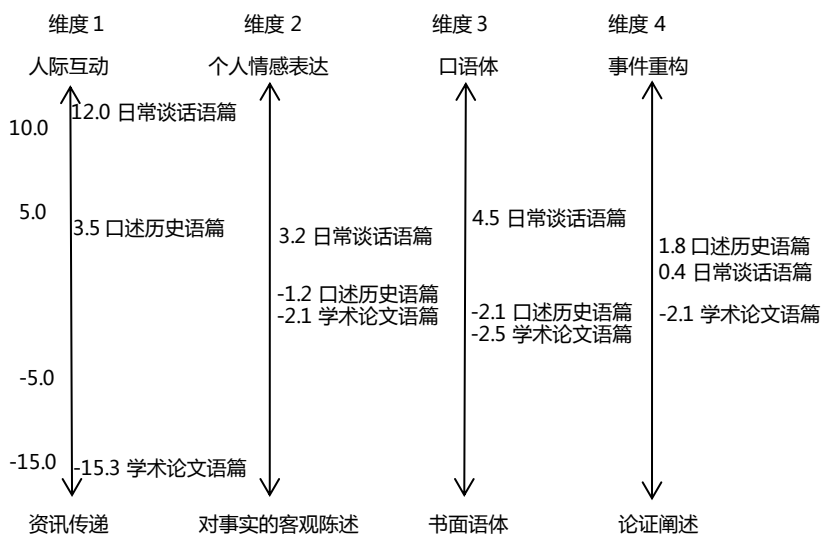


图 2 口述历史、日常谈话、学术论文语篇维度分布对比

上接近于日常谈话语篇,而不同于学术论文语篇。口述历史语篇和日常谈话语篇都具有人际互动性。

#### 例(1): 口述历史语篇

记者: 下边几个纵队?

陈海峰: 都合并了, 那都撤销, 他原来是什么 11 纵队、12 纵队、10 纵队。都重新编。

记者: 整编之后, 情况是什么样的?

陈海峰: 整编以后呢, 就是有 11 万人。就是我们苏浙……

记者: 6 个纵队是吗?

陈海峰: 不止。那个时候你看嘛, 1 纵队走了, 2、7 纵队还有, 6 纵队是原来 24 军的, 那个陶勇的 4 纵, 这两个主力换到华中。还成立 10 纵、11 纵、12 纵。11 纵、12 纵(是)新成立的, 10 纵这司令是谢祥军嘛。

#### 例(2): 日常谈话语篇

(手机屏幕上正播放着一个两岁左右男孩说话的视频)

妻子: 这是什么? 他说什么? 我听不懂。

丈夫: 划拳。

妻子: 教小孩儿这个, 真不好。

丈夫: 这是培养独立性!

妻子: 什么?! 什么人划拳? 酒鬼!

丈夫: 你说话真难听。你太不会说话了。

#### 例(3): 学术论文语篇

北京青少年法律援助研究中心调查显示, 农村留守女童性侵案件中, 犯罪嫌疑人为老年人的比例为 82.2%, 与受害人为熟人关系的占 65.74%。近年来教师侵犯儿童的案件数量也不断增加。2013 至 2014 年曝光的 192 起性侵儿童案件中, 师生关系达 42 起; 邻居关系紧随其后, 达 36 起。第一, 由于大量青年壮年劳动力外出, 农村中居住主体多为中老年人, 留守女童的周边环境为老年人居多, 其中尤以留守老人作案者居多。<sup>[12]</sup>

例(1)摘自口述历史语篇, 例(2)摘自日常谈话语篇。这两例中, 都出现了正向的语言参项, 如第一、第二、第三人称代词、指示

代词、语气词、疑问句、感叹句等。上述语言参项的共现都与人际互动有关。例(3)摘自学术论文语篇, 例中的陈述句、复句等负向语言参项的共现, 都与资讯传递有关。

因为口述历史语篇和日常谈话语篇都采用了对话的方式, 身处现场的交际者都采用了面对面的、即时交流的方式, 所以它们都具有人际互动性。而学术论文语篇的作者与读者是无法进行面对面的交流的, 他们只能通过书面形式进行单向的、无即时反馈的交流。因此, 学术论文语篇不具有人际互动性。

从维度分值上, 我们还可以看出口述历史语篇的人际互动性弱于日常谈话语篇。这是因为日常谈话语篇的交际目的是维系人际关系, 所以它的互动性很强; 而口述历史语篇的交际目的则是讲述历史, 因此, 它的互动性不及日常谈话语篇。

#### (二) 口述历史语篇重在事实的客观陈述

在维度 2 上, 口述历史语篇和学术论文语篇的维度分值均分布于负向, 分别为 -1.2 和 -2.1; 而日常谈话语篇的维度分值则分布于正向, 为 3.2。这说明口述历史语篇在这个维度上接近于学术论文语篇, 而不同于日常谈话语篇。口述历史语篇和学术论文语篇都注重对事实的客观陈述, 而非个人情感或立场的表达。

#### 例(4): 口述历史语篇

陈海峰: 新四军的卫生部长, 那么后面我还要介绍到他。那么宫乃泉是新四军军医处的医务主任, 医务主任兼医政科长。他是辽宁的神经医学院毕业的。崔义田、宫乃泉, 齐仲桓是新四军四师卫生部长, 也当过三师卫生部长。还有一个孙志诚是药剂师。崔、宫、齐三个不是, 这是辽宁医学院的。上海医科大学(在)国际上是有名的, 那么是吴之理, 志愿军的卫生部长; 章央芬, 协和医科大学的副校长。

#### 例(5): 学术论文语篇

作为现代国家治理之重要主体的行政机关, 它代表公共利益对社会进行治理, 以形成符合社会正义之生活关系、规划及推动基本建设, 引导及维持合于公意之政治发展。行政机关主要依照法律规

定,通过对公共利益的集合、维护和分配来实现上述目的。<sup>[13]</sup>

例(4)摘自口述历史语篇,受访者讲述了新四军医疗机构和人员的情况,例中出现了陈述句以及表示人名、地名、机构名的名词。例(5)摘自学术论文语篇,例中出现了陈述句、普通名词等。在上述两例中,陈述句、名词等负向语言参项共现的情况,说明了口述历史语篇和学术论文语篇都注重对事实的客观陈述。这与语篇的交际目的有关。口述历史语篇的交际目的是通过受访者的叙述来挖掘史实,学术论文语篇的交际目的是介绍学术成果。因而对这两类语篇而言,对事实的客观陈述尤为重要。

而日常谈话语篇的交际目的则是用语言来维持并发展对话双方的关系,所以它注重个人情感和态度的表达。因此,在日常谈话语篇中频繁共现的是正向语言参项,如例(2)中的状中短语、非主谓句等。

从维度分值上来看,口述历史语篇在对事实的客观陈述上弱于学术论文语篇。这是因为口述历史语篇中的受访者在述说历史事件、历史人物时,难免带有主观色彩。而学术论文语篇则要求真实、客观。因而,学术论文语篇对客观陈述的要求也就更高。

### (三)口述历史语篇具有书面语体的倾向

在维度3上,口述历史语篇和学术论文语篇的维度分值均分布于负向,分别为-2.1和-2.5,而日常谈话语篇的维度分值则分布于正向,为4.5。这说明口述历史语篇在这个维度上接近于学术论文语篇,而不同于日常谈话语篇。口述历史语篇和学术论文语篇都具有书面语体的特征。

#### 例(6):口述历史语篇

解说:虽然调查结果显示赵总编的历史并无反动之处,但是洪炉回报社之后还是就一些小问题进行了揭发。

#### 例(7):学术论文语篇

同时,网络语言也兼容了口语同文字双方的一些特性,网络即时交往大都要求使用文字,然而却必须发挥出口语功能,即时交流时间短暂,且要求对电脑设备的

熟练使用,因此,网络语言的应用同时要求书面语言的文采、口语的生动及电脑的快捷。<sup>[14]</sup>

#### 例(8):日常谈话语篇

甲:我也感觉好像是小的芒果比较甜诶。

乙:对呀。

丙:但大大的芒果一般来说也是甜的。

甲:哇,你干吗?

乙:开胃啊。

例(6)摘自口述历史语篇,例(7)摘自学术论文语篇。在这两例中,都出现了复句、单句平均层数等负向语言参项,它们的共现与书面语体有关。例(8)摘自日常谈话语篇,例中出现了单句、非主谓句等正向语言参项,它们的共现都与口语体有关。

这与口述历史语篇的生成环境有关。口述历史语篇是有准备的、经过加工的语篇。一般来说,在做口述历史访谈前,采访者要做充分的准备工作,包括收集、比对、甄别材料,拟定采访提纲等。访谈时,采访者和受访者面对镜头展开对话,他们在镜头前都会调整自己的语言。访谈后,还要有专人将录音转写成文本,并对文本进行考证。可见,口述历史语篇是准备程度较高、经过编辑加工的言语成品,所以它具有书面语体的倾向。学术论文语篇的作者在动笔之前,同样也需要查阅文献,写完之后还要反复修改、润色加工。而日常谈话语篇大多是即时生成的。参与者通常既没有时间思考,也无法对已经说出的话进行编辑。因此,日常谈话语篇是典型的口语体。

从维度分值上,我们还可以看出口述历史语篇的书面化程度低于学术论文语篇。这是因为,学术论文是典型的书面语体,它的准备程度、加工程度,通常都比口述历史语篇更高。

### (四)口述历史语篇具有叙事性

在维度4上,口述历史语篇和日常谈话语篇的维度分值均分布于正向,分别为1.8和0.4;而学术论文语篇的维度分值则分布于负向,为-2.1。这说明口述历史语篇和日常谈话语篇一样,具有叙事性,而学术论文语篇则重在论证。

#### 例(9):口述历史语篇

陈海峰:上海党领导下,我哥哥是老



地下党员,1934年就在那。我是第一个过去,所以我见陈毅,到了皖南的时候,到这个班之前,见过四次,面对面的。他也动员我,学完了一定要回一支队,结果后来就留在皖南了(陈海峰笑),派到三支队去了。

记者:那个时候陈毅已经到江南了,是吧?

陈海峰:去了去了。后来他来了新四军,1938年上半年来了以后,马上和我们接上了,我们这个部队就编到他那里去了。

例(10):日常谈话语篇

甲:那天可热闹了。腾格尔还有那个,那个谁,都来了,去那个民大的那达慕大会。可那天雾霾太重了,晚上回来的时候可难受了。还有摔跤呢。

乙:哇哦。

甲:在那儿摔了三十多分钟,两个人都没有劲儿了。

丙:你,你,你有尝试吗?

丁:大会在哪儿开的?

甲:在中央民大。

丙:哦。

例(11):学术论文语篇

自20世纪90年代中期我国便有传统媒体开始了探索之旅。如今近20年已过去,就在一个又一个后起的商业新兴媒体如搜狐网(1998年成立)、腾讯网(1998年成立)、淘宝网(1999年成立)、百度搜索(2000年成立)、360安全卫士(2005年成立)等飞速成长起来,并如空气一般影响了现代人的学习、生活、工作之时,我国传统媒体在发展新兴媒体业务方面虽然不时也有一些声势浩大的举措,但成功的案例却寥寥无几。<sup>[15]</sup>

例(9)摘自口述历史语篇,例(10)摘自日常谈话语篇。在这两例中,用于叙事的时间名词、地名等正向语言参项的共现,用于叙事。例(11)摘自学术论文语篇,例中出现了表示并列、转折等关系的负向语言参项连词,用于说明逻辑关系。

这同样与口述历史语篇的交际目的有关。

口述历史语篇的交际目的是讲述历史,日常谈话语篇也常用于叙事。而学术论文语篇则重在论证。这就是学术论文语篇在维度4上的分值比口述历史语篇、日常谈话语篇低很多的原因。

从维度分值上,我们还可以看出口述历史语篇的叙事性比日常谈话语篇更强。这是因为口述历史语篇的交际目的是挖掘史实,而日常谈话语篇的交际目的则在于维系人际关系。

## 六、余论

对话体研究而言,多维度分析不仅是一种极具操作性的研究方法,而且在方法论上也具有极其重要的意义。多维度分析以大规模语料库为基础,穷尽式地提取语言参项作为语体标记,突破了以往语体计量分析语料规模较小、语言参项较单一的局限,使语体研究更精确、更全面。

在对语体进行量化分析的同时,多维度分析也非常注重质化分析。多维度分析认为语言参项是功能性的,也就是说,某些语言参项之所以倾向于在某种语体中出现,是因为它们特别适合这种语体的交际目的和情景语境。因此,研究者在分析语体时,既要描写语言参项,还要分析情景语境,更要从功能的角度探究情景语境与语言参项之间的关系,通过对两种或多种语体的对比,分析语言参项产生的原因。

本文是对多维度分析的一次有益的尝试。我们采用多维度分析的方法来分析口述历史语篇的语体特征,根据语料,参考相关研究成果,从口述历史语篇、日常谈话语篇、学术论文语篇中提取了56个语言参项;通过因子分析,确定了29个语言参项,初步建立了一个适用于汉语口述历史语篇的语言参项分析框架。通过因子分析,我们得出“人际互动与资讯传递”“个人情感表达与对事实的客观陈述”“口语体与书面语体”“事件重构与论证阐释”四个因子,并将其作为基本维度。

通过与日常谈话、学术论文语篇维度分值的对比,我们发现口述历史语篇具有以下语体特征:口述历史语篇倾向于人际互动、重在对历史事实的客观陈述、具有书面语体的倾向、具有叙事性,上述语体特征是由口述历史语篇

的交际方式、交际目的以及生成环境决定的。

口述历史语篇与日常谈话语篇、学术论文语篇既有相同之处，同时又存在着一些区别。与日常谈话语篇的相同之处是，它们都具有互动性和叙事性；二者的不同之处是，口述历史语篇重在事实的客观陈述，而日常谈话语篇则重在个人情感和立场的表达；口述历史语篇具有书面语体的倾向，而日常谈话语篇则是典型的口语体。同时，我们意外地发现口述历史语篇与学术论文语篇也存在相同之处。口述历史语篇与学术论文语篇的相同之处是，二者都注重对事实的客观陈述。它们的不同之处是，口述历史语篇重在人际互动，而学术论文语篇则重在资讯的单向传递；口述历史语篇重在事件的叙述，而学术论文语篇则重在论证阐释。而这三类语篇的语体差异形成的原因是它们在交际方式、交际目的、生成方式上的差异。

(陆俭明先生对本文提出宝贵意见，谨致谢忱。)

#### 注释：

- ① 口述历史语料来自崔永元口述历史研究中心、“人民记忆”“我的抗战”“西南联大抗战史”等微信公众平台、袁成毅和丁贤勇主编的《烽火岁月中的记忆：浙江抗日战争口述访谈》(2007年，北京图书馆出版社)以及凤凰网等网站，网址为 [http://news.ifeng.com/a/20140706/41034594\\_0.shtml](http://news.ifeng.com/a/20140706/41034594_0.shtml) 等。
- ② 根据张斌关于句子的理论，对单句进行短语类型标注和层次分析，对复句进行语义逻辑分析，根据各分句之间的逻辑关系划分复句层次。
- ③ Python 是常用的程序设计语言之一。通过编程，Python 可以对中文进行处理。我们主要依据 Python 的分句功能，对语料进行分句。
- ④ 根据张斌对句子的定义，在转写语料，尤其是日常谈话语料时，将能够表达相对完整意思，具有一定语气，前后有明显停顿的词或短语看作一个句子。
- ⑤ 语言参项，即参与统计的语言变量，可以是涵盖词汇、语法、语音等层面的语言特征。
- ⑥ 因子分析是一种降维(即数据简化)的分析方法，即通过研究原始的大量数据之间的内部依赖关系，对这些数据进行浓缩、归纳，提取出少量能反映原始数据的因子。
- ⑦ 因子旋转是因子分析中的一个核心步骤。其目的是使得到的因子结构尽量简化，以突出各个因子中典型的变量，这样更容易发现因子的作用，并进一步对维度进行命名。
- ⑧ 在使用 SPSS 软件对原始数据进行降维计算时，选用的是基于主成分模型的主成分分析法。主成分分析法将原来多个变量划分为少数几个综合指标，是常见的一种降维处理方法。
- ⑨ 碎石图：横坐标表示因子个数，纵坐标表示特征值，因为图中的特征值点以圆圈的形式表现出来，似碎石一般，故名碎石图。碎石图用于因子分析，以直观地评估哪些因子占数据

中变异性的绝大部分。一般而言，曲线越陡，该段代表因子所占数据的变异性越大，就越能涵盖整体信息，显著性也越大。

- ⑩ 为便于统计，且使统计数据具有可比性，这里的均值是将原始数据进行归一化处理，经 SPSS 软件计算所得，归一化的计算方法为原频率 $\times 100$ ，即以 100 个词(或句子)为单位的频数。
- ⑪ 根据层次分析法对单句和复句进行层次划分。复句的层次划分仅限于不同分句之间逻辑关系的划分，不包括各分句内部层次关系的划分。一般来说，层次越多，句子结构越复杂。单句平均层次数以及复句平均层次数，体现了句子的复杂程度。
- ⑫ 因子负荷量为正的语言参项对因子具有正向的影响，即因子负荷量越大，正向的特征越显著；而因子负荷量为负的语言参项则对因子具有负向影响，即因子负荷量的绝对值越大，负向的特征越显著。语言参项的因子负荷量的绝对值越高，该语言参项对因子的贡献就越大。为清晰地表示每个语言参项对不同的因子的贡献值，这里按照语言参项的因子负荷量的大小排列，而不考虑它们之间的上下位关系。
- ⑬ 表中“初始特征值”“提取平方和载入”“旋转平方和载入”栏下的三列分别表示特征值、方差贡献率和累计方差贡献率，其中方差贡献率用来表示所提取因子能够解释的信息量；“提取平方和载入”栏中的累计方差贡献率表明提取六个因子，能够共同解释 75.794% 的信息，丢失的信息量较少；“旋转平方和载入”表示旋转后的因子在特征值、方差贡献率和累计方差贡献率的情况，累计方差贡献率没有改变，仍为 75.794%，表明因子旋转后没有改变原有的共同度，但却重新分配了各个因子解释原有变量的方差，改变了各因子的方差贡献率，使各因子更易于解释。
- ⑭ 在多维度分析中，平均维度分值的计算方法是，先用正向的语言参项的标准化频率之和，减去负向语言参项的标准化频率之和，由此得出每个文本在各维度上的维度分值。然后，再计算得出不同语篇在各维度上的平均维度分值。如果维度分值为正，就表明语篇具有正向的语体特征；如果维度分值为负，就表明语篇具有负向的语体特征。这里的标准化频率的计算，采用的是 SPSS 软件最为常用的标准化方法，即 z-score 标准化(zero-mean normalization)，也叫标准差标准化，经过处理的数据符合标准正态分布，即均值为 0，标准差为 1。

#### 参考文献

- [1] 唐纳德·里奇. 大家来做口述历史——实务指南[M]. 王芝芝, 姚力, 译. 第 2 版. 北京: 当代中国出版社, 2006.
- [2] 赵雪, 王莉. 口述史、民族志、新闻访谈话题的对比——基于词共现网络的分析[J]. 云南艺术学院学报, 2018(2): 13-19.
- [3] 韩礼德, 哈桑. 英语的衔接[M]. 张德禄, 王钰纯, 韩玉萍, 等, 译. 北京: 外语教学与研究出版社, 2007.
- [4] BIBER D. Variation across speech and writing[M]. New York: Cambridge University Press, 1988.
- [5] BIBER D, CONRAD S. Register, genre and style[M]. New York: Cambridge University Press, 2009.
- [6] CONRAD S. Investigating academic texts with corpus-based techniques: an example from biology[J]. Linguistics and

- Education, 1996 (8): 199-326.
- [7] 桂诗春. 基于语料库的英语语言学语体分析[M]. 北京: 外语教学与研究出版社, 2009.
- [8] 张学谦. 台语口语及书面语体的多面向分析[J]. Language and Linguistics, 2000 (1): 89-117.
- [9] 张斌. 现代汉语描写语法[M]. 北京: 商务印书馆, 2010.
- [10] 张玉, 郭会雨, 陈建青. 我国数据挖掘研究现状分析——基于共词分析视角[J]. 情报科学, 2011 (10): 1589-1593.
- [11] 赵雪. 刍议广播电视语言中的长短句[J]. 现代传播, 1996 (2): 87-89.
- [12] 王伟. 法律视角下留守儿童遭受性侵案件分析[J]. 法制与社会, 2015 (22): 57-58.
- [13] 张扩振. 论行政诉讼原告资格发展之历程与理念转换[J]. 政治与法律, 2015 (8): 87-95.
- [14] 康庆业. 传播学视角下的网络语言研究[J]. 语文建设, 2014 (24): 39-40.
- [15] 朱鸿军, 农涛. 媒体融合的关键: 传媒制度的现代化[J]. 现代传播, 2015 (7): 6-11.

### Multidimensional Analysis of Oral History Text

ZHAO Xue<sup>1</sup>, LU Jin-fang<sup>2</sup>, GONG Chun-hui<sup>3</sup>, ZHANG Yan-xue<sup>4</sup>

(1. College of Humanities, Communication University of China, Beijing 100024, China; 2. The Secondary Specialized School of Shaoxing, Shaoxing 312000, China; 3. Media College, Shandong University of Political Science and Law, Ji'nan 250014, China; 4. Ireader Company, Beijing 100024, China)

**Abstract:** On the basis of comparing with daily conversation and academic papers with Multidimensional Analysis, this paper extracted four basic dimensions of oral history. These four basic dimensions were as followed: interpersonal interaction vs. information transmission, expression of personal feelings vs. objective statement of the fact, oral register vs. written register, event reconstruction vs. demonstrative explanation. The study showed the results as followed: the oral history text tended to be more inter-personally interactive; focused more on objective statement of the fact; was well-prepared and processed speech product with a tendency of written register; highlighted on the reconstruction of historical events, hence with the features in narratives. These features were determined by the ways and purposes of communication, and production circumstances of the oral history text.

**Keywords:** oral history text; multidimensional analysis; interpersonal interaction; objective statement; a tendency of written register; narrative

(责任编辑 周 芬)