

口译语料库中副语言信息的转写及标注: 现状、问题与方法

邹兵¹, 王斌华²

(1. 广东外语外贸大学 高级翻译学院, 广东 广州 510420;

2. 香港理工大学 中文及双语学系, 香港 九龙)

[摘要] 本文对口译副语言信息的相关概念进行了界定,并检视了国际范围内口译语料库中副语言信息的转写标注情况,发现现有研究对副语言信息的转写标注等口译语料库基础建设问题缺乏关注。本文基于笔者设计和建设口译语料库的经验,归纳了口译副语言信息转写及标注应注意的问题,并从标注工具、标注步骤和后期建设几个方面探讨了口译副语言信息的转写及标注方法。

[关键词] 口译语料库; 研究现状及问题; 副语言信息; 转写及标注

[中图分类号] H059 **[文献标识码]** A **[文献编号]** 1002-2643(2014)04-0017-07

Transcription and Annotation of Paralinguistic Information in Interpreting Corpora: The Status Quo , Problems and Solutions

ZOU Bing¹ ,WANG Bin-hua²

(1. School of Interpreting and Translation Studies , Guangdong University of Foreign Studies , Guangzhou 510420 , China;

2. Department of Chinese and Bilingual Studies , The Hong Kong Polytechnic University , Hong Kong)

Abstract: In this paper the authors first define some relevant concepts of paralinguistic information (PI) , and then review the status quo of PI transcription and annotation in existing interpreting corpora around the world. It is found that more attention is needed for research into this issue. The authors of this paper then , integrating their experiences in interpreting corpus design and construction , summarize the major problems that need to be considered in transcribing and annotating PI , and explore the methods of PI transcription and annotation in interpreting corpora in terms of tools , procedures and post-construction maintenance.

Key words: interpreting corpora; status quo and problems; paralinguistic information; transcription and annotation

1.0 引言

语料库应用于翻译研究已逾 20 个年头。当前

基于语料库的笔译研究多关注语言层面 ,即集中于译文语言特征研究和译者语体/文体风格研究。对基于语料库的口译研究而言 ,语言层面之外的副语

收稿日期: 2013-12-06

基金项目: 本研究得到香港理工大学科研项目(G-UA92)和广东外语外贸大学研究生科研创新项目(14GWCXXM-41)的资助。

作者简介: 邹兵(1986-)男,广东外语外贸大学高级翻译学院博士生。研究方向:翻译研究。

王斌华(1974-)男,博士,香港理工大学中文及双语学系助理教授(研究)。研究方向:口译研究、翻译研究。

言和超语言信息也值得关注,因为这些信息对于研究者分析口译产品的特点以及考察口译过程起着关键作用。而在口译语料库建设中,语言层面之外的信息转写和标注是一个难题,当前关于这一难题的讨论(包括转写内容、方法、工具、原则与标准等)却不多见,这一定程度上制约了语料库口译研究的发展。在本文中,笔者在检视世界范围内主要口译语料库的基础上,结合自身设计与建设口译语料库的经验,探讨口译语料区别于笔译语料的特有信息——副语言信息的转写和标注问题,希望能对口译语料库建设的标准化有所贡献。

2.0 口译语料中的副语言信息

语言学中的副语言信息概念最初由 Trager (1958) 在“Paralanguage: A First Approximation”一文中首次使用(梁茂成,1994: 128),指在与词汇和语法层面平行的信号层面上的非语言话语信息(陈瑞青、王巍巍 2011: 5)。

口译语料涉及的信息大致可分为三类,即语言信息(linguistic information)、副语言信息(paralinguistic information)和超语言信息(extra-linguistic information)。基于 Roach, et al. (1998)、Monti, et al. (2005)、张威(2009)等的观点,结合口译语料库的特点和设计需要,笔者对这三类信息界定如下:

1) 语言信息,即口译源语与译语中词句篇章各个层面的信息,包括:词性标注、句法标注、时间标记、句子段落标记、词句段对齐,等等。

2) 副语言信息,即源语和译语产出的同时所伴随产生的相关信息,包括:停顿、支吾语(犹豫)、填充语、重音、语音拖长、自我修正、打断、话语重叠、不完整句、幽默、肢体语言,等等。

3) 超语言信息,即与口译活动相关的环境信息,包括:①口译背景信息,如口译主题、口译场合、时间地点、源语语体、专业难度、技术设备等;②讲话人信息,如口音、语速、时长及字数、信息密度、国籍、性别、政治身份等;③口译员信息,如口译经历、专业

级别、口译形式、准备时间、国籍、性别、母语等;④口译听众信息,如知识背景、与会目的、双语水平等;⑤口译活动赞助人、组织者信息,等等。

这些信息是口译语料库建设时所应转写和标注的基本信息。关于语言信息和超语言信息的转写和标注,当前笔译语料库建设已经积累了较为标准化和可操作化的方法、工具和体系。口译语言信息的标注基本可以参照笔译或笔语语料库的标注体系(如词性标注集、句法标注集等),超语言信息的标注也基本可以在头文件(text header)中统一处理即可。(梁茂成、许家金 2012)

但是,口译副语言信息的转写和标注问题,目前尚未引起足够的重视,这从相关研究的缺失即可看出。有必要指出的是,副语言信息的标注对于口译语料而言有着特殊的意义,因为副语言信息“有利于判定具体口译策略的影响因素以及这些策略的应用效果”(张威 2009: 56)，“有助于揭示口译语体特征和译员风格的差异”(胡开宝、陶庆,2010: 52),有助于“研究非言语因素对口译的方向性和语言特征的影响”(李婧、李德超 2010: 101)。

从笔者设计和建设口译语料库的经验以及作为口译研究者和口译教师对口译语料库功能的期待来看,口译语料库建设之初的设计工作应引起极大的重视。因为口译语料库建设是一项耗时耗力的庞大工程,哪怕只是一个磁带小时的口译语料,往往需要花费数倍的时间来转写和标注,所以在口译语料转写之前就应当根据研制语料库的目的确定要关注的信息类别,在转写的同时标注这些信息。

3.0 现有口译语料库中副语言信息的标注

根据笔者掌握的资料,目前世界范围内已建和在建的口译语料库仅有十多个。此处对现有口译语料库的建设情况作一检视,重点关注其对口译副语言信息的转写及标注,详见表1的描述。

表1 现有口译语料库建设及其副语言转写标注情况^①

语料库	文献来源	语料来源及范围	语种及库容	副语言转写标注情况
EPIC	Russo et al. (2012)	欧洲议会口译语料	意大利语、英语、西班牙语; 17万余词, 21小时(未来计划新增27万词, 21余小时)	不完整句、拼读错误、无声停顿、有声停顿、无法辨识
CIAIR	Tohyama et al. (2004)	实验室状态下的职业译员口译录音	日语、英语; 近100万词, 182小时	声音特征、填充语、犹豫、修正、吞音或不正确发音、语音(元音/辅音)拖长、无法辨识
CECIC	胡开宝、陶庆(2010)	中国政府及国务院新闻发布会汉英平行语料、新闻发布会英语原创语料和政府工作报告汉英平行语料	汉语、英语; 54万余词	停顿、不完整句或言语打断、口误及修正、犹豫、非言语声音、发音含混、重复、语言错误或不规范
PACCEL	文秋芳、王金铨(2008)	2003-2007年中国18所高校英语专业大三四大学生的英汉汉英口译测试	汉语、英语; 210万字	(口译学习者) 错误类型
DIRSI-C	Bendazzoli (2012)	三场医学国际会议的现场同传	意大利语、英语; 13.6万词, 565分钟	言语事件、话轮转换、语段时长、语段长度、语速
FOOTIE	Sandrelli (2012)	2008年欧洲足球锦标赛16场新闻发布会现场同传	意大利语、西班牙语、法语、英语; (库容未知)	言语事件、话轮转换、话轮时长、话轮长度、话轮语速
CoSi/K6	House et al. (2012)	三场关于反对转基因食品的系列讲座的口译录音(两场交传、一场同传)	巴西葡萄牙语、德语; 3.6万词, 近6小时	明显的发音变异现象、停顿、笑声、咳嗽声、明显的重音、声音拖长、音量或声调变化、转写者评论
ComInDat	Angermeyer et al. (2012)	从三个社区口译语料库(DiK, IiSCC, SimDik)中抽取语料组建而成	英语、(标准)德语、海地语、葡萄牙语、波兰语、西班牙语、土耳其语、俄语; (库容未知)	传译策略(主要是源语与译语对应情况, 如增译、减译、压缩、替换、拆译等)、语言接触现象(主要是IiSCC子库)
DiK/K2	Bührig et al. (2012)	德国数家医院医患对话口译录音	德语、葡萄牙语、西班牙语、土耳其语; 17万词, 23小时	无法辨识/不确定、非语音现象、停顿、标点符号、话语修正、转写者评论
IiSCC	Angermeyer (2006)	纽约市小额法庭非正式法庭诉讼审理口译录音	海地克里奥尔语、波兰语、俄语、西班牙语; 40场口译	停顿、转写者评论、语言接触现象(如语码转换、语码混合、插入、仿造)、叙述角度(人称)变化
TIC	Cencini & Aston (2002)	电视广播节目口译影音	语种未知; 近4万词, 4小时	依据TEI转写规范, 包括停顿、话语重叠、笑声等
CorIT	Falbo (2012)	意大利近50年的电视节目同传和交传语料录音, 包含三个子库, 即脱口秀节目口译库、一级方程式赛车(F1)直播节目同传库、美国/英国首脑竞选直播节目同传库	意大利语、英语、法语、俄语、西班牙语、德语、葡萄牙语、塞尔维亚语、阿尔巴尼亚、波斯尼亚、日语; 2700场口译, 总库容未知	机构性会话特征(相邻对、反馈语、话轮转换、话语修正、话语重叠、打断、停顿等)
FPC	Straniero Sergio (2012)	CorIT中的F1直播节目同传子库	英语、意大利语; 340场口译, 近30小时	话轮转换、停顿、语音变异(口音)等

通过考察这些口译语料库的建设情况, 笔者发现, 其副语言信息转写及标注存在以下几方面问题:

1) 转写及标注内容选择不一。表1中较大型的口译语料库(如EPIC、CIAIR、CECIC、CoSi、CorIT)对副语言信息转写和标注较为全面, 其它口译语料库则选择性地转写和标注了停顿等部分副语言信息。各个语料库选择转写和标注的内容重合度较低, 没有体现对口译语料最基本的一些副语言特征的关注。

2) 转写及标注规范不统一。如EPIC、CECIC、

TIC按照TEI文本编码规范(见4.1.2), ComInDat及其子库(DiK, IiSCC, SimDik)、CoSi遵循HIAT口语文本转写标注规范(见4.1.2), CorIT则依据会话分析领域的转写规范, 其它语料库则根据需要自行设定了转写和标注符号。

3) 较少考虑语料库兼容问题。语料库之间的兼容体现在若干方面, 除了要转写标注一些基本特征以及要遵循统一的转写标注规范之外, 语料库文件格式还要能方便地转换为现在所普遍提倡的XML格式, 从而方便未来的跨库研究。而表1中只

有较少语料库(如 EPIC、CECIC) 考虑到了与其它语料库的兼容问题。

总体而言,当前多数口译语料库的开发者并没有很详细地描述其副语言信息的转写标注过程,再加上很多语料库本身并不公开,因此对于其它口译语料库的副语言信息转写标注无法提供很好的指导和借鉴。

4.0 口译副语言信息转写标注的问题与方法

上文的考察反映出,国内外学者在探讨口译语料库开发及建设时,都认为口译语料深加工是需要重点关注和解决的问题,但是少有学者论及口译副语言信息转写和标注的具体操作问题。下面笔者将结合自己设计和开发口译语料库的经验,探讨副语言信息转写标注应注意的问题和转写标注的方法。

4.1 口译副语言信息转写标注应注意的问题

4.1.1 转写及标注内容的选择

口译副语言信息层面可以转写标注的内容,除了上文(见 2.0)提及的类别之外,还可以包括口译学习者的口译错误和职业译员的口译策略。至于这些内容如何选择,则须遵循两点原则:

1) 明确建设语料库的最终目的。即在建库之初,要明确口译语料库使用方(包括口译研究者、口译学习者、口译教育者、口译实践者、机器口译研发者等)的需求。面向的服务对象不同,对口译副语言信息转写和标注程度的要求也有所不同。同时,转写和标注内容的选择也要考虑现有资源(包括语料规模、资金、人员等)以及所掌握技术的情况。

2) 体现对口译基本/共性特征的关注。即对口译活动表现的一些基本和共性的副语言特征表示关注。这样做的目的是增强各个领域、各种形式、各个语种的口译语料库的可比性和兼容性,有利于未来进行跨语料库的多语类多语种口译比较研究。笔者认为口译语料库应当转写和标注以下几类基本的副语言信息:①言语行为特征,包括停顿(又可分为无声停顿和有声停顿)、犹豫、填充语、不完整句、自我修正、打断、话语重叠、重复、口误、不规范用语;②明显的发声特征,如拼读错误、语音拖长、语音变异(口音)、语速、音量/调变化、笑声、咳嗽等;③明显的体态语特征,如眼神、手势等面部表情和肢体动作;④无法辨识的现象,如因设备影响而听不清、故意含混不清等现象;⑤突发事件;⑥转写者评论,如幽默、错误、策略等。

4.1.2 转写及标注体系与规范

当前各口译语料库遵循的规范大致有三类,即

TEI 文本编码规范、HIAT 口语文本转写标注规范和会话分析转写规范。

TEI 全称为 Text Encoding Initiative,是一个国际性的跨学科的编码标准,提倡使用可扩充置标语言 XML 对数据和语料语言及结构信息进行编码,现行版本 TEI P5 专辟一章说明如何转写语音语料^②。HIAT 全称为 Halbinterpretative Arbeitstranskriptionen (Semi-Interpretative Working Transcriptions),现已发展成为集转写标注格式规范和转写标注工具为一身的 EXMARaLDA 系统,主要致力于解决口语文本的转写和标注问题^③。语言学中的会话分析(Conversation Analysis, CA) 领域长期关注机构话语和日常会话的结构、策略和风格特点,并形成了一套比较系统和完整的会话转写规范。(Schiffrin, 1994)

对口译副语言信息而言,这三类规范各有优势,各大口译语料库对这三类规范也是各有青睐,但其彼此之间既有交叉也存在一定差异。随着未来口译语料库建设日益走向标准化,还是有必要根据口译活动自身的特点以及口译语料库“目标用户”的需要,研制出一套普遍适用且能被广泛采用的口译语料库副语言信息转写及标注体系或规范。(Cencini & Aston, 2002) 可以说,“语料转写是决定口译语料库代表性的一项关键工作,转写的程序与操作规范都可以成为研究课题”。(张威, 2013: 83)

4.1.3 其它相关问题

1) 转写标注者主观因素

在副语言信息转写和标注过程中,需要注意的一个重要问题是如何避免或尽量减少转写标注者的主观因素,因为这些主观因素往往容易导致语料标注前后不一致甚至相互冲突。为此,可以采取的措施有:①在转写标注之前专门进行集体培训,并进行试验性转写标注,在试验期间进行反复调试,直至完全符合要求后再正式参与转写标注工作;②专设核对和“质检”的角色,即时监控转写标注过程,随时发现问题随时更正;③在工具开发上尝试设计标准化的转写标注功能,对每一类副语言信息设置单独的转写标注模块,出现此类信息时直接点击选择,自动生成相应的转写标注符号,从而减少插入符号时出现的失误;④有些涉及口译错误和口译策略的副语言特征,本身便存在主观判别的风险,这便需要事先进行明确定义,在操作过程中一以贯之地执行。

2) 语料库的兼容问题

现有语料库大多存在重复建设、转写标注标准不统一以及经过调整也很难融合的问题,即语料库之间的兼容性太差。当前语料库建设普遍提倡在建库时即采用 XML 置标语言,或者所建语料库能够方

便地转换为 XML 文件格式,这对于语料库的标准化、网络化十分关键。对于口译副语言信息而言,具体的转写标注内容和符号需要研究者结合口译活动特点和口译研究需要,制定一个通行的可操作性强的操作准则和细则。口译语料库建设尚刚刚起步,在起步之初,如果各口译语料库的设计者在语料选取上能够尽量避免同质和重复,并且能很好地遵循通行的转写标注规范和体系,那么随着越来越多的语料库形成一个大的集合,未来的口译语料库研究一定会发挥越来越大的作用,远非现在的各自为战所能比拟。

3) 语料库的应用问题

语料库建好之后如何应用,这是在建库之初就应思考的问题。口译语料库中所转写标注的副语言信息如何应用,笔者认为可以从其最终目的出发进行考虑:①若为口译研究者服务,便要清楚认识到副语言特征可以说明什么问题,如停顿、支吾语、填充语等可能与译员当时当地的心理活动有关,若再针对这些副语言现象出现的规律提出一定的研究假说,与其它的实证研究手段(如 TAPs、ERPs、fMRI 等)结合进行三方验证,便能很好地解释和预测复杂口译过程中的一些现象,但需要注意的是进行语料库口译研究时,应特别重视方法论设计(Setton, 2002);②若为口译教育者和口译学习者服务,则要注重语料库调用的功能模块设计,如在课堂上同时检索呈现不同译员停顿的位置、时长、前后语境等信息,同时还要注意与口译多媒体教学平台的兼容问题。

4.2 口译副语言信息的转写及标注方法

4.2.1 转写及标注工具

口译副语言信息的转写和标注与语言信息和超语言信息不同,需要使用专门的工具和软件。副语言信息通常的转写方法是用“...”、“-”、“*”、“p”等符号指代某类副语言特征,各类副语言信息夹杂于口译输出文本之中,语言信息与副语言信息相互交织。这种做法的优点是便于线性转写操作,缺点在于:1) 文内的标点需要去掉或作特别处理(以免与标注符号弄混),为此不得不使用额外的符号区分语段间隔;2) 较难处理多种副语言信息出现在同一时间节点的情况,也较难处理话语重叠等副语言现象;3) 不便于实现转写和标注的可视化操作。

关于副语言信息的转写及标注,目前已经有一些较为成熟的工具和软件可供利用,如 Anvil、EX-MARaLDA Partitur Editor、Praat 等,这些软件各有优势。以 Anvil 为例,该软件开发的初衷是为肢体语言研究服务,其操作界面如图 1 所示。

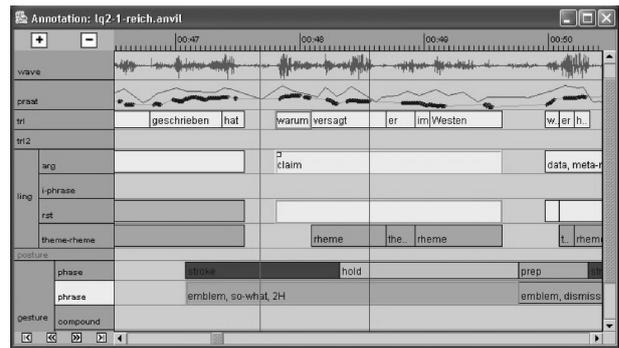


图1 Anvil 软件转写及标注界面

值得关注的是,Anvil 软件成功实现了副语言信息转写和标注的可视化操作。它允许对语言信息与副语言信息进行分层标注,不同类别的信息在不同的轨道(track)上进行标注,彼此之间互不干扰,而又通过线性时间轴相互联系。甚至不同类别的副语言信息(如肢体动作、语音高低长短、停顿等)还可进一步细分,在不同的轨道进行单独标注,这也避免了不同类别副语言信息在同一时间节点出现时不便标注的问题。讲话人的输出与口译员的输出也可各自占据一条轨道,因此讲话人与口译员话语重叠的问题也得到了解决。新轨道可以由转写标注者自行开辟,标注符号体系可以由转写标注者自行制定和导入,同时该软件还提供简单的数据统计分析功能,并且可以很方便地将转写标注好的语料导出成 XML 格式文件,因此也较好解决了与其它语料库的兼容性问题。

可惜的是该软件使用舒适度较低(Garg et al., 2004),而且支持的影音格式和输入语言有限。但这些都可以在未来通过对软件的不断更新进行完善,或者至少为口译副语言信息转写标注工具的研制提供了很好的思路 and 方向。现阶段口译副语言信息转写和标注可以依托现有工具可利用的功能,综合利用各个工具的长处。但未来大规模口译语料库的建设,还是有待于性能更加优良、更符合口译研究需要的副语言信息转写标注软件的研发。

4.2.2 转写及标注步骤

基于对上述问题的探讨,并根据自身建设口译语料库的经验,笔者总结了口译语料库副语言信息转写标注的步骤:

1) 明确建库目的,初步确定其未来应用领域,据此选定需要进行转写及标注的副语言信息类别(本文 4.1.1 建议的基本副语言特征应予标注);

2) 基于文本编码规范(TEI)、口语文本转写标注规范(HIAT)和会话分析(CA)领域的会话特征转写规范,编制符合当前口译语料库建设需要的副语言信息转写及标注符号体系(应尽可能使用现存规

范已有的标注符号);

3) 对口译影音语料进行头文件信息转写,要求尽可能多地涵盖该口译活动所涉及的超语言信息;

4) 根据第1)步所选取的副语言信息类别,设定转写标注软件(如 Anvil)中的转写标注轨道(track),有几类副语言信息就通过编写程序设定几个轨道;

5) 运用转写标注软件(如 Anvil),按照第2)步中所确定的副语言信息转写标注符号,对口译影音语料同时进行语言信息和副语言信息转写及标注,每个转写标注轨道对应一个类别的语言信息或副语言信息;

6) 从转写标注软件(如 Anvil)中导出已经转写标注好的语料的XML格式文件,并运用语料库建库工具(如 TEC Tools)建立口译语料库;

7) 运用语料库检索软件(如 BFSU ParaConc)以及相关统计分析软件(如 SPSS),基于所建立的口译语料库,开展相应的研究与教学工作。

4.2.3 后期建设

口译语料库的建设往往要在前期投入大量的时间和精力,但建库完成并不意味着建设工作的结束,后期建设同样要引起足够的重视。据笔者的经验,需要注意以下两方面的问题。首先,口译语料库的维护问题。前期建设过程中难免会出现纰漏,比如副语言信息标注位置错误、标注类别错误等,这就需要在语料库实际使用过程中不断发现问题,不断进行更正。有时建库者可能还要根据教学与研究需要,追加标注更多更为细化的副语言信息,这也是后期建设的重要工作。第二,口译语料库的扩充问题。口译语料库的建设是一个长期的过程,也是一个语料从少到多不断壮大的过程,因此后期语料规模扩大也是在建库之初就要考虑到的问题。有些建库者是长期依托团队力量,让每一届学生参与转写、标注等建库工作,这种情况下一定要注意副语言信息转写标注体系和方法的传承性。

5.0 结语

本文只是针对口译语料库副语言信息转写及标注问题的一项探索性研究。基于口译语料库开展口译研究的意义已经得到口译学界的广泛认同,但其应用前景尚待进一步拓展。现有研究多停留在使用词汇密度、词长、句长等书面语的参数研究口译语言特征等问题,对口译产品的口语体典型特征关注不够(王斌华,2012),对于口译特有的认知处理过程紧密相关的副语言信息关注不够。另外,开展语料库口译研究的前提是已经建设好经过一定程度加工的较高质量的口译语料库,而关于口译语料库建设的基础研究目前还相当欠缺。近年来,一些学者已

经开始关注口译语料库建设中的转写和标注问题,分享了各自建库方法和技术方面的经验,这对于后来者有着相当程度的参考和借鉴价值。诚如张威(2011:46)所言,“口译语料库的建设和相关研究也必将是口译教学与研究未来发展的一个核心”,或者更准确地说,至少在未来相当一段时期内,口译副语言信息的转写标注等口译语料库建设的基础类研究还是大有可为的,还需要更多研究者积极参与进来。

注释:

① 表1中语料库名称缩写的全称依次为:EPIC(European Parliament Interpreting Corpus); CIAIR(CIAIR Simultaneous Interpretation Corpus); CECIC(Chinese-English Conference Interpreting Corpus, 汉英会议口译语料库); PACCEL(Parallel Corpus of Chinese EFL Learners, 中国大学生英汉汉英口笔译语料库); DIRSI-C(Directionality in Simultaneous Interpreting Corpus); FOOTIE(Football in Europe, a corpus of press conferences of EURO 2008); CoSi/K6(Consecutive and Simultaneous Interpreting); ComIn-Dat(Community Interpreting Database Pilot Corpus); Dik/K2(Dolmetschen im Krankenhaus [Interpreting in Hospitals]); IiSCC(a corpus of interpreter-mediated interaction in New York Small Claims Court); TIC(Television Interpreting Corpus); CorIT(Italian Television Interpreting Corpus); FPC(Formula one grand prix Press Conferences)。

② 关于TEI的详细说明参见:<http://www.tei-c.org/index.xml>。文本编码的国际规范还有语料库编码标准(Corpus Encoding Standard, CES),但其在语音语料转写方面尚处探索阶段。

③ 关于HIAT的详细说明参见:http://www.exmaralda.org/hiat/en_index.html。

参考文献

- [1] Angermeyer, P. S. *Speak English or What? Codeswitching and Interpreter Use in New York Small Claims Court* [D]. New York University, 2006.
- [2] Angermeyer, P. S. et al. Sharing community interpreting corpora: A pilot study [A]. In T. Schmidt & K. Wörner (eds.). *Multilingual Corpora and Multilingual Corpus Analysis* [C]. Amsterdam/Philadelphia: John Benjamins, 2012. 275 - 294.
- [3] Bendazzoli, C. From international conferences to machine-readable corpora and back: An ethno-

- graphic approach to simultaneous interpreter-mediated communicative events [A]. In F. Straniero Sergio & C. Falbo (eds.) . *Breaking Ground in Corpus-based Interpreting Studies* [C]. Bern: Peter Lang ,2012. 91 – 118.
- [4] Bührig , K. et al. The corpus “Interpreting in hospitals”— Possible applications for research and communication trainings [A]. In T. Schmidt & K. Wörner (eds.) . *Multilingual Corpora and Multilingual Corpus Analysis* [C]. Amsterdam/Philadelphia: John Benjamins , 2012. 305 – 318.
- [5] Cencini , M. & G. Aston. Resurrecting the corp (us | se) : Towards an encoding standard for interpreting data [A]. In G. Garzone & M. Viezzi (eds.) . *Interpreting in the 21st Century — Challenges and Opportunities* [C]. Amsterdam/ Philadelphia: John Benjamins ,2002. 47 – 62.
- [6] Falbo , C. CorIT (Italian Television Interpreting Corpus) : Classification criteria [A]. In F. Straniero Sergio & C. Falbo (eds.) . *Breaking Ground in Corpus-based Interpreting Studies* [C]. Bern: Peter Lang ,2012. 155 – 186.
- [7] Garg , S. et al. Evaluation of Transcription and Annotation tools for a Multi-modal , Multi-party dialogue corpus [J/OL]. In *Proceedings of LREC 2004*. <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA459208>. [2013-06-30]
- [8] House , J. et al. CoSi – A Corpus of Consecutive and Simultaneous Interpreting [A]. In T. Schmidt & K. Wörner (eds.) . *Multilingual Corpora and Multilingual Corpus Analysis* [C]. Amsterdam/Philadelphia: John Benjamins , 2012. 295 – 304.
- [9] Monti , C. et al. Studying directionality in simultaneous interpreting through an electronic corpus: EPIC (European Parliament Interpreting Corpus) [J]. *Meta* ,2005 50(4) : 114 – 129.
- [10] Roach , P. et al. Transcription of prosodic and paralinguistic feature of emotional speech [J]. *Journal of the International Phonetic Association* , 1998 28(1 – 2) : 83 – 94.
- [11] Russo , M. et al. The European Parliament Interpreting Corpus (EPIC) : Implementation and developments [A]. In F. Straniero Sergio & C. Falbo (eds.) . *Breaking Ground in Corpus-based Interpreting Studies* [C]. Bern: Peter Lang , 2012. 53 – 90.
- [12] Sandrelli , A. Introducing FOOTIE (Football in Europe) : Simultaneous interpreting in football press conferences [A]. In F. Straniero Sergio & C. Falbo (eds.) . *Breaking Ground in Corpus-based Interpreting Studies* [C]. Bern: Peter Lang ,2012. 119 – 154.
- [13] Schiffrin , D. *Approaches to Discourse* [M]. Cambridge: Blackwell Publishers ,1994.
- [15] Straniero Sergio , F. Using corpus evidence to discover style in interpreters ’ performances [A]. In F. Straniero Sergio & C. Falbo (eds.) . *Breaking Ground in Corpus-based Interpreting Studies* [C]. Bern: Peter Lang , 2012. 211 – 230.
- [16] Tohyama , H. et al. CIAIR Simultaneous Interpretation Corpus [J/OL]. In *Proceedings of the O-COCOSDA 2004*. <http://ir.nul.nagoya-u.ac.jp/jspui/handle/2237/15081>. [2013-08-15]
- [17] Trager , G. Paralanguage: A first approximation [J]. *Studies in Linguistics* , 1958 ,13(1) : 1 – 12.
- [18] 陈瑞青 ,王巍巍. 口译中的副语言信息研究刍议 [J]. *外语艺术教育研究* 2011 (3) : 5 – 9.
- [19] 胡开宝 ,陶庆. 汉英会议口译语料库的创建与应用研究 [J]. *中国翻译* 2010 (5) : 49 – 56.
- [20] 李婧 ,李德超. 基于语料库的口译研究: 回顾与展望 [J]. *中国外语* 2010 (9) : 100 – 105 , 111.
- [21] 梁茂成. 副语言初论 [J]. *徐州师范学院学报* ,1994 (2) : 128 – 130.
- [22] 梁茂成 ,许家金. 双语语料库建设中元信息的添加和段落与句子的两极对齐 [J]. *中国外语* 2012 (11) : 37 – 42 , 63.
- [23] 王斌华. 语料库口译研究——口译产品研究方法的突破 [J]. *中国外语* ,2012 (3) : 94 – 100.
- [24] 文秋芳 ,王金铨. 中国大学生英汉汉英口笔译语料库 [M]. 北京: 外语教学与研究出版社 , 2008.
- [25] 张威. 口译语料库的开发与建设: 理论与实践的若干问题 [J]. *中国翻译* ,2009 (3) : 54 – 59.
- [26] 张威. 近十年来口译语料库研究现状及发展趋势 [J]. *浙江大学学报* 2011 (10) : 38 – 49.
- [27] 张威. 线性时间对齐转写: 口译语料库建设与研究中的应用分析 [J]. *外国语* ,2013 (2) : 76 – 83.