

# 语料库语言学的多维视角

## ——2011 中国语料库语言学大会综述

王艳伟

(上海交通大学,上海,200240)

**摘要:**本文对 2011 年 11 月 19 日至 20 日在北京外国语大学召开的“2011 中国语料库语言学大会”做了综述,简要介绍了主旨发言和各议题分组发言,以及“语料库语言学在中国”高层论坛的主要内容。文章指出了大会的主要特点以及呈现出的主要研究趋势和存在的一些问题。

**关键词:**2011 中国语料库语言学大会,语料库语言学

[中图分类号] H0-06

[文献标识码] D

[文章编号] 1674-8921-(2012)09-0071-02

由北京外国语大学中国外语教育研究中心主办,北京航空航天大学外国语学院协办的“2011 中国语料库语言学大会”于 2011 年 11 月 19 日至 20 日在北京外国语大学召开。这是中国语料库语言学研究会成立后的首届全国性学术研讨会。会议邀请了 4 位国内外知名专家做主旨发言,另有来自内地和香港的 85 所高校和科研院所及研发公司的 130 余位教学和科研工作者出席了会议并做了分组发言。11 月 20 日下午会议还组织了由五位语料库研究专家参与的“语料库语言学在中国”高层论坛。

### 1. 主旨发言

东京外国语大学的 Yukio Tono 教授做了题为“Linking CEFR to Learner Corpus Research: Identifying Criterial Features of L2 Development”的发言,他将语料库语言学与二语习得相结合,利用近邻聚类(variability-based neighbor clustering)和随机森林(Random Forest)的统计方法从学习者语料库中提取可供识别 CEFR(欧洲语言教学评估共同纲领)各语言水平等级的标准性特征(criterial features)。香港理工大学的 Martin Warren 教授的报告“Extending our Understanding of Phraseology”基于 John Sinclair 的“成语原则”,探讨和分析短语学的范围亦或短语倾向的内部运作机制。报告具体描述了利用 ConcGram1.0 软件从语料库中自动提取多元词语序列(n-grams),提出短语学变异(包括成分变异和位置变异)的分析框架,探讨和解释了一些重要的短语型式,并讨论了同现词语序列技术(concgramming)对于关涉(aboutness)和关键性(keyness)研究的启示。

卫乃兴教授在题为“Exploring Phraseology across

English and Chinese: Ideas and Methods”的发言中阐述了他所带领的研究团队基于平行语料库研究英汉对应短语序列的理念和方法,具体介绍了 John Sinclair(1996)的扩展意义单位模型以及将 Alternberg(1999)的相互对应率测量工具与 Tognini-Bonelli(2002)识别跨语言可比意义单位的三步分析法相结合的方法论,最后强调了语义选择趋向和语义韵在实现英汉对应词项不同程度的对等中所起的关键作用。许家金副教授的报告“Pear/Pair Stories: Maximising Comparability in CIA”尝试利用语料库方法调查外语学习者的母语在概念和语言学两个维度上对其英语口语叙事话语的影响。该研究利用 Chafe(1980)的“梨电影”(Pear Film)视频剪辑分步收集具有高度可比性的叙事语篇,并将这些梨故事(pear stories)进行不同向度的配对(pairing)来获得更多跨语言比较研究的机会,从而将 Sylvaine Granger(1996)首创的经典中介语对比分析法(CIA)朝着语际、语内、跨文类和跨话步的分析扩展和延伸。

### 2. 各议题分组发言

本届大会旨在加强国内语料库语言学界的学术交流与合作,探讨学术发展动态与关键技术以促进语料库语言学的学科建设和发展。会议将收到的论文摘要按 10% 的退拒率进行筛选,最终采纳了 87 篇论文并具体分为 9 个议题在四个分会场进行分组报告和讨论。

#### 2.1 语料库与语言描写研究

该类议题论文数占有所有议题论文数的 11.5%。其中实证研究占绝大多数,研究话题较为广泛,涉及词汇、型式语法研究、双语词典编纂、话语标记、对话行为和评价性语言研究;研究视角有词汇语法、词源学和协同语言学视角的研究视角;既有单语的研究,也有跨语言的研究。

**作者简介:**王艳伟,上海交通大学外国语学院在读博士。研究方向为语料库语言学。电子邮件:yanweizw@sjtu.edu.cn

## 2.2 语料库加工处理技术研究

该类议题论文只有4篇,在会议所有议题论文中所占比例最小,仅为4.6%,涉及的研究话题有基于句法依存关系的特定事件提取方法,用于多词法律术语检索的口语对话系统,平行语料库检索技术和基于海量生语料库的中文专家系统。

## 2.3 中介语语料库建设与研究<sup>①</sup>

该类议题论文共17篇,占总数的19.5%。其中关于中介语语料库建设的论文1篇,语料库在语言测试中的应用的论文1篇,语料库驱动的实证研究论文2篇,其他基于语料库的中介语研究论文12篇。研究话题涉及中介语使用(单个词、近义词、反义词、报道动词、虚化动词、立场副词、连接副词、被动结构、有灵主语)和习得(意愿与预测情态词、语言干涉)、口语(连接词、相邻时态默认、会话修补策略)和书面语的诸多方面。

## 2.4 平行语料库与翻译研究

该类议题的研究论文共18篇,占总数的20.7%。按研究内容可分为平行语料库建设及其关键技术(基于网络的平行语料库集成、平行词块抽取、双语句对齐、语言特异组合发现法)、语言对比(中介共性、被动句欧化特征、双语词汇对等项的建立、政治文本中的无主分句)和翻译研究(政治隐喻、词汇特征、显化、译文特征、译者风格)三大类。

## 2.5 汉语语料库建设与研究

该类议题论文所占比例较小,为8.0%。从研究话题来看,论及汉语语料库建设的论文远远少于基于汉语语料库的特定字词、短语、框式和从句研究,基于历时语料库的研究也远远少于基于共时语料库的研究,基于树库的研究更是远远少于基于普通语料库的研究。

## 2.6 多模态语料库建设与研究

该类议题在会议所有议题中所占比例也很小,仅为4.6%,具体的研究话题仅限于手势语和口语。

## 2.7 专门用途语料库研究<sup>②</sup>

该类议题论文共11篇,占总数的10.3%。与其他议题不同的是,该议题中讨论专门用途语料库建设和开发的研究居多(6篇),其中既有单语语料库(建筑英语、IT英语和航空航天英语),也有双语平行语料库(旅游英语和口译英语)和可比语料库(CROWN/CLOB),涉及的特定学科领域除了上述领域外,还有侧重于进行基于语料库的分析和对比的医学英语和民歌歌词。后者中既有技术词汇和高频词簇的分析,也有语篇分析、对比和翻译研究。

## 2.8 语料库与话语语用研究

该类议题基于话语建构现实的哲学观,论文数占有所有议题论文数的10.3%,内容涵盖通用话语的

研究(英语中的家庭关系、孔子的身份、中国印象)、媒体话语的研究(关于地震、反恐、圣诞节新闻报道的用词、主题词分析),以及特定话语类型的研究(话语标记、情态模糊限制语、适当性自我修正式语)。

## 2.9 语料库与语言教学

该类议题论文数占总数的10.3%。研究话题广泛,涉及语料库在语言教学众多领域的应用,如自建语料库在大学英语备课、教材评估、教材元话语建构和词汇教学中的应用,基于语料库的自助写作教学模式、会话教学和同义词辨析,还有平行语料库在旅游英语课程中的应用。

# 3. “语料库语言学在中国”

## 高层论坛内容梗概

11月20日下午,大会举办了“语料库语言学在中国”的高层论坛,由五位中国语料库语言学研究领域的领军专家做概括性发言,将本届语料库语言学大会推向了高潮。论坛上首先发言的是教育部语言文字应用研究所的冯志伟教授。冯教授谈了双语平行语料库的建设与应用,以及语言学对于计算语言学的理论支撑作用,最后呼吁计算语言学家应关注语言学研究。北京外国语大学中国外语教育研究中心的王克非教授围绕他主持的国家社科基金重大课题“大规模双语语料库的建设与加工”,给与会者分享了很多他在双语语料库建设的取样、研究思路及现存问题方面的思考。北京航空航天大学外国语学院卫乃兴教授就其所做的一系列研究和研究范式以及语料库研究的一些理论视角做了深入探讨。北京外国语大学中国外语教育研究中心的梁茂成教授谈了他对语料库研究中一些问题的思考,并介绍了北外语料库团队所做的一些研究项目。解放军外国语学院濮建忠教授讨论了语料库语言学的研究思路、研究领域和素质要求。

本届大会反映出国内语言研究的当前趋势是专门用途语料库、语料库应用于话语分析和外语教学的研究态势喜人,多媒体网络资源和技术、统计和自然语言处理技术等语料库相关技术的应用前景广阔。大会也反映出一些存在的问题,如研究多以书面语语料库为主,基于口语尤其是多模态语料库的研究相对薄弱;中介语语料库、平行语料库或翻译语料库的研究蒸蒸日上,而语料库加工处理技术研究和汉语语料库的建设和研究却相对滞后。大会也预示出国内语料库语言研究的未来走向:语料库研究将呈现多视角、多层次的特点。展现在我们面前的语料库语言学研究是一个可供学人们纵横驰骋的广阔领域。

(责任编辑 甄凤超)