

语料库语言学

外国语言学及应用语言学文摘

CORPUS

LINGUISTICS

RESEARCH

HIGHLIGHTS

北京外国语大学 中国外语教育研究中心 主办
外国语言研究所

2013年第2卷第1期

CORPUS
LINGUISTICS
RESEARCH
HIGHLIGHTS



外国语言学及应用语言学文摘 • 语料库语言学

主 办：北京外国语大学中国外语教育研究中心

主 编：梁茂成

电子邮箱：bfsucrg@sina.com

北京外国语大学
中国外语教育研究中心、外国语言研究所

外国语言学及应用语言学文摘

语料库语言学

(半年刊, 内部交流资料)

主 办: 北京外国语大学中国外语教育研究中心

主 编: 梁茂成

副 主 编: 李文中 许家金 熊文新

执行编辑: 许家金

文献收集: 吉 洁

封面设计: 吉 洁

编审委员会

文秋芳

王克非

梁茂成

李文中

许家金

熊文新

目 录

编者按.....	(1)
----------	-----

全文转载 (8 篇)

汉语复杂名词短语的独“的”偏好	
——整合语料库和认知语言学的证据.....	张立飞、严辰松 (2)
类型学概括与二语学习者汉语关系从句产出研究.....	李金满、吴芙芸 (9)
指量结构与汉语关系从句共现时的不对称分布及其原因	
——一项基于真实口语语料库的研究.....	盛亚南、吴芙芸 (22)
学术论文中篇章性句干的型式和功能研究.....	张 乐、卫乃兴 (30)
大学生英语现在完成时习得的语料库考察	
——检验“突显度”与“形式优先”假说的新证据.....	章柏成 (38)
中国英语学习者写作产出词汇的心理语言属性发展趋势	
——一项基于 WECCL 和 MRC 数据库的研究.....	李 海、张淑静 (43)
译者的意象图式与合成概念化	
——基于语料库方法的《红楼梦》“社会脸”翻译研究.....	谭业升 (51)
基于 R-gram 的语料库分析软件 PowerConc 的设计与开发.....	许家金、贾云龙 (56)

摘要转载 (20 篇)	(62)
-------------------	------

2013 年上半年国内刊发的语料库相关论文题录信息.....	(70)
--------------------------------	------

资讯

International Journal of Applied Linguistics 刊文介绍中国语料库研究.....	(83)
A tale of two C's: Comparing English varieties with Crown and CLOB.....	(86)
ToRCH 2009 平衡汉语语料库建成.....	(95)

全文转载论文英文提要.....	(96)
-----------------	------

Table of Contents

Editor's note: From data to theory, and from description to interpretation.....	(1)
--	-----

Research highlights (Part I): Full-text articles

<i>The one-de-preference of Chinese nominal phrases: Converging evidences from the corpus and cognitive linguistics.....</i>	ZHANG Lifei & YAN Chensong (2)
<i>Typological generalizations and Chinese relative clauses produced by second language learner.....</i>	LI Jinman & WU Fuyun (9)
<i>Demonstrative-classifier positioning in Chinese relative clauses and its underlying reasons: A spoken corpus study.....</i>	SHENG Yanan & WU Fuyun (22)
<i>Patterns and functions of textual sentence stems in research articles</i>	ZHANG Le & WEI Naixing (30)
<i>Undergraduates' acquisition of English present perfect tense: A corpus-based study – New evidences for the Saliency and Form-acquisition-priority Hypotheses</i>	ZHANG Baicheng (38)
<i>On the development of psycholinguistic word information in Chinese EFL learners' written discourse: A study based on WECCL and MRC database</i>	LI Hai & ZHANG Shujing (43)
<i>Translator's image schema and integrated conceptualization: A corpus-based study of "social face" in Hong Lou Meng and its translation.....</i>	TAN Yesheng (51)
<i>The design and development of the R-gram based corpus analysis tool 'PowerConc'</i>	XU Jiajin & JIA Yunlong (56)

Research highlights (Part II): Abstracts.....	(62)
--	------

A list of journal articles published in China in the first half of 2013.....	(70)
---	------

Short notices

<i>Corpus research activities in China covered by International Journal of Applied Linguistics.....</i>	(83)
<i>A tale of two C's: Comparing English varieties with Crown and CLOB.....</i>	(86)
<i>The release of ToRCH 2009 Chinese balanced corpus.....</i>	(95)

English abstracts of the full-text articles in <i>Corpus Linguistics Research Highlights</i>.....	(96)
--	------

编者按

从数据到理论，从描述到阐释

语料库研究素以量化描述见长，然而若止于量化，终于描述，语料库语言学这株苗圃新花，将注定是昙花一现，也必定营养不良，难以长盛。不少人主张应结合现有理论阐释量化数据。遗憾的是，这方面的成功案例少之又少。本期《语料库语言学文摘》（以下称《文摘》）所收 2013 上半年刊发的一些论文值得格外关注。无论是针对本族语或中介语的研究，都在数据的理论提升方面有诸多可圈可点之处。

本期选出的全文转载论文，多数都注重语料数据和现有语言理论的结合。比如，与认知语言学/功能语言学（张立飞、严辰松）、二语习得理论（章柏成）、篇章体裁分析（张乐、卫乃兴）、语言类型学视角（李金满、吴芙芸）、心理语言学研究（李海、张淑静）、社会认知（谭业升）等领域的相互关照。可见语料库语言学除去坚守自身理论内核外，还具有包纳其他语言学理论的开放性。语料库语言学与相关研究领域的互通，表明其具有广泛的适应性和扩展空间。

《文摘》全文收录论文的作者，多数并非专攻语料库语言学的学者，他们有自己的理论立场。这些学者的加入，使得语料库语言学研究丰满了许多。语料库语言学不再仅仅热衷于词块、主题词、搭配、语义韵这些传统议题。随着更多学者带着各自理论视角投入语料库研究，可以预见语料库语言学将会不断孳生新的研究增长点。

《文摘》收录论文中体现出的研究动向，借用研究方法相关的概念，即它们或多或少都采用了某种方式的“三角验证”（triangulation）。这种译法，多少有些生硬。同古汉语一样，此处的“三”是虚指，表达“多”的含义。因此，triangulation 称为“多重验证”或“多重论证”更合本义（有关 triangulation 的类型，可参看 Denzin 1970: 301）。诚然，《文摘》八篇选文中，很多更应看做量化研究和定性研究的混合型研究设计（参看 Hashemi 2012）。多重验证是为加强论证的信度，混合设计（尤其是“后继型混合设计”，sequential design）则更多是为拓展研究的深度和广度。

基于语料库的语言研究中，语言学理论的引入，有助于深化语言描写层次，挖掘现象背后机制的同时，还肩负着与相关领域进行理论对话的使命。比如，张立飞、严辰松文，利用平衡汉语语料库，就汉语中的复杂名词短语，特别是“的”字连用和单用的语言事实，对语法界权威观点（刘丹青 2008）提出质疑，修正了此前对独“的”复杂名词短语使用规律的描述，并从认知功能角度做出合理解释。在研究结论处，张立飞、严辰松认为“语法虽然有着功能或认知上的动因，并不是绝对的，而是呈现出一种概率性”。可见“计算机时代的田野调查者”（许家金 2011）很有必要与“坐而论道的语言学家（armchair linguist）”互学互辩。照此下去，语料库方法强于形式描写，弱于理论阐释的状况自会改观。

参考文献

- Denzin, N. 1970. *The Research Act: A Theoretical Introduction to Sociological Methods*. New Jersey: Transaction Publishers.
- Hashemi, M. 2012. Reflections on mixing methods in applied linguistics research. *Applied Linguistics*, 33(2): 206-212.
- 刘丹青, 2008, 汉语名词性短语的句法类型特征, 《中国语文》(1): 3-20.
- 许家金, 2011, 语料库语言学研究——计算机时代的“田野调查”, 《中国社会科学报》, 7月5日第14版。

汉语复杂名词短语的独“的”偏好 ——整合语料库和认知语言学的证据

张立飞, 严辰松

(解放军外国语学院 科研部 河南 洛阳 471003)

摘要: 对语料库的分析表明, 汉语的复杂名词短语有只使用一个“的”的偏好, 表现在: 包含多个定语的名词短语只带一个“的”的比例极高; 单个形容词(如“重要”)在复杂名词短语中不带“的”比例要远高于在普通名词短语中。复杂名词短语的独“的”偏好产生的认知机制是独“的”构式“D 的 H”由于固化程度高而被激活用来对包含多项定语的名词短语进行范畴化。复杂名词短语中“的”的隐现受象似性原则和经济原则这两个相互竞争的动因的制约: 前者追求形式和意义的一一对应, 倾向于保留所有的定语标记“的”, 后者寻求以最小的认知努力实现语义的表达, 倾向于只保留一个“的”。两者竞争的结果是在不影响语义表达的提前下尽量少用“的”。

关键词: 独“的”偏好; 名词短语; 象似性原则; 经济原则; 竞争的动因

中图分类号: H030 **文献标识码:** A **文章编号:** 1000-5544(2013)01-0030-07

Abstract: A corpus-based study indicates that Chinese nominal phrases with multiple determiners prefer only one determiner-marker *de*. The supporting evidences include: 1) In actual language use, the majority of nominal phrases carry only one *de* in places where more than one *des* are possible; 2) The adjective *zhongyao* is more likely to omit *de* in nominal phrases with multiple connotative determiners than in those with only one connotative determiner. The cognitive process behind Chinese nominal phrase's one-*de*-preference is the activation of highly entrenched schema D *de* H to categorize nominal phrases with multiple connotative determiners. The omission or presence of *de* in nominal phrases is determined by two competing motivations, i. e. the Principle of Iconicity and the Principle of Economy. The former stresses the one-to-one relationship between form and meaning and tends to allocate one *de* for every connotative determiner, while the latter seeks to achieve the communication goal with the least cognitive effort and thus hopes to keep only *de* for every nominal phrase. The result of the competition is a compromise of the sort, i. e. using as few *des* as possible while conveying the intended meaning.

Key words: one-*de*-preference; nominal phrases; principle of iconicity; principle of economy; competing motivations

1. 引言

众所周知, “的”作为汉语的定语标记并不是强制出现的, 有时可以隐去, 因此对于“的”在名词短语中隐现规律的探讨便成为学界的热门话题。众多学者纷纷就此问题提出自己的看法(如张敏 1998; 陆丙甫 2003; 刘永生 2004; 王光全、柳英绿 2006; 徐阳春 2008; 赵越 2009)。但有一类与“的”的隐现有关的语言现象未受到应用的重视, 即: 包含多个修饰成分, 且有可能使用两个或两个以上“的”的名词短语, 通常只带一个“的”。例如:

(1) 白莎丽独自来到了“夏威夷沙龙”, 她(的)这身大胆暴露的打扮, 果然令人为之侧目^①。

汉语名词短语中“的”的这种使用倾向, 可称之为“独‘的’偏好”(one-*de*-preference)。但是, 带多个定语的名词短语, 有时所有的定语都带“的”。例如:

(2) 我不愿意把它写成一部流水账似的庸俗的言情小说去骗取那些纯情少女们廉价的眼泪, 这是需要功夫的事

情。

那么, 为什么包含多项定语的名词短语会倾向于只使用一个“的”? 其中“的”的隐现是有规可循, 还是完全偶然的? 本文试着对上述问题做出解答: 首先, 概述前人有关名词短语以及其中“的”的隐现的探讨; 其次, 借助语料库的证据来证实名词短语的独“的”偏好; 再次, 指出独“的”偏好是由于构式“D 的 H”固化程度高, 容易被激活用来对包含多项定语的名词短语进行范畴化而造成的; 最后, 提出名词短语中“的”的隐现是象似性动因和经济动因相互竞争的结果。

2. 定语的分类与“的”的隐现

2.1 定语的内涵和外延之分

汉语中可做定语的成分极多, 如名词、动词、形容词、区别词、介词短语、定语从句、指示词以及数量词等, 且它们做定语时都置于核心名词之前且缺少形态变化, 这就使得核心名词之前的位置格外“拥挤”。例如:

(3) [我][那件][昨天下午买的][很厚的][棉麻]衬衫。(刘丹青 2008: 4)

于是,许多学者都试图对定语进行分类,希望找到不同定语形式和意义之间的对应关系。朱德熙(1956)根据意义将定语分为限制性的和描写性的。前者用来对事物进行限制,并作为其分类的依据,如“白纸”中的“白”;后者则用来描写事物的状况或情态,如“雪白的纸”中的“雪白的”。之后,朱德熙(1982)又从形式上区分了组合式定语(包括领属性定语、数量定语和所有带“的”的定语)和粘合式定语(其余各种不带“的”的定语)。朱先生分别从形式和意义上对定语进行了分类,但“两大意义分野同两大形式分野之间显然没有对应关系”(陆丙甫 1988: 102)。陆丙甫(1988)提出:粘合式定语是称谓性的,构成某种事物的称谓方式;组合式定语是非称谓性的,指名词所指事物具有修饰语所指的性状特征。组合式定语又可从意义上区分出限制性的和描写性的。定语意义上的称谓性和非称谓性的区分,不仅使粘合式和组合式这种形式对立有了更坚实的基础,也避免了把定语简单地分为限制性和描写性(特别是把粘合式定语也归入描写性或限制性的)所导致的复杂化(详见陆丙甫 1988: 103-104)。陆的做法虽说实现了定语形式和意义的对应关系,为分析定语引入了清晰的思路,但由于过于追求形式和意义的对应,把所有不带“的”的定语(不包括领属性和数量定语)都纳入称谓性定语的范畴,为充分解释某些不带“的”的定语带来了困难。例如:

(4) 较整体车流行驶速度低的汽车,给跟随其后的车辆行驶带来很大麻烦。

例(4)中,下画线部分名词短语的结构应该是“[[跟随其后的车辆][的]行驶]”。若按照陆的观点,“车辆行驶”由于不带“的”,应该是称谓性的,构成某种事物的称谓方式。但“车辆行驶”很难说是某个特定概念的称谓。而且,整个名词短语的结构会因此变成“[[跟随其后的]车辆行驶]”,语义上很难说得通。在语言的实际使用中,这样的例子大量存在,如“不同的所有制界限、巍峨的石坝脚下、劳动者提供的劳动数量、第一个跨越海峡进京参赛的大型体育代表团成员”等。

刘丹青(2008: 3-4)没有刻意追求形式和意义的工整对应,而是从语义上将定语分为内涵定语(connotative determiner)和外延定语(denotative determiner)两大类^②:

内涵定语由实词性/开放性语类充当,给整个名词增加词汇性语义要素(即内涵),包括描写性和限制性定语,由名词、区别词、形容词、动词、介词短语及定语从句等充当。如“木头桌子、慢性疾病、吃的点心、修车工具、他买的书、看门的老头”等。……外延定语由指称和量化成分充当,用来给名词语赋以指称、量化属性,表明它在真实世界或可能世界中的具体所指范围,即在不改变内涵的情况下指明其外延,由指示词、冠词、数量词语、量化词语(全量或分量词语)充当。如“这人、那支笔、三本书、所有人、一些商店、上述人

员”等。

定语的内涵、外延之分是语义上的,也是形式上的:一方面,两类定语各自由特定的词类来充当;另一方面,两类定语与定语标记“的”有着不同的对应关系。“的”可以看成是内涵定语和外延定语的基本界限,内涵定语都可以用(但不一定必须)带“的”,而外延定语不能带“的”(刘丹青 2008)。

我们赞同刘丹青对于定语的分类体系。这样,定语不带“的”的问题就简化为内涵定语带不带“的”的问题,使研究的对象更加明确^③。

2.2 概念距离与“的”的隐现

对于定语“的”的隐现规律,许多学者都做过研究,但仍未有一致看法。难点在于定语的语类或简单的句法规则不足以预测“的”的出现与否,因为作为定语标记的“的”并不是纯句法性的,它的隐现受句法、语用、韵律等条件的制约(刘丹青 2008)。众多研究中,最具代表性的、被普遍接受的便是张敏(1998)提出的“概念距离”说。

张敏(1998)借用 Haiman(1985)提出的“概念距离”的概念来解释定中结构中“的”的隐现:从构成成分的语言距离来看,不带“的”的定中结构中定语与中心语之间的距离要比带“的”的定中结构中的更近,因此它们之间的概念距离也更近。概念距离的远近等同于概念整合程度的高低,或概念联系的松紧。张的“概念距离”说能够较好的解释定中结构中“的”的隐现,且符合我们的语感。例如:

(5) a. 白纸热水

b. 白的纸热的水

从形容词与中心语之间的形式距离看,例(5a)小于(5b)。语言距离的远近对应定语与中心语结合紧密程度的松散:例(5a)的定语与中心语之间不能插入数量成分,如“*白一张纸”;例(5b)则可以,如“白的一张纸”。最后,从概念距离来看,例(5a)的“热水”表达的是单个概念的名称,概念“热”作为分类的属性与概念“水”整合在一起,两者概念距离很近;例(5b)“热的水”表达的不是一个整体性的概念,“热”只是对“水”的属性进行说明。

利用“概念距离”来解释定中结构中“的”的隐现不仅符合我们的语感,解决语言使用中的实际问题,也能很好地与其他学者的观点兼容。上文提到,陆丙甫(1988)曾用“称谓性”来解释粘合式(不带“的”)的定中结构,而称谓性又与较近的概念距离相对应。两个语言成分之间的概念距离越近,它们的关系就越紧密,越容易为人们所充分了解,其心智中的固化程度(degree of entrenchment)就越高,因此容易作为一个整体概念被激活和提取,自然语言也就倾向于用专门的、固定的名称对其进行称谓。

3. 名词短语的独“的”偏好

3.1 语料库的证据

3.1.1 方法和步骤 下面利用语库方法,通过定量分

析来证实汉语名词短语的独“的”偏好。基本思路是:分析那些包含两个或两个以上定语的名词短语,对比其中只出现一个“的”的和出现两个或两个以上“的”的。若前者数量远大于后者,则说明汉语名词短语的独“的”偏好确实存在。反之,独“的”偏好应为谬误。有两点需要补充:1)由于“这”、“那”、“三本”等外延定语不能带“的”,不存在“的”的隐现的问题,因此不做考虑;2)对于只包含一个定语的名词短语中“的”的隐现情况,由于前人的论述已比较充分,且无法确定关键词来对语料库进行检索,此处也不予考虑。

本研究选用由肖忠华编制的兰开斯特汉语语料库(The Lancaster Corpus of Mandarin Chinese,简称 LCMC)^④。对语料库的检索和对语料的处理按以下步骤进行:

1)以“的”为关键词对 LCMC 语料库进行检索,得到所有包含“的”索引行。

2)对索引行进行筛选,剔除那些“的”字为非定语标记的索引行。

3)对索引行中包含定语标记“的”的名词短语进行手工标注。标注分以下几个层次:

A.名词短语所带定语(包括带“的”和不带“的”)的数量^⑤。

B.对于复杂名词短语,判断定语的结构关系,即是多层定语(定语本身是内含定语的名词短语)还是多项定语(几个定语依次修饰核心名词)。为避免使问题复杂化,本研究只考查多项定语。

C.对于复杂名词短语,标注每个定语:

- ①是否带“的”;
- ②是何语法类别(名词性<N>、动词性<V>、形容词性<A>、介词性<P>、从句<C>);
- ③相对于核心名词的位置(以核心名词为起点,从右至左依次为 P1、P2、P3……)。

4)利用统计软件 SPSS 对标注后的索引行做统计分析。

3.1.2 统计结果和分析 按上述方法和步骤,我们以“的”为关键词以语料库进行检索,共得到包含“的”的检索行 50852 例,并随机抽取了 1/5 共计 10186 例作为考查对象。10186 个用例中,除去 867 例非定语标记和多层定语标记的用例,其余用例中“的”的总体分布如下表 1 所示:

表 1. 定语标记“的”的总体分布

	普通名词短语	带两个定语的名词短语	带三个定语的名词短语	总计
数量	8106	1154	59	9319
百分比	87.0%	12.4%	0.6%	100%

表 1 说明,作定语标记时,“的”绝大多数出现在带一个定语的名词短语中,占 87%,出现在复杂名词短语中的比例只有 13%。当名词短语包含两个或两个以上定语时,有可能所有定语都带“的”,也有可能只有部分定语带“的”。例如:

(6) 贝多芬音乐中那雄浑的低音,巨大而辉煌的交响

轰鸣,以及从中透出的金属般的光辉、英雄气概、神圣的激情和晚年的沉思,是代表了整个人类最纯正的伟大情感。

(7) 认清经营类别及如何经营,是建立商店特有形象的重要(的)条件,进而去掌握特色,表现个性。

表 2. 包含两个内涵定语的名词短语中“的”的隐现

	带一个“的”	带两个“的”	总计
数量	863	291	1154
百分比	74.8%	25.2%	100%

从表 2 可以看出,当名词短语包含两个内涵定语,即有可能带两个“的”时,大多数情况下(78.2%)只带一个“的”,只有少数情况下(25.2%)才带两个“的”。

再看名词短语包含三个内涵定语的情况。

表 3. 包含三个内涵定语的名词短语中“的”的隐现

	带一个“的”	带两个“的”	带三个“的”	总计
数量	12	34	13	59
百分比	20.4%	57.6%	22.0%	100%

与包含两个内涵定语的名词短语不同,包含三个内涵定语的名词短语多数情况下(57.6%)带两个“的”,只带一个“的”和三个“的”都带的比例大致相等(分别为 20.4%和 22.0%)。综合考虑两个和三个内涵定语的情况,名词短语中“的”的隐现可用下表 4 来描述:

表 4. 复杂名词短语中“的”的隐现

	带一个“的”	带两个或两个以上“的”	总计
数量	875	338	1213
百分比	72.1%	27.9%	100%

由表 4 可知,有可能带两个或两个以上“的”的复杂名词短语,多数情况下(72.1%)只带一个“的”,其余“的”则被隐去。这说明汉语的复杂名词短语倾向于只带一个“的”,此即独“的”偏好。

名词短语的独“的”偏好在语感上体现为“的”字用多了会给人以繁冗、啰嗦、单调之感。诚然,“的”作为汉语中唯一的定语标记,对于调节成分间的节奏、明晰修饰语与中心语之间的语义关系至关重要,但“若是不加节制,出现太频,则不但听起来琐碎,看来纷繁,而且可能扰乱了文意”(余光中 1996: 183)。“的”字泛滥往往出现在英译汉作品中,这是因为英语中有多种结构对应“的”的功能,如形容词、动名词、定语从句、介词词组等,而且形容词本身词尾也有多种变化,如 childish、mathematic、ladylike、intellectual、monotonous、friendly。若在翻译时一律用“的”来处理的话,就会造成“‘的’到底,不肯罢休”的情况。黄忠谦、许萍(1997)通过对二十余篇译作的调查,发现汉译偏正结构时,无论定语有多长,其“的”字使用数不得超过 3 2 个正常,1 个最好。刘丹青(2008)敏锐地观察到,多层定语和多项定语都有排斥“的”在一个名词短语内多次出现的倾向。

3.2 语境因素与“的”的隐现

下面以带两个定语的名词短语为例,分析何种语境下“的”最容易被省去,主要从两个方面进行探讨:一是定语的

语法类别,即是名词性的、动词性的、形容词性的、介词性的、还是从句;二是定语相对核心名词的位置,是距离核心名词较近(P1)还是较远(P2)。各类定语在不同位置上时“的”的隐现如下表5所示:

表5. 各类定语在不同位置上时“的”的隐现

类别	位置	数量	带“的”	不带“的”
形容词性	P2	147	136	11
	P1	588	141	447
	总计	735	277	458
名词性	P2	565	491	74
	P1	290	92	198
	总计	855	583	272
动词性	P2	89	87	2
	P1	174	45	129
	总计	263	132	131
介词短语	P2	50	47	3
	P1	9	9	0
	总计	59	56	3
从句	P2	303	302	1
	P1	93	93	0
	总计	396	395	1

先只看定语类别,而不考虑其所处的位置。在语料库中,形容词性定语共出现735次,其中带“的”的277次,占37.7%,不带“的”的458次,占62.3%;名词性定语共出现855次,其中带“的”的583次,占68.2%,不带“的”的272次,占31.8%;动词性定语共出现263次,其中带“的”的132次,不带“的”的131次,两者比例相当,各占50%;介词短语做定语共59次,带“的”的56次,占94.9%,不带“的”的3次,仅占5.1%;定语从句共出现396次,带“的”的395次,占99.7%,不带“的”仅有1次,占0.3%。从“的”字省略的难易程度来看,五类定语形成一个等级:形容词性>动词性>名词性>介词短语>从句。其中,形容词性和动词性定语不带“的”的比例都在50%上,说明它们较容易隐去“的”;名词性定语只有31.8%不带“的”,“的”字省略较难;介词短语和从句只有极个别的可以不带“的”,“的”字最难隐去。形容词性、动词性和名词性定语“的”字省略难易程度的不同可以从它们本身语义结构的特点进行阐释。按照Langacker(1987)的观点,动词和形容词表征的是事物间的关系,是概念依赖的,对它们语义结构的描述必须借助相关事物的概念;而名词表征的是事物,是概念独立的,对其语义结构的描述可相对独立地进行,无须借助相关事物的概念。既然从语义结构上看动词和形容词对名词有依赖性,那么它们与名词之间的语义关联就更紧密,按照“距离象似性”原则,“的”字也就更容易省略,而名词是概念独立的,它做修饰语时,对中心名词的依赖程度相对较低,因而两者概念距离较远,因此需要保留“的”来明晰两个名词间的修饰关系,“的”字也就较难省略。

介词短语和从句做定语时,“的”字也可隐去,这有点出乎意料,因为一般认为介词短语和从句做定语时必须带“的”。下面举例详细分析:

(8) 在我国社会主义现代化建设的关键时期,小平同志的到来,是对深圳特区(的)最大的关怀和支持,是对深圳人民最大的鼓舞和鞭策。

(9) 昔日在全国显赫一时(的)辽宁引为自豪的丹东手表、金笔、营口幸福牌钢琴、友谊牌洗衣机、沈阳的机床等,如今都坠入市场萎缩的困境。

例(8)和例(9)的一个共同点就是省略“的”字的介词短语和从句都位于距中心名词较远的P2位置。这样,位于P1位置的带“的”的修饰语相当于一个缓冲,避免介词短语或从句直接与中心名词直接接触从而导致修饰关系不可解读。如果去掉P1位置的定语,“对深圳特区关怀和支持”和“昔日在全国显赫一时丹东手表”,一个意义有变化,一个不合语法。

再看定语相对于中心名词的位置。总体上看,P1位置上省略“的”的比例(67.1%)远大于P2位置(19.1%)。这说明距离中心名词越近,越容易省略“的”,即“的”字有“蒙后略”的倾向。这与刘丹青(2008:11)的观点正好相反,他指出,多层定语和多项定语排斥“的”字多次出现的共同做法是“尽前省略”,即位置在前的定语,尤其是在定语之首的领属语,最容易或最需要省略“的”。“的”字“蒙后略”的倾向还具体体现在形容词性、名词性和动词性这三类定语上:形容词性定语在P1位置省略“的”的比例为76.0%,名词性定语为68.3%,动词性定语为74.1%,而它们在P2位置省略“的”的比例分别为7.5%、13.1%、2.2%,均远远小于在P1位置。

3.3 个案分析:重要

上文,我们借助语料库的证据,证明汉语的复杂名词短语构成一个特殊的上下文语境,此语境下名词短语倾向于只保留一个定语标记“的”。这一论断还须排除一种可能性:倘若只带一个定语的普通名词短语,其隐去“的”的概率与带两个或两个以上定语的复杂名词短语一样的话,那么后者就不存在任何特殊之处,其中“的”省略只是“的”在一般情况下省略规律的自然延续。为证伪这一可能性,我们需要证明:定语在复杂名词短语中隐去“的”的比例要高于普通名词短语。为此,我们以形容词“重要”为例进行个案分析,考查其在普通名词短语和复杂名词短语中“的”的隐现规律,具体步骤是:以“重要”为关键词对LCMC进行检索,筛选出“重要”做定语的索引行,并进行标注,记录两个方面的信息,即“重要”是否带“的”以及“重要”所在的名词短语带一个还是多个定语。具体统计结果如下表6所示:

表6. “重要”做定语时“的”的隐现

类别	带“的”	不带“的”	总计
普通名词短语	106	94	200
	53%	47%	100%
复杂名词短语	28	201	229
	12.2%	87.8%	100%

表6显示,“重要”做定语时在两类名词短语上的分布

大致相同 在普通名词短语中出现 200 次 在复杂名词短语中出现 229 次。例如:

(10) 这段时期,他不仅为中共党的建设、共产国际运动做出重要(的) 贡献,而且,在中文拉丁化上、语言理论上也取得令人惊喜的成果。

(11) 第四,生育家庭,即一个人结婚、生育后组建的家庭,它标志着一个新的独立消费单位的出现,会相应产生许多新的消费需要,往往会成为消费品的重要(的) 购买者。

但就“的”的隐现而言,“重要”在这两类名词短语中的表现则差异明显:在普通名词短语中,“重要”带“的”与不带“的”比例相差不大,分别为 53% 和 47%,“的”的隐现有一定的任意性;但在复杂名词短语中,“重要”带“的”与不带“的”比例则分别为 12.2% 和 87.8% 说明在包含多项定语复杂名词短语中,“重要”强烈倾向于不带“的”与只包含一个定语的名词短语的差别显而易见。可见,带多个定语的复杂名词短语确实构成一个特殊的上下文语境,该语境下名词短语倾向于只保留一个定语标记“的”。

4. 独“的”偏好的认知阐释

我们认为 汉语名词短语独“的”偏好的产生是典型的定中构式“D(定语)的 H(中心名词)”被激活用来对复杂名词短语进行范畴化的结果。认知语言学秉承基于使用的语言观(usage-based model of language),认为语言使用建构语言知识(严辰松 2010)。人们对实际语言的体验中,利用范畴化等基本认知能力,对体验到的语言实例进行认知处理,形成一系列抽象和复杂程度不一、形式和意义兼具、在心目中得到表征的实体或图式构式(张立飞 2010)。Langacker(1987)和 Bybee(2010)都把具体的语言使用看成是范畴化的过程,其间存于心智的图式被激活,或对感知到的语言形式做出语义解读,或根据想要表达的意义产出适当的形式。Bybee(2010)认为有两个因素决定心智中的哪个图式被激活对使用中的语言进行范畴化:一是图式的固化程度;二是图式与实际语言表达在形式和意义上的相似程度。其中,图式的固化程度取决于该图式所例示的实际语言表达的使用频率。语言的每一次使用都会在心智中留下印迹,使得下次提取变得更加容易。潜藏于语言实例中的图式会在频率的作用下固化程度不断加深,从而变得更加容易激活。使用事件的范畴化有两种结果:一是完全识别(full recognition)即图式与实际表达完全吻合;二是部分识别(partial recognition)即图式与实际表达只是在形式或意义上部分吻合(严辰松 2008)。

再来看“的”的分布情况。表 1 显示,“的”出现在普通名词短语中的次数为 8106,占其总出现频数的 87%,这说明独“的”构式“D 的 H”的固化程度要远大于带两个或两个以上“的”的构式“D 的 D 的 H”或“D 的 D 的 D 的 H”。因此,在实际语言使用中,构式“D 的 H”会被优先激活来对名词短语进行范畴化。若名词短语只带一个定语,则是完全识

别;若名词短语带多个定语,则是部分识别。当“D 的 H”被用来对带多个定语的复杂名词短语进行范畴化时,有两种选择:一是在碰到首个定语时便请出“的”。为了与范畴化图式保持形式上的一致,其余位置上的“的”便被隐去。此时,若其余位置的定语与中心名词语义关系密切,可构成一个概念整体,名词短语则被完全识别;若两者概念距离较远,名词短语则在语义上被部分识别。后一种情况会催生一些语义不通的名词概念。例如:

(12) 但是,由于美国和南朝鲜联合举行敌视北方的“协作精神”军事演习,破坏了通过对话形成的缓和(的) 气氛,造成了对话的中断。

上例中“缓和气氛”并不是一个耳熟能详的名词概念,它的产生是构式“D 的 H”由于固定较高被优先激活用来对名词概念“通过对话形成的缓和的气氛”进行范畴化,后者为了满足其结构上的特点而隐去后一个定语的“的”而造成的。

独“的”偏好的“蒙后略”策略甚至还会作用在包含多层定语的名词短语上。例如:

(13) 此种观点认为,妻管严的大量出现,是整个社会的男性(的)进取心和创造力大幅度下降的一种标志,它不仅严重挫伤了男性的尊严,而且使丈夫们不得不以诸如攒“私房钱”等形式来加以消极反抗。

例(13)中,下画线部分原本应是一个多层定语,其结构为[[[整个社会的]男性]的]进取心]。从语义上看,“男性”是“整个社会的”修饰的对象,它们之间的语义关系显然更为密切。省略它们之间的“的”所形成的“整个社会男性的进取心”显然更加合理,既不破坏成分间的语义关系,又符合“距离象似性”原则。“蒙后略”策略形成的“整个社会的男性进取心”不仅混乱了成分间的语义关系,而且从概念上说,“男性进取心”本身就经不起推敲,不能作为一个概念整体被看成是“进取心”的一个类别。

二是保留 P1 位置的“的”以突显修饰语和中心语之间的界线。例如:

(14) 石敢瘦小的身材叫乔光朴(的) 魁伟的体架一衬,就像大人拉着一个孩子。

需要注意的是,承前省略“的”的定语中,绝大部分表达的是领属关系(如例(14)所示),这或许是刘丹青(2008: 11)认为“定语之首的领属语,最容易或最需要省略‘的’”的原因。

此外,刘丹青(2008)指出补上“的”后的名词短语会显得重复、啰嗦,而这并非句法语病,只是语用瑕疵。我们认为,所谓的“语用瑕疵”本质上还是一种句法语病。语感上的重复、啰嗦源自实际语言表达与被激活的范畴化图式之间的不完全匹配。听话人会觉得“王主任的年纪很轻的在湖北老家经商的远房亲戚”之类的表达很啰嗦是因为他所接触到的名词短语大多数都只带一个“的”^⑥。

5. 动因竞争与“的”的隐现

至此 我们还需要回答为什么汉语名词短语的独“的”偏好只是语言使用中的一种倾向,而非不可违逆的句法规则?即,为什么并不是所有复杂名词短语都只保留一个“的”?为什么带三个定语的名词短语“的”的省略要比带两个定语的名词短语困难(比较表2和表3)?我们认为,这些看似任意的现象实际是语言使用中的象似性原则(Principle of Iconicity)和经济原则(Principle of Economy)这两种动因相互竞争的结果。

语言符号的象似性原则基于这样一种认识,即语言结构从某种程度上反映了经验的结构。因此,语言的象似性可定义为语言表达在外形、长度、复杂性以及构成成分之间的关系上平行于它所表达的概念、经验或交际策略(张敏 1998: 143)。语言的象似性有多种体现。凯撒大帝的名言“veni, vidi, vici”体现了语序象似性,即:若其他一切条件相同,那么子句在叙述中的顺序一定和它们所描述的事件的次序相同。还有一种象似性动因就是同构(isomorphism),即语言符号能指和所指之间的一一对应。从某种意义上说,语言符号就像是示意图(diagram),语言结构的构成成分之间的关系以示意图的方式反映它所表达的意义或概念结构的元素之间的关系(Haiman 1985)。

语言使用的经济原则要求说话人利用最小的认知努力获取最大的交际效果,它最直接的体现就是语言形式的简化。但简化的形式并不意味着概念简化,许多语言中都存在概念复杂但形式简单语言单位。很多时候,语言表达形式上的复杂程度反映的是使用者对它的熟悉程度:使用频率越高、使用者越熟悉的语言表达,其形式就越简化,意义也越隐晦。英语的一些缩略结构,如I'm, I'll, can't, don't, won't等都是在经济原则的驱动下、由于较高的使用频率而产生的。

象似性原则和经济原则是一对“天生的敌人”:前者强调形式和意义的对应关系,用简单的形式表达简单的概念,用复杂的形式表达复杂的概念;而后者则追求用简单的形式表达复杂的概念。象似性原则和经济原则相互竞争,并且各有胜负,从而造成语言的使用带有一定任意性的假象,其中最常见的例子就是反身代词的使用或删略。例如:

(15) ? a. He shaved himself.

b. He shaved \emptyset .

(16) a. He killed himself.

* b. He killed \emptyset .

例(15)中的shave属于内向动词(introverted verb),其语义结构决定动作的受事与施事同指,这就使其动作的受事(即宾语)具有高度的可预测性,此时例(15b)隐去做宾语的反身代词并不会影响语义的表达,因而经济原则胜出。反而是补上反身代词的例(15a)让人感觉有些怪异,使人联想到“他在车祸中失去了双手,但现在居然能自己刮胡子”

之类的语用蕴涵。例(16)中的kill则属于外向动词(extroverted verb),其语义结构虽然也包含一个受事,但该受事通常不与施事同指,因此为了明确语义,例(16a)需要依照象似性原则将动词的概念结构在形式上体现出来,而受经济原则驱动的例(16b)则语义不明。

再回过头来看带多个定语的名词短语中“的”的隐现。激活构式固化程度高的、只包含一个“的”的构式“D的H”来对名词短语进行范畴化显然是受经济原则的驱使,因为固化程度高的构式容易被激活,需付出的认知努力较少。但经济原则会使语言形式与意义之间的象似性受到磨损,隐去定语标记“的”会使得定语与中心名词之间的概念上修饰关系无法在形式上得到体现,从而无法保持形式与意义之间的同构关系,语义变得不透明,增加听话人解读的难度。相反,保留所有的定语标记“的”则可最大程度地保证名词短语形式与意义之间的一一对应关系(即同构),使其语义变得透明,容易解读,但会加重说话人的认知处理负担。两种动因竞争的结果是,在不影响语义表达的情况下尽量少用“的”,证据有:

1) 带两个定语的名词短语由于内部语义关系(主要是定语对中心名词的修饰关系)并不特别复杂,受经济原则的驱使倾向于只保留一个“的”;

2) 带两个定语的名词短语也有一部分有两个“的”,使得形式成分间的内部关系与语义成分间的内部关系保持对应,即形式与意义同构,这是象似性动因使然。另外,在291例带两个“的”的名词短语中,有57例(19.6%)两个定语均为介词短语或从句,142例(48.8%)有一个定语为介词短语或从句。例如:

(17) 随即,她和小赵一起又采摘了一朵裹得很紧的大约二天之后才能开的花蕾。

(18) 1937年在南京时,曾在国民党电台里当过广播员,在青年战地服务训练班与国民党教官关系暧昧,还参加过国民党三青团举办的一个党一个主义一个领袖的讲演比赛。

上文谈到,介词短语和从句做定语时,“的”很难被省略,而两个“的”都保留的名词短语中,接近70%至少有一个定语为介词短语或从句。这再次说明,一般来讲,还两个定语的名词短语倾向于只保留一个“的”,而只有当省略“的”会影响语义的表达时(特别是定语为介词短语或从句时),才会两个“的”都保留,维持形式和意义的同构。

3) 带三个定语的名词短语一方面只带一个“的”的比例大幅减少(比较表2和表3),另一方面带两个“的”的比例占绝大多数。这是因为随着定语的增多,名词短语内部的语义关系也变得更为复杂,保持语义透明性,使形式和意义同构的压力也增大,使得省略“的”的难度加大。在象似性原则和经济原则的双重作用下,说话人对“的”的省略更多地采用折衷的做法,即“去一留二”,既达到了省力的目的,又不至于影响语义的表达。

6. 结语

实际言语交际中,语言表达的选择虽然有着认知以及功能的动因,但同时也受语境以及背景知识的制约,而且有些情况下认知或功能上的动因过于抽象、宽泛,无法对语言表达的选择产生立竿见影的效果。这使得真实交际场景中说话人对于语言表达的选择难以预测,存在变数。因此,语法虽然有着功能或认知上的动因,并不是绝对的,而是呈现出一种概率性。语法的规律性有时并不能从单个的语言表达中观察得到,只有对大量语言实例进行全盘考察,借用频率对语言事实进行描述,语法的规律性才能显现出来,其背后的功能或认知动因才能被揭示(DuBios 1985)。汉语复杂名词短语的独“的”偏好作为一种语法现象,并不体现为非此即彼的强制性规则,而是揭示了语言使用中的一种倾向,通过对频率的观照才能显现,较好地诠释了语法的概率性本质。

汉语名词短语中“的”的隐现受象似性原则和经济原则这两个相互竞争的动因的制约。象似性原则追求形式和意义的一一对应,倾向于保留名词短语中所有的定语标记“的”;经济原则寻求以最小的认知努力实现语义的表达,倾向于只保留一个“的”。竞争的结果是在不影响语义表达的情况下尽量少用“的”,故而造成了名词短语的独“的”偏好。本文论证充分说明,“语言因为交际的目的而存在,语言的结构也是在交际使用中磨合成形的”(张伯江 2005: 23),也即语言使用建构语言知识。

* 本文初稿曾在“第11届国际认知语言学大会”(西安外国语大学 2011年7月11日-17日)和“2011中国语料库语言学大会”(北京外国语大学 2011年11月19日-20日)上宣读,在此对与会专家提出的宝贵意见表示感谢!文中错误概由作者负责。

注释:

- ① 如无特别说明,文中例句均来自兰开斯特汉语语料库。例句括号中的“的”为有可能出现的“的”,由笔者添加,下同。
- ② 内涵定语和外延定语的概念源自陆丙甫(1988),但刘丹青(2008)对这两个术语的使用显然与原作者不同。陆把内涵定语和外延定语看成是非称谓性定语的下位概念。
- ③ 为表述方便,如无特别说明,下文“定语”一律指“内涵定语”,“普通名词短语”指只包含一个内涵定语的名词短语,“复杂名词短语”指包含两个或两个以上定语的名词短语。
- ④ 对该语料库的详细介绍请参阅网址: <http://www.lancs.ac.uk/fass/projects/corpus/LCMC>。
- ⑤ 由于我们是以“的”为关键词对语料库进行检索,因此检索行中只包含一个定语的名词短语必定带“的”。
- ⑥ 除了隐去“的”以外,汉语复杂名词短语保持其独“的”偏好的另一做法是利用虚指的“那”取代“的”,如“乔光朴抡起眼睛,闪电般地扫过全场,最后和霍大道那穿透一切的目光相遇了”。

参考文献

- [1] Bybee, J. *Language, Usage and Cognition* [M]. Cambridge: Cambridge University Press, 2010.

- [2] DuBios, John W. Competing motivations [A]. In J. Haiman (ed.), *Iconicity in Syntax* [C]. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1985: 343-365.
- [3] Haiman, J. *Natural Syntax: Iconicity and Erosion* [M]. Cambridge: Cambridge University Press, 1985.
- [4] Langacker, R. W. *Foundations of Cognitive Grammar (Vol. I): Theoretical Prerequisites* [M]. Stanford: Stanford University Press, 1987.
- [5] 何晓炜. 双及物结构与重型名词短语移位研究[J]. 外语教学, 2008(2): 6-11.
- [6] 黄忠廉, 许萍. 汉译偏正结构中“的”字最佳用量探析[J]. 修辞学习, 1997(6): 40-41.
- [7] 刘丹青. 汉语名词性短语的句法类型特征[J]. 中国语文, 2008(1): 3-20.
- [8] 刘永生. 从句层面看领属性“N1/P + 的 + N2”结构中“的”字的隐现[J]. 修辞学习, 2004(6): 51-52.
- [9] 陆丙甫. 定语的外延性、内涵性和称谓性及其顺序[A]. 中国语言杂志社编. 语法研究和探索(四) [C]. 北京: 北京大学出版社, 1988: 102-115.
- [10] 陆丙甫. “的”的基本功能和派生功能——从描写性到区别性再到指称性[J]. 世界汉语教学, 2003(1): 14-29.
- [11] 王光全, 柳英绿. 定中结构中“的”字的隐现规律[J]. 吉林大学社会科学学报, 2006, 46(2): 115-121.
- [12] 徐阳春. 也谈人称代词做定语时“的”字的隐现[J]. 中国语文, 2008(1): 21-27.
- [13] 詹宏伟. 名词短语的无格位现象及其解释[J]. 外语教学, 2005(2): 26-28.
- [14] 严辰松. 从“年方八十”说起再谈构式[J]. 解放军外国语学院学报, 2008(6): 1-5.
- [15] 严辰松. 语言使用建构语言知识——基于使用的语言观概述[J]. 解放军外国语学院学报, 2010(6): 1-7.
- [16] 余光中. 论的的不休[A]. “翻译学术会议——外文中译研究与探讨”论文集[C]. 香港: 1996: 178-192.
- [17] 赵越. “N1的N2”比喻结构中“的”的隐现[J]. 汉语学习, 2009(4): 61-66.
- [18] 许艾明. 基于语料库的英汉中动句式修饰语之对比研究[J]. 西安外国语大学学报, 2011(4): 9-12.
- [19] 张伯江. 功能语法与汉语研究[A]. 刘丹青主编. 语言学前沿与汉语研究[C]. 上海: 上海外语教育出版社, 2005: 23-45.
- [20] 张立飞. 论频率对语言结构的建构作用[J]. 解放军外国语学院学报, 2010(6): 8-14.
- [21] 张敏. 认知语言学与汉语名词短语[M]. 北京: 中国社会科学出版社, 1998.
- [22] 朱德熙. 现代汉语形容词研究[J]. 语言研究, 1956(1): 83-111.
- [23] 朱德熙. 语法讲义[M]. 北京: 商务印书馆, 1982.

作者简介: 张立飞 解放军外国语学院科研部讲师, 博士, 主要研究方向为认知语言学和语料库语言学。

严辰松 解放军外国语学院科研部教授, 博士, 博士生导师, 主要研究方向为认知语言学和英汉对比。

收稿日期 2012-06-14

责任编辑 孙毅

类型学概括与二语学习者 汉语关系从句产出研究^{*}

上海财经大学 李金满 上海外国语大学 吴芙芸

提要: 汉语虽同英语一样具有 VO 语序,但其关系从句却罕见地前置于核心名词,这对英语本族语者习得该结构可能会构成潜在困难。本研究从语言类型学的三个概括出发,通过分析英语本族语者产出的汉语中介语语料库数据,考察了不同类型汉语关系从句的分布情况及其与类型学概括的关系。研究发现:二语学习者产出的汉语关系从句模式与建立在自然语言调查基础上的三个类型学概括有不同程度的契合;三个类型学原则的影响力强弱不一,彼此之间相互作用;生命性原则影响力最强,可及性序列其次,嵌入性最弱;并且生命性的强势导致了语料库分布和后两者的部分偏离。

关键词: 二语学习者、汉语关系从句、类型学概括

[中图分类号] H319 [文献标识码] A [文章编号] 1000-0429(2013)01-080-13

1. 引言

作为体现语言递归性的一种重要语法手段,关系从句(relative clause)普遍存在于世界各地诸多语言中,但其与核心名词的位置却依语言类型的不同而有所不同。大规模语言类型学调查(Dryer 2011)显示:在 VO 语中,关系从句几乎一律后置于名词(NRel),前置语序极为罕见,仅有 5 种:汉语、汉语方言(粤语、客家话)、白语(属于深受汉语影响的藏缅语)和阿眉斯语(属南岛语,一般认为亦受汉语影响)。由此可见,汉语关系从句前置于名词的现象的确非常特殊。

* 感谢国家社科基金青年项目“语言类型学视角下的汉英关系从句使用处理模式对比研究”(10CYY043)和吴芙芸主持的上海市浦江人才计划项目“填充词-语隙依存关系的加工机制——一项跨语言的研究”(10PJC066)的资助。

既然汉语关系从句在语序类型比较中如此“另类”，那么这种特殊性是否会在言语产出、理解、习得等方面有所体现呢？近些年围绕汉语关系从句加工的研究似乎证实了这一点。Hsiao & Gibson(2003)提出汉语与其他语言不同，其主语关系从句比宾语关系从句更难加工；不过，随后进行的相关行为实验和认知神经研究却结果不一，有同有异(详见吴芸芸 2011)。虽然是否汉语中宾语关系从句更容易处理迄今尚无定论，但这仍对已获得大量实证研究支持的主语关系从句加工优势论(参见 Gibson & Wu in press)提出了挑战。而在二语习得方面，比如把汉语作为二语的外国人士自然产出汉语关系从句情况如何，其分布模式是否符合普遍的主语关系从句优势及其他共性假设，相关研究尚付阙如。本文以英语国家留华学生为考察对象，针对他们产出的汉语关系从句进行调查，以了解类型独特的汉语关系从句在二语中的使用特点。

2. 研究背景

语言类型学和二语习得两个领域相互借鉴，加强对话，有利于推动彼此的理解和发展(Giacalone Ramat 2003)。类型学研究成果可以为二语习得研究提供丰富的理论基础和方法启示，而二语习得领域的发现也可以反哺类型学调查研究。鉴于关系从句是语言类型学和二语习得两领域对话成果尤其丰硕的部分(Comrie 2003:19)，以下先简要介绍与关系从句有关的三个类型学理论假设。

2.1 可及性等级

语言类型学研究中最最重要的一个语法等级序列是“名词短语可及性等级”(Noun Phrase Accessibility Hierarchy, NPAH)(Keenan & Comrie 1977; Comrie & Keenan 1979)，事关名词关系从句化(relativization)的类型。Keenan 和 Comrie 深入研究了全世界不同地区不同类型的五十余种语言中的关系从句结构，发现这些语言都可以形成主语关系从句，其他类型的关系从句则不然。他们认为，主语关系从句是可及性最高、最容易形成的无标记关系从句类型，并根据先行词在从句中的语法地位，建立了著名的 NPAH 序列：主语 > 直接宾语 > 非直接宾语 > 领属语。随后进行的类型学调查研究表明，可及性等级假说的确广泛适用于各类关系从句在自然语言中的分布(Comrie 1989)。

NPAH 是语言类型学和二语习得交汇广泛的研究区域。基于这一可及性假设，许多二语习得研究者围绕关系从句进行了多角度探讨。绝大多数研究显示，二语学习者习得、产出各类关系从句的难易程度和先后顺序与可及性等级序列约摸一致，特别是主语 > 直接宾语(Gass 1979; Eckman *et al.* 1988; Izumi 2003; Li 2007 等)。

2.2 嵌入性概括

第二个类型学概括与关系从句在主句中的嵌入位置相关。据一些语言类型学研究的观察,很多语言中的从句倾向于旁置于主句两端,而不是内嵌于主句中间(Kuno 1974; Dryer 1980)。

Kuno(1974)根据人类记忆系统的感知能力,提出了“感知难度假设”(Perceptual Difficulty Hypothesis)。该假说认为,由于短期记忆的局限性,嵌入主句中间位置的关系从句中断了主句的解析过程,加重了人们工作记忆的负担,因而处理起来比旁置于主句前后两端位置的关系从句要困难;而位于句首或句末的关系从句由于没有中断主句,因此不影响主句的处理,使用和习得起来理应相对容易。随后的一些围绕一语(Bates *et al.* 1999; Kidd & Bavin 2002)、二语(Izumi 2003; Mellow 2006; 李金满 2008 等)的研究也证实了这一观点。

2.3 生命性等级

语言类型学研究中另一个制约语言表征的重要等级序列当推生命性等级:人 > (非人)生命体 > 无生命体¹(Comrie 1989)。作为名词短语的固有特性和普遍存在的概念属性,生命性在许多语言现象中都颇具影响力。很多研究(如 Comrie 1989)显示,名词短语的生命性与其在自然语言句法中的语法身份关系密切:主语倾向于为有生命体,而直接宾语倾向于为非生命体。具体而言,在典型及物构式中,最自然的模式是施动主语的生命性、限定性高,受动宾语的生命性、限定性低。任何与这种模式不同的表达都可能需要显性或隐性的符号标记。

除了制约语言的具体外在表达形式之外,生命性也可能体现为某一语言项使用频率的高低或者加工的难易。比如,针对不同类型语言的语料库考察和心理实验就证实了生命性等级对关系从句的这种影响力(德语、荷兰语: Mak *et al.* 2002; 英语: Traxler *et al.* 2002, 2005; Gennari & MacDonald 2008, 2009; 汉语: Wu *et al.* 2012; 吴芙芸 2011; 学习者英语: 李金满、王同顺 2007 等)。

3. 研究方法

以下我们就从上述三方面入手,通过语料库调查,详细探讨将汉语作为二语(含外语)学习的英语本族语者对语序类型独特的汉语关系从句的使用情况。换言之,本文研究的具体问题是:英语本族语者产出的汉语关系从句分布情况是否符合可及性、嵌入性、生命性这三个类型学原则?

3.1 语料库简介

本研究采用北京语言大学崔希亮(2008)主持编制在线发布的“HSK 动态作文语料库”中的部分数据。“HSK 动态作文语料库”是母语非汉语的外国人参加

高等汉语水平考试(HSK 高等²)作文考试的答卷语料库,收集了1992-2005年的部分外国考生的作文答卷。在线语料库1.1版语料总数达到11569篇,作文题目总数为30,共计424万字。由于本研究主要考察母语为英语的外国留华学生产出汉语关系从句的情况,我们从在线库中抽取了母语为英语的四个国家³留学生的作文429篇,具体国籍和篇章数目分别为:澳大利亚,123;加拿大,80;美国,118;英国,108。

3.2 语料提取

提取汉语关系从句之前,首先须对汉语关系从句进行充分界定。本研究根据研究目的,借鉴刘丹青(2005:194,201-209)结合汉语特点对Comrie和Smith《语言描写问卷》中关系从句所做的说明,并参照Wu(2009)的做法,确定研究对象应满足以下几个条件:

1) 必须含有标句词“的”,且“的”字的用法不含领属关系。

2) 必须是有空位(gap)的关系从句,即从句所修饰的中心名词必须在从句中占据一个句法位置(如从句内主语或宾语等),并与该位置上的空位或复指代词同标(co-index)。

此外,我们还考虑了必须包含的情况(3)和必须剔除的两种情况(4和5)。

3) 当从句中只有动词短语、没有名词做主语时,如“(昨天/在商场)买的一本书”,认定为关系从句。这是由汉语中主语成分省略比较自由的特点决定的。

4) 形容词短语(比如“一个快乐的假期”)和“家中/学校里的人”等表方位处所的结构排除在关系从句之外。形容词作定语的情况虽然也有理由看作关系从句,但在汉语“调查时可以不必要放在关系从句部分处理”(刘丹青2005:202)。

5) 话题结构或所谓的“主谓谓语句”排除在关系从句之外。汉语是话题结构丰富的语言,允许主谓齐全的小句在核心名词前做定语,比如“成绩很好的学生、节奏比较强烈的流行歌曲”等。此类话题型从句类似于英文中whose引导的领属语关系从句,不在本研究考察范围之内。

确定好待考察的关系从句种类后,我们首先运用“‘的’+国籍+证书级别(A级、B级)”搭配式对“HSK动态作文语料库”进行字符串检索,抽取出所有包含“的”字的句子,具体数目(A级,B级)如下:澳大利亚(188,739);加拿大(107,461);美国(79,372);英国(217,698)。然后,我们根据上面确定的判定标准,通过人工手动逐句辨别,剔除掉不合格项,找出所有符合要求的限定性主语、宾语关系从句,总计308例。

3.3 语料标注

关系从句全部提取完成后,我们根据本研究从类型学角度考察的三个因素

(可及性、嵌入性、生命性),对其逐一进行标注。标注内容具体如下:

1) 根据核心名词在关系从句中的句法身份和从句动词类型,将所有关系从句标注为三类:及物动词主语关系从句(S1)、非及物(包括不及物动词、系动词)主语关系从句(S2)、直接宾语关系从句(O);

2) 根据关系从句在主句中的嵌入位置,作如下标注:关系从句内嵌于主句中间记作 E;关系从句旁置于主句前端(或左端)记作 L;而不属于这两种方式的则归入其他类(比如做表语的名词、做介词宾语的名词、领属结构中的名词、独立名词等);

3) 根据名词的生命性,对主句中核心名词和从句内名词的生命性进行标注。所有人类和非人类的有生命名词记作 A,其他无生命名词记作 I。

全部标注完成后,我们统计各个单项在语料库中的使用频率,计算所占比例,并运用 SPSS 进行统计检验,确定不同种类之间是否具有显著性差异。

4. 结果和讨论

在本节中,我们从可及性、嵌入性、生命性几个角度逐一描述探讨英语本族语者产出汉语关系从句的分布特点。

4.1 可及性

表 1 显示了主语、宾语关系从句的总体分布态势。在所提取的 308 个关系从句中,证书级别为 A 级的学生共产出 68 例,B 级学生则产出 240 例。这种不均衡的分布态势可归因于 A 级学生数量较少,收入作文篇数亦相应较少。

表 1. 主语、宾语关系从句的总体分布情况

	主语 S			(及物) 宾语 O	总数
	及物 主语 S1	非及物 主语 S2	总数		
A 级	20(29.41%)	23(33.82%)	43(63.24%)	25(36.76%)	68(100%)
B 级	71(29.58%)	75(31.25%)	146(60.83%)	94(39.17%)	240(100%)
总数	91(29.55%)	98(31.82%)	189(61.36%)	119(38.64%)	308(100%)

从核心名词在关系从句内的语法身份(即关系从句类型)来看,不管是 A 级还是 B 级,主语关系从句总体均多于宾语关系从句,且比例基本一致:A 级中主语关系从句为 43 句,占 63.24%;B 级中主语关系从句为 146 句,占 60.83%。因此,总体而言,学习者产出宾语关系从句确实较少,所占比例不到 40%(119/308)。二项统计检验显示,汉语主、宾语关系从句的使用频率均具有显著性差异(A 级: $p < .05$;B 级: $p = .001$;总数: $p < .001$)。这与可及性等级序列

(NPAH)相吻合,并与前人基于汉语一语(如 Hsiao & Gibson (2003)和 Wu (2009)的 Treebank 汉语新闻语料库、Pu (2007)的汉语小说、口述默片语料库等)和英语为二语的中国学习者(如 Li (2007)的作文语料库)调查得出的结果均一致。该发现意味着可及性等级序列的广泛适用性,即无论母语中关系从句的取向如何,对于二语学习者而言,他们产出主、宾语关系从句的总体情况均符合“名词短语可及性等级”这一类型学概括。

不过,如果我们排除约占关系从句总数三分之一的非及物主语关系从句(S2),仅考察从句动词为及物的情况,即 NVN 句式,那么 A 级、B 级中的宾语关系从句在数量上都超过了主语关系从句($O > S1$):A 级, $25 > 20$;B 级, $94 > 71$;总数, $119 > 91$;但差异均不显著($p > .05$)。这说明汉语学习者在产出汉语及物关系从句时,没有显示出对主语关系从句的偏好。这与 Hsiao & Gibson (2003)、Wu(2009)基于 Treebank 汉语新闻语料库、Sheng & Wu(2012)基于汉语口语语料库针对及物关系从句所得出的结果不同。

至于二语学习者产出的及物关系从句中宾语关系从句为何在数量上反而多于主语关系从句,我们认为可能有以下几个原因。1) 母语迁移的作用。汉语宾语关系从句的 NVe_i 结构与英语简单主句 SVO 语序相同,对二语学习者而言,或许母语与二语的构式相似会引发母语句式正向迁移,从而造句更容易。与此相反,汉语主语及物关系从句是 e_iVN 结构,对于母语英语不允许主语缺省的二者而言,没有母语迁移的可能,理解和产出此类结构较难,因此倾向选择规避。2) 语料特性的限制。受二语水平和作文话题⁴ 要求制约,二语学习者产出的关系从句句式较短,相对简单,有可能将宾语关系从句作为一个语块来使用,比如“自己说的话/子女做的事/我喜欢的东西”等,从而导致宾语关系从句的产出增多。3) 话题新旧信息的引入干扰。通常情况下,命题作文要求二语学习者围绕某一话题或人物深入探讨或延伸展开,主语通常是已知的核心话题,或是在之前语境中已经提到的旧信息(Givón 2001),包括高度可及的人称代词(“我、你、她、他”)(Real & Christiansen 2007; Roland *et al.* 2012)。因此,二语学习者更倾向于用主语作为前景信息,引入表达新信息或新话题的宾语(Fox & Thompson 1990),从而产出更多的宾语关系从句。4) 句中名词短语的生命性,我们下文再论。

4.2 嵌入性

4.2.1 嵌入性

表 2 汇总了关系从句在主句中的分布态势。从总体情况来看,三类关系从句(即旁置、内嵌、其他)约各占 1/3,其中旁置关系从句略少于内嵌在主句中间

的关系从句,后者又略少于其他位置的关系从句⁵,说明汉语学习者对于关系从句在主句中的位置并没有表现出明显偏好($p = .588$)。这一点从 B 级学生产生的关系从句中也可得到进一步印证:旁置和内嵌的比例几乎相同,都约为 32%。

表 2. 关系从句在主句中的位置情况

	旁置 L(S)	内嵌 E(O)	其他	总数
A 级	17(25%)	26(38.24%)	25(36.76%)	68(100%)
B 级	78(32.5%)	77(32.08%)	85(35.42%)	240(100%)
总数	95(30.84%)	103(33.44%)	110(35.71%)	308(100%)

鉴于本节主要考察汉语关系从句修饰主句主语(即旁置)和主句宾语(即内嵌)的情况,以下我们剔除其他位置的关系从句,并将 A、B 级学习者产生的关系从句合并起来。表 3 显示了不同类型关系从句在主句中的位置分布情况。

表 3. 嵌入不同主句位置的主、宾语关系从句分布情况

	及物 主语 S1	不及物 主语 S2	(及物) 宾语 O	总数
旁置 L(S)	38(63.33%)	26(45.61%)	31(38.27%)	95(47.98%)
内嵌 E(O)	22(36.67%)	31(54.39%)	50(61.73%)	103(52.02%)
总数	60(100%)	57(100%)	81(100%)	198(100%)

如表 3 所示,单从汉语关系从句在主句中的嵌入位置来看,内嵌在主句中间的关系从句比旁置于主句左端的关系从句总体数量稍多(103:95),但无倾向性($p = .490$)。这说明,英语本族语者产出汉语关系从句时没有表现出要避免从句内嵌的倾向,甚至汉语关系从句修饰主句宾语的频次还多于主句主语。这与类型学概括不太一致。

为何通常认为较难处理的内嵌从句在本研究中出现较多?究其原因,可能一是二语学习者的特性或学习策略使然。学习者倾向于将一些常见的简短从句(如“这项新措施可以减少[抽烟的]人”、“他个人是有权力做[自己喜欢的]事”)作为一个语块习得、存储和使用,这样内嵌的从句实际上不会形成真正意义上的主句加工中断,从而不会带来困难。二是受到母语的影响。在英语中,关系从句倾向旁置于句子末端(即修饰主句宾语的关系从句较多)。受母语使用习惯的影响,二语学习者在产出汉语关系从句时,可能也会倾向将其往主句后面放,从而一定程度上导致内嵌。

4.2.2 嵌入性和可及性

综合考虑可及性和嵌入性,我们发现,当主语关系从句为非及物动词句式

(S2)时,从句内嵌于主句中间(如“吸烟已成为[受到严格限制的]活动”)还是旁置于主句左端(如“[刚出生的]孩子们怎能学会走路或说话呢”)差别不明显,虽然内嵌关系从句稍多(31:26),但没有显著倾向($p = .597$)。

当从句为及物主语关系从句(S1)时,旁置从句多于内嵌从句(38:22),符合内嵌从句感知难度高的观点;而当从句为宾语关系从句时,内嵌从句(如“我会领悟到一些[自己从未领悟到的]想法”)明显多于旁置从句(如“[他们呼出的]二手烟污染了整个区域的空气”)(50:31),占 61.73%,不符合“感知难度假设”。对及物关系从句的列联表检验显示,从句嵌入主句的位置与从句类型关联密切($p = .003$):主语关系从句倾向于旁置于主句左侧(即修饰主句主语的主语关系从句(SS1)多于修饰主句宾语的宾语关系从句(OS1)),而宾语关系从句倾向于内嵌于主句中间(即修饰主句宾语的宾语关系从句(OO)多于修饰主句主语的主语关系从句(SO))。

四类及物关系从句的产出频数排序如下:OO > SS1 > SO > OS1,即内嵌在主句中间的宾语关系从句(OO)产出最多,其次是左置的主语关系从句(SS),然后是左置的宾语关系从句(SO)和内嵌的主语关系从句(OS)。该排序与基于汉语一语语料得出的排序(Pu (2007):SS > OS > OO > SO; Wu (2009):SS > SO > OS > OO)并不完全一致。我们认为,这一分布模式与名词的生命性也存在密切联系,将在 4.3.3 小节中一并探讨。

4.3 生命性

4.3.1 生命性

对于生命性,我们首先考察核心名词的生命性及其在关系从句中担任的语法角色情况,如表 4 所示。

表 4. 有/无生命核心名词在关系从句中的语法身份情况

		有生命 A	无生命 I	总数
主语 S	主语 S1	69(75.82%)	22(24.18%)	91(100%)
	主语 S2	40(40.82%)	58(59.18%)	98(100%)
	总数	109(57.67%)	80(42.33%)	189(100%)
宾语 O		17(14.29%)	102(85.71%)	119(100%)
总数		126(40.91%)	182(59.09%)	308(100%)

在全部 308 例关系从句中,182 例修饰无生命核心名词,占 59.09%;126 例修饰有生命核心名词,占 40.91%。这意味着汉语二语学习者倾向于使用关系从句来修饰无生命名词($p = .002$)。

从关系从句的提取类型来看,126 个含有生命核心名词的关系从句中,109

例为主语关系从句(S),占 86.51%;相对而言,182 个含无生命核心名词的关系从句中,102 例为宾语关系从句(O),占 56.04%。核心名词的生命性与其在关系从句中的语法身份关联密切($p < .001$):有生核心名词倾向于在从句中做主语(即施事),而无生核心名词倾向于在从句中做宾语(即受事)。这说明二语为汉语的学习者所产出的关系从句亦符合类型学对生命性的概括。

结合从句动词来看,我们更容易发现这种分布特点。当从句为及物动词(即 NVN 句式)时,核心名词的生命性与其在关系从句中的语法身份显著相关($p < .001$):75.82%的主语关系从句(S1)修饰的名词为有生命体;85.71%的宾语关系从句(O)修饰的名词为无生命体。二者倾向性泾渭分明。

回顾 4.1 小节,正是由于无生命核心名词居多(共 182 例),且大部分(共 102 例)集中在宾语关系从句(O),因此,含及物动词的关系从句中(即 NVN 句式),宾语关系从句(O)在数量上超过了主语关系从句(S)(119:91)。

4.3.2 生命性和可及性

本节继而考察及物关系从句内名词的生命性。表 5 列出了当同时考虑两个名词生命性的时候,主、宾语关系从句的产出情况。

表 5. 不同生命性组配下的主、宾语关系从句分布情况

核心名词 从句内名词	有生命 A		无生命 I	
	有生命 A	无生命 I	有生命 A	无生命 I
主语 S1	15(46.88%)	54(100%)	4(4.35%)	18(56.25%)
宾语 O	17(53.13%)	0(0%)	88(95.65%)	14(43.75%)
总数	32(100%)	54(100%)	92(100%)	32(100%)

由表 5 可以发现,当关系从句的两个名词生命性相同时,关系从句产出相对较少,均为 32 例;当关系从句的两个名词生命性不同时,关系从句产出较多,共有 146 例。这说明关系从句更倾向于拥有两个生命性相异的名词而不是生命性相同的名词($p < .001$),与吴芙芸(2011)的调查结果一致。单看两名词生命性相同的关系从句分布时,主语、宾语关系从句的产出区别不大:两名词同为有生命时,宾语关系从句产出多 2 例;同为无生命时,主语关系从句产出多 4 例。

当核心名词与关系从句内名词的生命性不同时,主语、宾语关系从句产出的情况则大相径庭:当核心名词有生命、从句内名词无生命时,54 例关系从句皆为主语关系从句;而当核心名词无生命、从句内名词有生命时,92 例关系从句中宾语关系从句占 95.65%,共 88 例,主语关系从句仅有 4 例。而细观这 4 例句子,均为带有“让、令”等显性标记的致使句式(如:“代沟”问题是一个经常[让晚辈和长辈都头痛或烦恼的]事情。),这在语义层面上凸显了无生(核心)名词充任主语

施事的可能性,但其更多呈现为有标记结构。综上可见,从句中两个名词的生命性组配极大地影响了关系从句的使用类型。

4.3.3 生命性、可及性和嵌入性

进一步将可及性、生命性、嵌入性三个因素一起纳入考察,则及物关系从句的各项详细分布情况如表 6 所示。

表 6. 及物关系从句详细分布情况

		有生命 A 核心名词		无生命 I 核心名词		总数
		无生命 I	有生命 A	无生命 I	有生命 A	
主语 S1	旁置 L(S)	22(65%)	6(38%)	10(45%)	0(0%)	38(27%)
	内嵌 E(O)	12(35%)	4(25%)	3(14%)	3(4%)	22(16%)
宾语 O	旁置 L(S)	0(0%)	5(31%)	3(14%)	23(33%)	31(22%)
	内嵌 E(O)	0(0%)	1(6%)	6(27%)	43(62%)	50(35%)
总数		34(100%)	16(100%)	22(100%)	69(100%)	141(100%)

在 4.2.1 节我们同时考虑可及性和嵌入性两个因素时,得出的频数序列为 $OO > SS1 > SO > OS1$,与两个类型学假设在总体上都不够吻合。但当我们结合生命性一起看时,就会发现这个分布模式主要是由两个名词论元的生命性因素所导致。

第一、按照生命性原则,有生命名词倾向于(在主句和/或关系从句中)充当主语,而无生命名词倾向于(在主句和/或关系从句中)充当宾语;能在两个小句中同时满足这种倾向的正是 OO 和 SS,因此这两个关系从句类型出现较频繁。

第二、本调查研究中,核心名词为非生命体偏多,因此 OO 的使用频次超过了 SS。

第三、SO 和 OS 两个关系从句类型都只在一个位置满足了生命性倾向,因此使用频次相对于 SS 和 OO 偏少。

第四、SO 符合避免内嵌的倾向,而 OS 违背了该倾向,因此 $SO > OS$ 。

此外,当核心名词和从句内名词生命性不同时,方可同时满足一个有生命名词充当主语、另一个无生命名词充当宾语的条件。因此,关系从句无论可及程度高低(即主语关系从句还是宾语关系从句),或线性嵌入程度深浅(即旁置还是内嵌),都是 AI 或 IA 这种生命性相异组配的关系从句使用频率最高。差别仅仅在于,AI 组配在主语关系从句中居多,而 IA 组配在宾语关系从句中居多。由此可见,生命性原则和可及性、嵌入性都联系紧密,也就是说,生命性原则实际上影响了关系从句产出的类型以及关系从句在主句中的位置。

5. 结论

综上所述,我们对母语为英语的二语学习者产出的汉语关系从句做出如下归纳:

1) 就可达性而言,本研究中英语本族语者产出的汉语关系从句总体上与可达性等级序列一致:主语关系从句多于宾语关系从句。但在带有及物动词的关系从句中,主、宾语关系从句的产出并没有显著差异;数量上后者稍多于前者。

2) 就嵌入性而言,本族语为英语的二语习得者总体上对于关系从句在主句中的嵌入位置没有明显倾向,以往研究中发现的避免关系从句内嵌的现象在本研究中并没有得到证实。事实上,内嵌的从句在数量上还稍多一点。但是,分类观察显示及物主语关系从句仍有左置的倾向($SS > OS$);宾语关系从句则相反,内嵌更多($OO > SO$)。

3) 就生命性而言,本研究完全支持生命性原则:有生命名词有强烈的倾向充当句子施事主语,而无生命名词有强烈的倾向充当句子受事宾语。

我们认为,如第四节所析,1)和2)中分别与可达性、嵌入性不吻合的现象在很大程度上是因为受到了生命性的干扰。由于本调查中无生命核心名词数量占大多数,因此依据生命性原则,宾语关系从句自然增多,在主句中的宾语位置(即内嵌位置)出现也增多,从而导致了偏离。此外,其与语料库题材、母语迁移的影响、新旧信息的引入、二语水平等因素亦或相关。至于英语本族语者产出的汉语关系从句多数修饰无生命名词到底是一种普遍倾向,还是同研究语料的性质(如二语、语种、语体等)有关,还需要更广泛语言的检验。

基于本研究调查,我们发现,虽然汉语关系从句在语序上类型独特,但二语学习者产出的关系从句模式与建立在自然语言调查基础上的三个类型学原则仍有不同程度的吻合。而且,不同类型学原则的影响力有强有弱,彼此之间相互竞争,相互制约,综而观之能更好地理解语言现象。在本研究中,生命性原则的制约力表现最强劲,可达性序列其次,嵌入性最弱;也正是生命性的强势导致了关系从句分布模式与后两者的部分偏离。

注 释

1. 感谢金立鑫先生对初稿的意见。Animacy 有时也译作“生命度”,其不同语言中的区分有粗细不同。
2. HSK(高等)适用于接受过 3000 学时及以上现代汉语正规教育的学习者。
3. 英语国家人士并不绝对等同于母语为英语的人士,虽然他们的母语绝大多数为英语。由于现有的语料库以国籍标注来区分汉语学习者,本研究只能以国籍为参数来提取。
4. 语料库中的作文话题为议论文和记叙文,主要包括“如何解决代沟问题、吸烟对个人健康和公众利益的影响、我看流行歌曲、如何看待安乐死、我对离婚问题的看法、记对我影响最大的人”等 30 个题目。
5. 这与前期一语调查有同有异。本研究中位于其他位置上的关系从句比例与 Wu(2009)基于 Treebank

汉语新闻语料库的数据一致,约为 1/3;而 Pu(2007)基于汉语小说的此类从句所占比例较小,仅 9%。有趣的是,本研究中位于主句主语、宾语位置上的关系从句比例相近,与 Pu(2007)约摸一致(49%:42%),与 Wu(2009)则不同,其研究中前者多于后者(44%:24%)。

参考文献

- Bates, E., A. Devescovi & S. D'Amico. 1999. Processing complex sentences: A cross-linguistic study [J]. *Language and Cognitive Processes* 14: 69-123.
- Comrie, B. 1989. *Language Universals and Linguistic Typology* (2nd edition) [M]. Chicago: The University of Chicago Press.
- Comrie, B. 2003. Typology and language acquisition: The case of relative clauses [A]. In A. Giacalone Ramat (ed.), *Typology and Second Language Acquisition* [C]. Berlin: Mouton de Gruyter. 19-37.
- Comrie, B. & E. Keenan. 1979. Noun phrase accessibility revisited [J]. *Language* 55: 649-664.
- Croft, W. 2003. *Typology and Universals* (2nd edition) [M]. Cambridge: CUP.
- Dryer, M. 1980. The positional tendencies of sentential noun phrases in Universal Grammar [J]. *Canadian Journal of Linguistics* 25: 123-195.
- Dryer, M. 2011. Relationship between the order of object and verb and the order of relative clause and noun [OL]. <http://wals.info/chapter/96> (accessed 05/04/2012).
- Eckman, F., L. Bell & D. Nelson. 1988. On the generalization of relative clause instruction in the acquisition of English as a second language [J]. *Applied Linguistics* 9: 1-20.
- Fox, B. & S. Thompson. 1990. A discourse explanation of the grammar of relative clauses in English conversation [J]. *Language* 66: 297-316.
- Gass, S. 1979. Language transfer and universal grammatical relations [J]. *Language Learning* 29: 327-344.
- Gennari, S. & M. MacDonald. 2008. Semantic indeterminacy in object relative clauses [J]. *Journal of Memory and Language* 58: 161-187.
- Gennari, S. & M. MacDonald. 2009. Linking production and comprehension processes: The case of relative clauses [J]. *Cognition* 111: 1-23.
- Giacalone Ramat, A. 2003. Introduction [A]. In A. Giacalone Ramat (ed.), *Typology and Second Language Acquisition* [C]. Berlin: Mouton de Gruyter. 1-18.
- Gibson, E. & H. Wu. In press. Processing Chinese relative clauses in context [J]. *Language and Cognitive Processes* iFirst: 1-31.
- Givón, T. 2001. *Syntax: A Functional Typological Introduction* (Vol. 2) [M]. Amsterdam: John Benjamins.
- Hsiao, F. & E. Gibson. 2003. Processing relative clauses in Chinese [J]. *Cognition* 90: 3-27.
- Izumi, S. 2003. Processing difficulty in comprehension and production of relative clauses by learners of English as a second language [J]. *Language Learning* 53: 285-323.
- Keenan, E. & B. Comrie. 1977. Noun phrase accessibility and Universal Grammar [J].

- Linguistic Inquiry* 8: 63-99.
- Kidd, E. & E. Bavin. 2002. English-speaking children's comprehension of relative clauses [J]. *Journal of Psycholinguistic Research* 31: 599-615.
- Kuno, S. 1974. The position of relative clauses and conjunctions [J]. *Linguistic Inquiry* 5: 117-136.
- Li, J. 2007. An Empirical Study of Chinese EFL Learners' Behavior on English Relative Clauses: Bridging Typology and SLA [D]. Ph. D. Dissertation, Shanghai Jiaotong University.
- Mak, W., W. Vonk & H. Schriefers. 2002. The influence of animacy on relative clause processing [J]. *Journal of Memory and Language* 47: 50-68.
- Mellow, J. 2006. The emergence of second language syntax: A case study of the acquisition of relative clauses [J]. *Applied Linguistics* 27: 645-670.
- Pu, M. 2007. The distribution of relative clauses in Chinese discourse [J]. *Discourse Processes* 43: 25-53.
- Real, F. & M. Christiansen. 2007. Processing of relative clauses is made easier by frequency of occurrence [J]. *Journal of Memory and Language* 57: 1-23.
- Roland, D., G. Mauner, C. O'Meara & H. Yun. 2012. Discourse and relative clauses: Building a more complete model of expectations in language processing [J]. *Journal of Memory and Language* 66: 479-508.
- Sheng, Y. & F. Wu. 2012. Audience design affects classifiers positioning in Chinese relative clauses [R]. Poster presented at the 25th Annual CUNY Conference on Human Sentence Processing, New York, March 2012.
- Traxler, M., R. Morris & R. Seely. 2002. Processing subject and object relative clauses: Evidence from eye movements [J]. *Journal of Memory and Language* 47: 69-90.
- Traxler, M., R. Williams, S. Blozis & R. Morris. 2005. Working memory, animacy, and verb class in the processing of relative clauses [J]. *Journal of Memory and Language* 53: 204-224.
- Wu, F. 2009. Factors Affecting Relative Clause Processing in Mandarin [D]. Ph. D. Dissertation. University of Southern California.
- Wu, F., E. Kaiser & E. Andersen. 2012. Animacy effects in Chinese relative clause processing [J]. *Language and Cognitive Processes* 27: 1489-1524.
- 崔希亮, 2008, HSK 动态作文语料库 [OL], <http://202.112.195.192:8060/hsk/login.asp> (2011 年 05 月 01 日读取)。
- 李金满, 2008, 中国学习者英语关系从句使用行为研究 [J], 《现代外语》(4): 406-414。
- 李金满、王同顺, 2007, 当可及性遇到生命性 [J], 《外语教学与研究》(3): 191-200。
- 刘丹青, 2005, 语法调查与研究中的从属小句问题 [J], 《当代语言学》(3): 193-212。
- 吴芙芸, 2011, 基于经验还是基于工作记忆? 来自汉语新闻语料库中关系从句生命度格局的证据 [J], 《语言科学》(4): 396-408。

收稿日期: 2012—04—07; 修改稿, 2012—08—13; 本刊修订, 2012—11—25

通讯地址: 200433 上海市 上海财经大学外语系(李)

200089 上海市 上海外国语大学语言研究院(吴)

指量结构与汉语关系从句共现时的不对称分布 及其原因*

—— 一项基于真实口语语料库的研究

上海外国语大学 盛亚南 吴芙芸

提要:本研究从《鲁豫有约》的真实口语语料出发,分析了指量结构与关系从句共现时的分布态势及其内在驱动因素。研究结果表明:在主语关系从句中,指量结构倾向于前置;在宾语关系从句中,指量结构倾向于后置。验证了前人通过书面语料得到的分布趋势。说明说者会采用利他的言语产出策略:在主语关系从句中,提前给出提示线索,以帮助听者构建目标句式,从而更倾向产出前置的指量结构;而在宾语关系从句中,为了帮助听者规避语义冲突或潜在歧义,说者更倾向产出后置的指量结构。该结论支持基于听众设计的利他说。

关键词:言语产生、口语语料、关系从句、指量结构

[中图分类号] H030 [文献标识码] A [文章编号] 1003-6105(2013)02-0150-08

1. 问题的提出

汉语的基本语序是 VO,但关系从句的核心名词却置尾,在语言类型学上较为罕见(Greenberg 1966;Dryer 1992),因此汉语关系从句的研究得到了广泛关注。此外,汉语较印欧语言缺乏形态标记,而量词要求与所修饰的名词在语义上匹配。考虑到汉语关系从句和量词的独特性,已有一些学者关注指量结构在关系从句中的位置分布(Chao 1968;Tsao 1986;Zhang 2006;唐正大 2007;Pu 2008;徐赳赳 2008;Wu 2009;吴芙芸 2011a;Ming & Chen 2010)。指量结构和关系从句共现时,主要有两种位序:一种是指量结构前置(1a);另一种是指量结构后置(1b)。

(1) a. 主语关系从句,指量结构前置

那个[t_i遇见星探的]女孩_i长得很漂亮。

b. 主语关系从句,指量结构后置

[t_i遇见星探的]那个女孩_i长得很漂亮。

本文仅考察“空位关系从句”(gapped relative clauses),即中心词源于从句内的某一位置,并与原位置上的语迹同标。根据中心名词在从句中的提取位置,本文重点关注两种关系从句类型:主语关系从句(1)和宾语关系从句(2)。

(2) a. 宾语关系从句,指量结构前置

那个[星探遇见 t_i的]女孩_i长得很漂亮。

b. 宾语关系从句,指量结构后置

[星探遇见 t_i的]那个女孩_i长得很漂亮。

针对汉语指量结构和关系从句共现时的位序分布这一议题,已有的研究主要基于小说、新闻等书面语料(唐正大 2007;Pu 2008;Wu 2009;Ming & Chen 2010),得到了较为一致的结论:指量结构在主语关系从句中倾向于前置,在宾语关系从句中则倾向于后置。唯一的例外,可能是徐赳赳(2008)。该文通过

* 本文系浦江人才计划资助(10PJC066)“填充词-语隙依存关系的加工机制——一项跨语言的研究”和上海外国语大学校级一般项目科研基金(2012114025)“指量结构在汉语关系从句的产出模式及内在动因:一语与二语学习者的对比研究”的部分研究成果。两位作者对本研究的贡献相当。感谢匿名审稿专家和编辑部提出的修改意见。许余龙、金立鑫、李金满先生对初稿提出过批评意见,谨表感谢。

考察《京华时报》的叙述文语料,发现指量结构¹更倾向前置于关系小句。但该文将不含空位的附加语关系从句²、名词补语从句³亦算作关系小句,占411例关系小句总数的49.4%(徐赳赳2008:8)。其结论是否能推及到空位关系从句,尚不可知。

已有的语料库研究鲜有涉及口语,可能仅有两项(唐正大2007;Pu2008),但是样本很小,且结果不一。Pu(2008)采用口头描述无声短片的实验任务,验证了书面语料所呈现的不对称分布态势。但该研究中的被试仅产出10例关系从句与指量结构共现的结构。唐正大(2007)基于北京方言会话和关中方言录音口语语料则发现,不论关系从句的类型,指量结构都倾向后置于从句。这有异于前人和其书面语料的研究结果。但唐文并未给出动词的具体分类,结论还有待进一步考证。

对于为何指量结构会根据关系从句类型呈不对称的分布态势,学者们尚未给出统一的解释。唐正大(2007)在优选论框架下提出至少两项制约条件,即“名词短语身份尽早确立”和“非歧义结构”,在书面语中前者优于后者⁴,而在口语中后者优于前者。我们认为,这些制约条件具有一定的认知、功能理据,但依语体差异做不同排序,似乎缺乏内在动因。Pu(2008)试图从语用及认知加工的角度进行解释,但是否普遍适用,尚需更多的语料检验。Wu(2009)提出“尽早出现”和“语义冲突规避”两项原则,但似乎是针对汉语现象而提出的特设性(ad hoc)规则,且更像是语言生成的原则,尚需言语实时产生或加工方面的证据。

综上所述,虽然基于书面语料的研究得到了相对统一的结果,但书面语多为几经修改后的最终版本,可能并不能真实地反映言

语生成的过程。因此,其分布趋势是否具有普遍性,还有待口语语料的进一步验证。在已有的为数不多的口语语料研究中,目前至少还存在两方面的争议:1)指量结构在口语语料中与关系从句共现时,其位序是否与在书面语料中的分布态势一致,即是否存在语体差异?2)如果的确存在指量结构依关系从句类型呈不对称分布的态势,那么其背后的驱动内因是什么?本文试图通过一项基于真实口语语料的研究来回答以上问题,并进一步揭示言语生成的内在机制。

2. 理论预测

在具体分析语料之前,我们根据当前主要的两大言语生成理论,对指量结构在汉语关系从句的位置分布做出不同的预测。

言语生成的一个重要理论是基于说话者内部认知限制(Speaker-internal Constraints),或称“利己性”假说(Zipf 1949; Ferreira & Dell 2000; Arnold *et al.* 2004)。该理论认为,说话者在言语生成过程中会优先使用可及性高、比较容易提取的语言结构,以便赢取时间,编码相对较难的结构,从而减轻说话者言语生成过程中的负荷。人类语言倾向于将长成分后置,而关系从句长于含有指示词的指量结构(参见Hawkins(1983:88-113)的“长度顺序原则”)。据此,不论关系从句类型,指量结构应该倾向于在关系从句之前出现。

而言语生成的另一派理论——基于听众设计(Audience Design,又称为“利他性”假说,见Clark & Murphy 1982; Brennan & Clark 1996; Temperley 2003)则认为,说话者在言语产生过程中会尽量采取一些利他的策

¹ 徐文(2008)将修饰中心词的指量结构具体细分为数量词和指示代词两种情况;前者共40例,35例为指量结构前置;后者共15例,无一例外均为指量结构前置。

² 如“采取[以他人名义签订虚假债权转让合同]的手段”(引自徐赳赳2008:4,例9)。

³ 如“但[购买新电脑]的喜悦很快就被冲淡了”(引自徐赳赳2008:7,例29)。

⁴ 唐文(2007:144)的书面语料中共有404例含指量结构的关系从句,不论从句提取类型,指量结构前置的占76%。但若具体区分从句的提取类型,指量结构则呈不对称分布态势(同上:143)。

略,如歧义规避、尽早说出提示词等,来帮助听话者更容易地理解话语。应用于汉语关系从句,该理论预测,说话者为了给听话者理解加工关系从句提供线索,会先说出指量结构,即前置于主语关系从句(1a)。但在宾语关系从句中,前置的指量结构会给听话者带来:1)潜在歧义,如(2a)中的量词‘个’与邻近名词‘星探’和中心名词‘女孩’都匹配;或2)语义冲突,如“那扇足球打碎的窗户”,量词‘扇’与邻近名词‘足球’不匹配,而与中心名词‘窗户’匹配。因此,说话者会有意将指量结构后置于关系从句。简言之,该理论预测:主语关系从句中,指量结构倾向于前置;宾语关系从句中,指量结构倾向于后置。

综上所述,两大言语生成理论对指量结构与关系从句的位置分布做出了不同的预测。以下我们转入对口语语料库的分析,通过细分关系从句类型及从句内动词类别,考察指量结构与主、宾语关系从句共现的分布态势,并与上述的理论预测进行对照,以期探究内在原因。

3. 语料库分析:指量结构与关系从句的位置分布

本研究采用访谈节目《鲁豫有约》中的对话部分作为口语语料。语料收集了5卷主持人与各行业嘉宾访谈的录音文本材料⁵,共22期节目,约含309,848汉字。考虑到旁白部分可能源于节目前期策划或后期制作的书面语料,本研究未作分析。

3.1 分析方法

本研究采用吴芙芸(2011a)的方法,从语料库中人工提取了356例“空位关系从句”,均至少含有一个动词短语。按照以下五个特征进行标注:(i)有无指量结构;(ii)指量结构与关系从句的相对位置(前置、后置);

(iii)核心名词在主句中的语法位置(主语、宾语、介词宾语、系表、存现动词宾语);(iv)核心名词在从句中的语法位置(主语、宾语);(v)从句谓语类型(不及物动词、及物动词,其中及物动词又分为状态动词和动作动词)。

为了与前人的语料库研究和实时加工等实证研究具有可比性,本文从三个层面上分析语料。首先整体考察356例关系从句中指量结构的出现情况,然后分析128例含指量结构的关系从句中量词分布情况,最后关注60例谓语为及物动词的关系从句。

3.2 数据统计结果

3.2.1 有无指量结构分布

首先考察356例关系从句中是否出现指量结构。分析结果显示,不含指量结构的关系从句有228例,明显多于含有指量结构的128例关系从句($\chi^2_{(1, N=356)} = 28.09, p < .0001$)。这与前人的书面语语料库结果一致(Pu 2008; 吴芙芸 2011a),即量词与关系从句共现的频率不高。结合本文重点,以下讨论128例含有指量结构的关系从句。

3.2.2 指量结构在不同类型关系从句中的分布态势

表1给出了指量结构在128例关系从句中的位置分布情况。首先,在128例关系从句中,主语关系从句(共100例,占78.13%)显著多于宾语关系从句(共28例,占21.88%)($\chi^2_{(1, N=128)} = 40.5, p < .0001$)。这表明,蕴含指量结构的关系化也符合Keenan和Comrie(1977)的名词短语可及性等级。其次,指量结构的位置与关系从句的类型存在相关性,并呈不对称的分布态势:在100例主语关系从句中,指量结构倾向于前置(共85例,占85%, $\chi^2_{(1, N=100)} = 49, p < .0001$);在28例宾语关系从句中,指量结构倾向于后置(共20例,占71.43%, $\chi^2_{(1, N=28)} = 5.14, p < 0.05$)。这与前人的书面语料库结果一致(唐正大 2007; Pu

⁵ 5卷录音文本材料分别为《红伶》、《沉浮》、《男角》、《人气》和《开心果》,由新浪读书共享(<http://ishare.iask.sina.com.cn/>)提供下载。网络txt文本与在线对话内容完全相符,但无具体页码,故本文所引例句仅标明其来源节目档期的首播时间。

表 1 指量结构在 128 例关系从句中的位置分布情况

中心名词在主句中的成分	前置指量结构				后置指量结构				总计
	主语关系从句		宾语关系从句		主语关系从句		宾语关系从句		
	#	%	#	%	#	%	#	%	
主语	6	46.15	0	0	1	7.69	6	46.15	13
宾语	16	53.33	4	13.33	5	16.67	5	16.67	30
系表	38	69.09	1	1.82	9	16.36	7	12.73	55
介词宾语	17	80.95	3	14.29	0	0	1	4.76	21
存现动词宾语	8	88.89	0	0	0	0	1	11.11	9
总计	85	66.41	8	6.25	15	11.72	20	15.63	128

2008; Wu 2009; Ming & Chen 2010; 吴芙芸 2011a)。

本文重点关注中心名词在主句中作主语和宾语的情况(即修饰主语和修饰宾语的关系从句),共 43 例。如表 1 的前两行所示,指量结构的不对称分布趋势再次出现:在 28 例主语关系从句中,指量结构倾向于前置(共 22 例,占 78.6%),在 15 例宾语关系从句中,指量结构倾向于后置(共 11 例,占 73.3%)。

我们注意到,如表 1 最右列所示,关系从句的中心名词在主句中更多地充当系表(共 55 例,占 42.97%)、宾语(共 30 例,占 23.43%),而较少作主语(共 13 例,占 10.16%)。这有异于前人通过书面语料库得到的普遍分布,即以关系从句的中心名词在主句中作主语和宾语居多(Pu 2008; Wu 2009; 吴芙芸 2011a)。究其原因,我们认为主要源于语体间差异,具体的解释如下。

第一、本研究的口语语料来自于电视访谈节目,在主持人与访谈嘉宾面对面交谈的过程中,说话者为了让听话者更好的理解其所言所指,会更倾向于使用系表结构(见例 3-4),来限定、描述他(她)所指的人或事物,以便为听话者提供更多的背景知识(Clark & Brennan 1991)。

- (3) 他是[我那时候唯一喜欢 t_i 的]一个明星_{io}。
(《鲁豫有约:红伶》——赵薇,2006-06-08)
(4) 他特别逗,他说莫文蔚是一个[t_i 特别喜

欢买单的]人_{io}。

(《鲁豫有约:红伶》——莫文蔚,2007-05-23)

第二、在嘉宾访谈这种语境中,对话双方通常围绕某一人物或话题深入探讨或延伸展开,主语通常是先前提到的旧信息(Gernsbacher 1990; Givón 2001),或已为双方所知。因此更倾向于用主语作为前景信息,并结合指量结构,引入宾语(或表语),来表达新信息或新话题(Fox & Thompson 1990)。

第三、就信息加工过程中的言语感知难度(perceptual difficulty)而言,通常认为相对于旁置于主句左端的(修饰主句主语的)关系从句,嵌入主句中间的关系从句(即修饰系表和修饰主句宾语)中断了主句的构建、解析过程,对于说话者和听话者均可能会加重工作记忆的负荷,从而增加加工困难(Kuno 1974)。但是,口语(相对于书面语)的句子较短,而且,一旦使用了量词——特别是前置于从句的量词,意味着与其相匹配的名词即将出现。前置量词不仅有助于产出者有效地从工作记忆中提取尚未出现的核心成分,还有助于理解者渐进式地构建关系从句(Wu 2009)。的确,指量结构在修饰系表和修饰主句宾语的关系从句中均有前置的倾向,而在修饰主句主语的关系从句中,则无前置倾向。

3.2.3 指量结构在含及物动词关系从句中的分布态势

表 2 显示了指量结构在 60 例包含及物动词的关系从句中的分布态势。首先,在 60

例关系从句中,主语关系从句(共41例,占68.33%)显著多于宾语关系从句(共19例,占31.67%)($\chi^2_{(1, N=60)} = 8.067, p = .005$)。其次,指量结构的位置随着关系从句类型的变化呈不对称分布:在41例主语关系从句中,指量结构前置的共36例(占87.8%),指量结构后置的仅5例(占12.2%),前置倾向显著($\chi^2_{(1, N=41)} = 23.439, p < .0001$);而在19例宾

语关系从句中,指量结构后置的共13例(占68.42%),指量结构前置的仅6例(占31.58%),但由于基数较小,后置倾向未达到统计意义上的显著($p = 0.108$)。

同样,如果将修饰主句主语和修饰主句宾语的关系从句合并起来,考察指量结构在主语、宾语关系从句的位置取向,我们还是可以得到以上不对称分布态势。

表2 指量结构在含有及物动词关系从句中的分布态势

中心名词在 主句中的成分	前置指量结构		后置指量结构		总计
	主语关系从句	宾语关系从句	主语关系从句	宾语关系从句	
主语	4	0	1	3	8
宾语	5	4	3	5	17
系表	12	0	1	3	16
介词宾语	10	2	0	1	13
存在动词宾语	5	0	0	1	6
总计	36	6	5	13	60

3.3 讨论

3.3.1 不对称分布及原因

结合数据结果(表1和表2),我们不难看出,无论是否区分中心名词在主句中的成分或关系从句内动词类型,都得到了指量结构的位置根据关系从句类型变化而呈现不对称的分布态势:主语关系从句中,指量结构倾向于前置(5);宾语关系从句中,则倾向于后置(6)。这种分布态势符合基于听众设计的言语生成理论(见第二小节)的预期,但与基于说话者内部认知限制的“利己说”的预测不符。

(5) 主语关系从句,指量结构前置

比如说她给我打着打着困了,我一想,完了!这戏没吸引住一个[t_i要睡觉的]人_i!
(《鲁豫有约:沉浮》——邹静之,2006-02-08)

(6) 宾语关系从句,指量结构后置

其实我们看到[内地发行 t_i 的]那个电影版本_i是已经经过配音的了,如果要是本来你说的粤语,可能会更搞笑吧?
(《鲁豫有约:男角》——梁家辉,2006-06-20)

对于指量结构前置于主语关系从句的倾向,我们结合基于听众设计的理论,可做出三方面的解释。第一,从说者角度,就“长度顺序原则”而言,含有指示词的指量结构比关系从句更简短(Hawkins 1983),因此,说者会选择先说出指量结构。

第二,从方便听者角度,提前给出指量结构这个提示线索,可帮助听者预期性地期待一个与其相匹配的名词,进而构建目标句式即关系从句。

第三,无论从说者方便自身筹划言语角度,还是从方便听者理解言语角度,在主语关系从句中,由于位于句首的语法成分是动词(如例5“要睡觉”),而非名词,前置的指量结构与紧邻其后的动词之间不会产生句法歧义。

综上所述,主语关系从句中指量结构前置优于指量结构后置。尽管后者也合乎语法,但将指量结构紧邻中心名词(如例5中的“人”)、置于关系从句之后的同时,也失去了量词或指量结构对构建关系从句的句法提示

作用(Wu 2009)。

基于听众设计的理论亦可解释指量结构后置于宾语关系从句的倾向。在宾语关系从句中,虽然指量结构前置对说者而言可及性高,但是该构式可能会给听者带来两方面的困难:

1) 语义冲突。以(6)为例,前置量词(‘个’)只可以修饰中心名词(‘电影版本’),而不可以修饰毗邻的从句内名词(‘内地’)。在该语境下,‘内地’指的是大陆地区,与港澳台地区相对,且中国境内只有一个‘内地’,因此不能说“那个内地”。这种局部的量名不匹配会干扰词汇通达,从而引起听者理解困难。

2) 潜在歧义。仍以(6)为例,即使听者将前置的“那个”理解为指示词短语 DP 中的名词缺省(NP-deletion),并将句子分析为简单主句,即“其实我们看到那个 e...”,那么在听到下一个词‘内地’之后,听者必须放弃前面的组块,做句法回溯,并重新分析之前的成分,最终构建一个正确的宾语关系从句。

再进一步推之,如果前置量词既可以修饰从句内名词又可以修饰中心名词,如“小明捡起了那只足球打碎的花瓶”,仍会给听话者带来潜在的歧义。由于前置的指量结构直接与其后的从句内主语‘足球’构成一个指示词短语 DP,听者更容易误入“花园小径”(Frazier 1987),构建简单主句的结构(“小明捡起了那只足球”),而不是关系从句。

鉴于以上两种可能存在的句法处理过程和策略,可以推断,为了方便听者理解言语,说者在宾语关系从句中更倾向产出“指量结构后置”构式(如例 6 所示),以帮助听者规避潜在歧义和语义冲突。

3.3.2 宾语关系从句中的前置指量结构及原因

值得注意的是,在 28 个宾语关系从句中,指量结构前置的语例共有 8 个(占 28.57%)。这些例子似乎对基于听众设计的解释提出了挑战。通过具体考察这 8 个“例外”,我们发现,从句内的主语名词都与代词

有关,主要有三类,以下是具体实例:

(7) 从句内主语是人称代词“我”(3 例)

如若我突然从他口里听到一个 [我没演过 t_i 的] 角色 i ,那我肯定会觉得很兴奋,会觉得他仿佛把我带到另外一个世界,另外一个国度里头一般。

(《鲁豫有约:男角》——梁家辉,2006-06-20)

(8) 从句内主语是人称代词“自己”(4 例)

对对对,丰富了,自由了,然后能做一件 [自己一直想做而没有机会去做 t_i 的] 事情 i 。

(《鲁豫有约:人气》——黄健翔,2007-02-02)

(9) 从句内主语是人称代词“她”(1 例)

她突然好像很孤立的感觉,仿佛独自面对着一群 [她不认识 t_i 的] 人 i ,所以那时候我跟我太太经常跟她在一起。

(《鲁豫有约:男角》——梁家辉,2006-06-20)

在这些语例中,第一人称代词“我”和第三人称代词“她”均高度可及,反身代词“自己”无一例外均与说话人“我”同指,亦高度可及,因此会直接影响到言语生成的难易程度。首先,可及程度越高,加工越容易,特别是对于说者而言,更易于从工作记忆中提取与指量结构毗邻的从句内名词。

其次,前置量词在语义上需与无生命、非人类的(中心)名词匹配,虽然在这 8 个语例中,毗邻量词的是与之不匹配的(有生)人称代词,但其就生命度等级而言(Silverstein 1976),与暂时存储在工作记忆中的无生(中心)名词反差极大,因此在词汇通达上不会产生“基于相似性的干扰”(similarity-based interference, Gordon *et al.* 2004)。

再次,从句内的主语名词均为有生,而中心名词(即宾语)均为无生,有研究显示:主语倾向为有生而直接宾语倾向为无生(Comrie 1989);“主语有生、宾语无生”的生命度格局在含关系从句的汉语语料中分布最广(吴芙芸 2011b);而且具有该生命度格局的主语、宾语关系从句均容易加工(Wu *et al.* 2012)。这在一定程度上亦减少了由前置指量结构引起的词汇提取困难。

因此,我们认为,这8个前置指量结构并不会给听话者带来真正意义上的潜在歧义或语义冲突。

4. 结语

本研究基于真实口语语料《鲁豫有约》,考察了指量结构在关系从句中的分布态势。研究主要发现:在主语关系从句中,指量结构倾向于前置;在宾语关系从句中,指量结构倾向于后置。这种不对称分布态势也符合前人通过书面语(唐正大 2007;Pu 2008;Wu 2009;Ming & Chen 2010)、口语(Pu 2008)语料库研究得到的结果,因此,并不存在语体间的差异。

本研究对言语生成的理论有所启示。研究结果支持了言语生成中基于听众设计的理论,但与基于说者的“利己说”不符。对于指量结构在汉语关系从句中的不对称分布态势,基于听众设计的“利他说”可系统地予以解释:在主语关系从句中,说话者倾向于将指量结构前置,以提前给出关系从句构式的线索;在宾语关系从句中,说话者倾向于将指量结构后置,以便帮助听话者规避由前置指量结构带来的潜在歧义或语义冲突。

本研究也有一定的局限性。鉴于本研究使用的口语语料样本容量有限,所得到的分布态势和支持基于听众设计假说的证据尚需来自在线实时加工实验数据的检验,这有待于进一步的研究。

参考文献

- Arnold, J. E., T. Wasow, A. Asudeh & P. Alrenga. 2004. Avoiding attachment ambiguities: The role of constituent ordering [J]. *Journal of Memory and Language* 51: 55-70.
- Brennan, S. E. & H. H. Clark. 1996. Conceptual pacts and lexical choice in conversation [J]. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22: 1482-1493.
- Chao, Y. R. 1968. *A Grammar of Spoken Chinese* [M]. Berkeley: University of California Press.
- Clark, H. H. & G. L. Murphy. 1982. Audience design in meaning and reference [A]. In J. F. LeNy & W. Kintsch (eds.), *Language and Comprehension* [C]. Amsterdam: North-Holland Publishing Company, 287-299.
- Clark, H. H. & S. E. Brennan. 1991. Grounding in communication [A]. In B. L. Resnick, M. R. Levine & D. S. Teasley (eds.), *Perspectives on Socially Shared Cognition* [C]. Washington, DC: American Psychological Association, 127-149.
- Comrie, B. 1989. *Language Universals and Linguistic Typology* [M]. Chicago: University of Chicago Press.
- Dryer, M. S. 1992. The Greenbergian word order correlations [J]. *Language* 68: 81-138.
- Ferreira, V. S. & G. S. Dell. 2000. Effect of ambiguity and lexical availability on syntactic and lexical production [J]. *Cognitive Psychology* 40: 296-340.
- Fox, B. A. & S. A. Thompson. 1990. A discourse explanation of the grammar of relative clauses in English conversation [J]. *Language* 66: 297-316.
- Frazier, L. 1987. Sentence processing: A tutorial review [A]. In M. Coltheart (ed.), *Attention and Performance XII: The Psychology of Reading* [C]. Hillsdale, NJ: Lawrence Erlbaum Associates, 559-586.
- Gernsbacher, M. A. 1990. *Language Comprehension as Structure Building* [M]. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Givón, T. 2001. *Syntax: An Introduction* (Vol. II) [M]. Amsterdam: John Benjamins.
- Gordon, P. C., R. Hendrick & M. Johnson. 2004. Effects of noun phrase type on sentence complexity [J]. *Journal of Memory and Language* 51: 97-114.
- Greenberg, J. H. 1966. Some universals of grammar with particular reference to the order of meaningful elements [A]. In J. H. Greenberg (ed.), *Universal of Language* [C]. Cambridge, MA: MIT Press, 73-113.

- Hawkins, J. A. 1983. *Word Order Universals* [M]. New York: Academic Press.
- Keenan, E. L. & B. Comrie. 1977. Noun phrase accessibility and universal grammar [J]. *Linguistic Inquiry* 8: 63-99.
- Kuno, S. 1974. The position of relative clauses and conjunctions [J]. *Linguistic Inquiry* 5: 117-136.
- Ming, T. & L. Chen. 2010. A discourse-pragmatic study of the word order variation in Chinese relative clauses [J]. *Journal of Pragmatics* 42: 168-189.
- Pu, M. 2008. Pragmatics of classifier use in Chinese discourse [J]. *Stellenbosch Papers in Linguistics* 38: 139-163.
- Silverstein, M. 1976. Hierarchy of features and ergativity [A]. In R. M. W. Dixon (ed.). *Grammatical Categories in Australian Languages* [C]. New Jersey: Humanities Press, 112-171.
- Temperley, D. 2003. Ambiguity avoidance in English relative clauses [J]. *Language* 79: 464-484.
- Tsao, F. F. 1986. Relativization in Chinese and English: A contrastive study of form and function [J]. *Journal of Chinese Language Teachers Association* 11: 13-47.
- Wu, F. 2009. *Factors Affecting Relative Clause Processing in Mandarin* [D]. PhD Dissertation, University of Southern California.
- Wu, F., E. Kaiser. & E. Andersen. 2012. The animacy effects in Chinese relative clause processing [J]. *Language and Cognitive Processes* 27: 1489-1524.
- Zhang, N. 2006. Representing specificity by the internal order of indefinites [J]. *Linguistics* 44: 1-21.
- Zipf, G. K. 1949. *Human Behavior and the Principle of Least Effort* [M]. Cambridge, MA: Addison-Wesley.
- 唐正大, 2007, 关系化对象与关系从句的位置——基于真实语料和类型分析 [J]. 当代语言学 (2): 139-150.
- 吴芙芸, 2011a, 试论量名不匹配构式在语料库中的低频出现率及内在原因 [J]. 现代外语 (2): 136-142.
- 吴芙芸, 2011b, 基于经验还是基于工作记忆? 来自汉语新闻语料库中关系从句生命度格局的证据 [J]. 语言科学 (4): 396-408.
- 徐赳赳, 2008, 关系小句的语法和篇章特征分析 [J]. 汉语学习 (5): 3-8.
- 收稿日期: 2012-01-04;
作者修改稿: 2012-07-02;
本刊修订: 2013-03-25
通讯地址: 200083 上海市虹口区大连西路 550 号
338 信箱上海外国语大学语言研究院 5 号楼 605 室
<fywu@shisu.edu.cn> (吴)
<yanansheng1@gmail.com> (盛)

“语言与认知”国际学术研讨会征文通知

为推动语言与认知的研究与发展,促进国内外合作与交流,电子科技大学外国语学院、电子科技大学神经信息教育部重点实验室将于 2013 年 8 月 17-18 日在四川省成都市联合举办“语言与认知”国际学术研讨会。主题为“语言与认知研究——继承与发展”,下设四个分议题:语言与认知的理论研究、语言与认知理论的实验研究、认知语言学的应用性研究、语言与认知的哲学研究。

大会将特邀国内外语言与认知学界的著名专家进行主旨发言。会议工作语言为中文或英文。请与会者提交中英文摘要和论文。论文摘要提交截止日期:2013 年 5 月 30 日,全文提交截止日期:2013 年 6 月 30 日。论文提交邮箱:ICLC2013@126.com。相关信息请登录网站查询: <http://fl.uestc.edu.cn>。

联系人:李宜博 15882258854,李自恒 13688319053。

电话:86-28-61831162 传真:86-28-61831230

电子科技大学外国语学院

2013 年 4 月

学术论文中篇章性句干的型式和功能研究

张 乐¹, 卫乃兴²

(1. 上海交通大学外国语学院, 上海 200240; 2. 北京航空航天大学外国语学院, 北京 100191)

摘 要: 本文采取语料库驱动的短语学研究方法, 以 JDEST 学术英语语料库为文本证据, 分析篇章性句干的型式和功能。研究表明, 篇章性句干实施多种人际行为, 是专业文本中表达人际意义的重要手段。而且, 作为语言使用中的单一选择单位, 其文本意义的实现依赖于词语、型式、修辞结构、语域等诸多因素的共选。

关键词: 篇章性句干; 学术论文; 型式; 人际功能; 成语原则; 共选

中图分类号: H315 **文献标识码:** A **文章编号:** 1002-722X (2013) 02-0008-08

0. 引语

篇章性句干 (textual sentence stem, 以下简称 TSS) 指学术英语中表达话语社团认可的篇章意义和态度意义的高频半固定句级序列。TSS 是语言自我映射功能 (reflexivity) 的一种实现形式。(Hockett, 1960) 所谓自我映射, 就是用语言谈论语言本身, 有时也称为元话语 (metadiscourse)。日常交流中包含了难以计数的命题信息, 必须利用各种元话语手段, 对其加以管理、组织和评判, 用有效、合理的方式传递命题内容, 与听者或读者建立和睦的关系。

近年来, 对学术英语元话语特征的研究方兴未艾。这些研究表明, 学术英语绝非一种完全客观、不带个人感情色彩的语篇类型。恰恰相反, 科研人员在论文撰写过程中需要考虑如何在树立个人观点的同时显示出对同行的尊重。这一类语言手段在学术英语中俯拾皆是。过去的研究多集中于单词层级和短语层级的表达, 较少涉及小句层级的半固定语言形式; 多借助有限数量文本, 而非大规模学术英语语料库; 多以个人直觉来判断元话语的典型性, 较少借助频数信息来描述。

Sinclair (1991) 的成语原则 (Idiom Principle) 为研究半固定元话语手段提供了新的理论视角和方法。按照这一理论模型, 语言选择和使用的基本单位是词组, 是形式与形式、形式与意义、词汇与结构等共现组成的多词序列。半固定句级短语如何以整体形式实现学术文本中的人际功能, 应是短语学

的重要研究内容。本文采取语料库驱动的短语学研究方法, 讨论小句层级上表达人际意义的半固定短语序列。

1. 研究设计

1.1 篇章性句干的界定

TSS 属于 Pawley & Syder (1983) 提出的词汇化句干 (lexicalized sentence stem, 简称 LSS)。Pawley & Syder 将 LSS 界定为作为整体储存在记忆中、表达文化上许可的标准概念的半固定小句。本文所说的 TSS 专指学术英语中小句层级的元话语手段, 其界定标准为:

- 1) 形式上包含主语、谓语两个必要元素以及其他可选元素;
- 2) 在学术英语语料库 (500 万词) 中至少出现 10 次;
- 3) 表达特定的篇章意义和人际意义;
- 4) 具备一定程度的词语变异范围。

本文使用 3 个术语来表示不同抽象度的多词序列: 词汇序列 (lexical sequence)、型式 (pattern) 和 TSS。词汇序列是语料库中实际出现的形式, 文内以斜体表示; 型式是词汇序列的语法抽象, 主要由词类标签构成, 也可包含不变的单词; TSS 是词汇序列的意义抽象, 由属于同一型式的、表达相似意义的词汇序列构成, 文内以尖括号表示。

1.2 研究问题

本文探讨 TSS 的型式、意义和功能特征, 具体探讨以下 3 个问题:

收稿日期: 2011-06-13

作者简介: 1. 张 乐 (1982-), 男, 浙江舟山人, 上海交通大学外国语学院博士研究生, 研究方向为语料库语言学;
2. 卫乃兴 (1955-), 男, 河南济源人, 北京航空航天大学外国语学院教授, 博士, 研究方向为语料库语言学和词法学。

1) 学术英语中有哪些典型的实现人际功能的型式、TSS 和词汇序列?

2) 专业人员如何使用 TSS 以显示对同行的尊重、获得同行认可?

3) TSS 功能的实现涉及哪些语言特征的共选?

1.3 研究工具和方法

本文使用上海交通大学在建的 JDEST 学术英语语料库。该库涵盖来自 40 余个学科领域的论文、教材等多种文本类型, 总计约 650 万词。使用 Wordsmith 等工具可检索出隶属每个型式的高频词

汇序列, 便于语义分类及语用功能分析。我们首先将语料库文本切分为不同长度的 n -gram 序列 (多数为 2 词至 6 词序列), 统计频数后筛选出含有主谓结构的, 然后利用检索软件的通配符功能提取所有实施评价功能型式的高频词汇序列。

2. TSS 的型式特征

从型式上看, TSS 可分为两类: 显性主语 TSS 和隐性主语 TSS。显性主语又可分为内容主语和虚位主语, 隐性主语主要涉及祈使句。表 1 显示了 TSS 隶属的 19 个型式及其在 JDEST 中发生的次数。

表 1. 篇章性句干的主要型式

	型式	TSS 举例	词汇序列举例	频数
内容主语	(1) pro v that	< I > < think > that	<i>I think that; we believe that</i>	1397
	(2) pro v to-inf	< I > < want > to	<i>I want to; I would like to</i>	186
	(3) pro modal v that	< we > < can > < see > that	<i>we can see that</i>	222
	(4) n v	< this paper > < DESCRIBE > ^①	<i>this paper describes</i>	119
	(5) n v that	< TABLE > < SHOW > that	<i>Table1. 1 suggests that</i>	3643
	(6) n v-link that	<i>the < question > is that</i>	<i>the problem is that</i>	134
	(7) n v-link to-inf	< the objective > is to	<i>the objective of this paper is to</i>	105
	(8) n aux v-ed pca ^②	< RESULT > BE < presented > in / here	<i>the result is presented in Table3</i>	67
	(9) pca pro v	in < this paper > , we < DESCRIBE >	<i>in this paper we will discuss</i>	180
虚位主语	(1) it v-link adj that	<i>it is < important > that</i>	<i>it is important that</i>	925
	(2) it v-link adj to-inf	<i>it is < important > to</i>	<i>it is important to note that</i>	1216
	(3) it (modal) v-link that	<i>it < appears > that</i>	<i>it appears that</i>	322
	(4) it aux v-ed (by n) that	<i>it BE < found > that</i>	<i>it is found that</i>	414
	(5) it modal be v-ed that	<i>it < should > be < noted > that</i>	<i>it should be noted that</i>	248
	(6) as v-ed pca / by	<i>as < described > < above ></i>	<i>as mentioned above</i>	701
	(7) as aux be v-ed	<i>as will be < seen ></i>	<i>as will be discussed</i>	26
	(8) as pro (aux) v	<i>as we (< will >) < see ></i>	<i>as we noted earlier</i>	44
	(9) there (modal) v-link	<i>there < may > be</i>	<i>there may be; there might be</i>	319
隐性主语	imperatives	-	<i>note that , let us consider</i>	521

Hunston & Francis (2000) 认为, 任何一个型式中经常出现的词汇总是有限的。但是, 这些词并不一定都是近义词, 型式和意义也不是一一对应的关系。更为实际的办法就是将每个型式成分的具体实现形式划分为若干意义类别。比如, 型式 *it v-link adj that* 的具体实现形式可归纳为若干个表达不同意义的 TSS, 如 *it is < important > that* (*it is important/essential/critical/vital that*)、*it is < clear > that* (*it is clear/obvious/apparent/evident that*)、*it is < possible > that* (*it is possible/likely/unlikely/probable that*) 等。也就是说, TSS 是以语义特征为依据对词汇序列进行分组并高度概括后所得到的抽象。

3. 篇章性句干的功能特征

受韩礼德语言元功能三分法的影响, 对元话语手段附加“篇章功能”或“人际功能”的标签似乎成为一种通行做法。诚然, 任何元话语都或多或少地偏向于实现某种功能, 但这仅仅是一种程度的差别, 强行区分反而容易掩盖语言使用动机的复杂性。本文将篇章性句干视为人际意义的重要表达,

尽管它们同时也具有篇章功能。^③

按照“以言行事”(Austin, 1962) 的观点, 专业人员总是用一定的语言形式实施具体行为。通过检查 JDEST 文本证据, 我们认定 TSS 大致实施如下 8 个子功能, 即认知行为 (cognitive acts)、研究行为 (research acts)、推断行为 (inference acts)、评价功能、模糊限制 (hedging) 功能、指示行为、文际互文行为和文内互文行为。^④下面逐一探讨这些子功能的主要实现型式及其在维护人际关系方面所起的作用。

3.1 认知行为

认知行为指在研究中涉及到的心理过程, 如“知道”“相信”“理解”“期待”等。认知行为 TSS 在 JDEST 中共出现 1453 次, 大多用于表达对所述问题的一般看法。它们的必要成分是一组表达心理活动的动词, 如 THINK、BELIEVE、SUPPOSE、KNOW、WANT、SAY、SUGGEST、ARGUE 等。1453 个实例主要集中于 5 类抽象的 TSS 方式:

(1) < I > < think > that

- (2) <I> <want> to
 (3) <we> <can> <say> that
 (4) *it is* <assumed> that
 (5) *it* <can> *be* <argued> that

(1)、(2) 和 (3) 通过第一人称代词和心理动词的组合来实现人际功能。*I* 和 *we* 是学术论文中重要的自称手段。普遍认为, 学术文本中应当尽量减少作者的直接介入, 力求保持研究的客观性。(如 Hyland, 2001b: 208) 但也有一种看法认为不应完全否认自称在学术论文中的人际价值, 理由有三: 其一, 第一人称代词往往能够凸显作者的个人主张和创造力, 是展示自我价值的重要手段; 其二, 通过强调作者的主观看法, 暗示研究结论并非绝对真理, 而是多种可能性之一; 其三, 第一人称复数能够拉近作者与读者之间的距离, 表明双方属于同一个话语社团, 有共同的目标、学术背景和看问题的方式。(Hyland, 1996a: 258; 2001a: 559; 2001b: 223) 这些语言特征在人文学科文章中尤为明显。以 *I think*、*I believe* 和 *I want to* 为例, 它们出现在自然科学文本和社会科学文本中的概率分别为每 10 万词 3.1/0.7/0.8 次和 7.3/3.1/2 次。一般来说, 社会科学研究的对象变量复杂, 测量标准不一, 因此使用第一人称代词有助于建立自信的作者形象。相反, 自然科学通常有一套普遍认可的实证分析方法, 强调对变量的数量控制和精确测试, 因此过多的自称反而影响结论的可信度。(Hyland, 2001b: 215-216)

(4) 和 (5) 常被称为外位结构 (extraposition)。语料库证据表明, 以 *I*、*we* 和 *one* 为主语的 TSS 一般都有相对应的包含虚位主语 *it* 的 TSS。比如, *we hope that* 和 *it is hoped that* 的意义大体相同, 但后者表达的态度显然更加客观、中立。

3.2 研究行为

研究行为主要用于宣布研究过程和研究目的。研究行为 TSS 的界定在很大程度上得益于 Swales (1990) 提出的 CARS (Create a Research Space) 模型。Swales 发现, 学术论文的引语部分常由 3 个语步组成: 语步 1——确立区域 (确立研究话题的重要性)、语步 2——确立位置 (评述以往同类研究) 和语步 3——占领位置 (试图填补研究空白)。该模型中的每个语步都有其典型的语言范例 (linguistic exponent)。卫乃兴 (2002) 通过对语言范例的分析, 发现其中有大量词语搭配序列, 且不乏小句级别的惯用语搭配。数据表明, 语步 3 常常有如下 4 个 TSS 作为其典型语言范例:

语步 3A: 概述研究目的

- (6) <the objective> *is to*
 (7) <the purpose> *of* <this study> *BE to*
 语步 3B: 宣布研究特点
 (8) *in* <this paper> , *we* <DESCRIBE>
 (9) <this paper> <DESCRIBE>

研究行为 TSS 在 JDEST 中共出现 404 次。所有的词汇序列都包含“研究行为”动词, 如 EXAMINE、DESCRIBE、DISCUSS、OBSERVE、ANALYSE、INVESTIGATE、EXPLORE、PRESENT、REVIEW、IDENTIFY、ASSESS、ESTIMATE、CREATE 等。(6) 的主要实现形式为 *the objective/aim is to*, 用于说明研究目标或实验动机。*my/our objective* 很少出现, 这表明概述研究目的较少涉及作者自称。(7) 中的 <this study> 可以是 *this study*、*this paper*、*this article*、*this chapter*、*this section* 等。系动词 *BE* 大多使用一般现在时, 唯有 *this study* 通常和 *was* 搭配使用。

同样, (8) 中的 <DESCRIBE> 也有时态变化。一般现在时的动词形式仍居支配地位, 但是, 语料库中大约有一半左右的 *in this study we* 和动词的一般过去式共现。由此可以进一步肯定, 在 (7)、(8) 中, *this study* 常被置于历史语境中, 这和 Swales (1990: 160) 的发现基本一致。此外, *in this paper/chapter we* 也常和一般将来时的动词形式共现。

与 (8) 相比, (9) 以更为中立的方式来说明研究特征, 即前者是一种“塌陷结构” (collapsed structure), 后者是“标准描述形式” (standard descriptive form)。(Swales, 1990: 160) 科研人员根据自己对客观性和交互性的平衡把握, 在两者之间做出取舍。

3.3 推断行为

所谓推断行为, 就是在一定的科学理论和观测事实的基础上, 用缜密的论证过程来验证初始假设, 并报道最终研究结论。毫无疑问, 研究结论是否新颖并具有说服力, 直接影响到整个学术产品的质量。但是, 报道研究结论的方式对加强说服力所起的作用, 丝毫不亚于结论本身。推断行为 TSS 在 JDEST 中共出现 1691 次, 其中大部分包含一个表达“得到结论”的推断动词, 如 SEE、CONCLUDE、INDICATE 等。下面我们根据 TSS 的不同主语类型分别讨论。

首先, 与认知行为 TSS 相似的是, 推断行为 TSS 亦可包含一个指代作者或读者的人称代词。如:

- (10) *we* <SEE> that

(11) <we> <can> <see> that

(10) 的主要实现形式有 *we see/found/note/find/show/conclude that*, 均用于报道研究发现。

(11) 的主语除了 *we* 外, 还可以是 *you* 和 *one*, 如 *you will/can see/find that* 和 *one can see that*。第二人称代词和不定代词在学术论文中可用来预测、表达读者的看法, 但由于其带有口语化特征, 且不利于拉近作者与读者之间的距离, 故总体频数并不高。

其次, 以下3个虚位主语 TSS 也常常实施推断功能:

(12) *it* BE <found> that

(13) *it can be* <seen> that

(14) *it is* <clear> that

(12) 和 (13) 是 (10) 和 (11) 的非人称形式, 它们的高频词汇序列有 *it was/has been found that* 和 *it can be seen/shown that*。(14) 中的 <clear> 有4个词语变异, 按频数从高到低分别为 *clear*、*obvious*、*apparent* 和 *evident*。之所以将 *it is* <clear> that 视为推断行为 TSS 的一个重要原因是它常和表达“证据来源”的介词词组共现, 而这几乎是所有推断行为 TSS 的共同特征, 如 “*From the analysis, we can see that...*” 以及 “*It is clear from our data that...*”。

最后, 推断行为 TSS 也可由某些表达“结果”“证据”等意义的抽象名词做主语。如:

(15) <RESULT> BE <presented> in/here

(16) <RESULT> is that

(17) <RESULT> <SHOW> that

(15) — (17) 分别说明研究结果显示在文中的位置、结果的具体内容和结果的功能分析。它们的主语一般为 *results* 和 *findings*, 与之共现的推断动词有 PRESENT、SHOW、INDICATE、SUGGEST、DEMONSTRATE 等。

不难看出, (10) — (14) 表明作者相信读者有能力推敲、厘清论证中的逻辑关系; 相反, (15) — (17) 限定了证据的性质和出处, 提供了一个确切的参考背景。非人称名词主语也表明作者将自己的观点置于次要地位, 其言下之意是: 无论谁来进行此项研究都会得到相同的结果。它们减轻了作者对命题内容所需承担的责任, 加强了推断过程的客观性, 因此较多出现在研究程序固定、评判标准清晰的硬科学文章中。

3.4 评价功能

评价功能指研究人员对命题内容赋予一定的价值, 表达对命题或事件的态度或评价意义。评价功

能 TSS 共出现 789 次, 集中于以下3个 TSS:

(18) *it is* <interesting> that

(19) *it is* <difficult> to

(20) *the* <question> is that

(18) — (20) 在如下3个方面有别于其他 TSS。首先, 它们的主语或为 *it*, 或为某些抽象名词, 不涉及其他人称代词。其二, 这些 TSS 的评价对象由 *that* 从句、带 *to* 的不定式从句或 *wh*-从句 (如 *the question is how...*) 引出。其三, 3个 TSS 的核心成分都是负载评判价值的词汇 (value-laden lexis)。其最大特点就是可以将这些评价词按照积极、消极和中性等语义标准归类。譬如, *reasonable*、*interesting*、*easy*、*useful* 等词表达积极意义; *not clear*、*unfortunate*、*problem*、*question* 等词表达消极意义; *fact*、*idea* 等词语趋于表达中性意义。

正因为评价功能 TSS 传递鲜明的态度, 故常见于论文的引语部分, 主要用于指出以往研究的缺陷。例如:

(21) Since little work has been done on speech acts in non-Western contexts, *it is not clear* just what sort of universal generalizations can be made...

(22) Lastly, in view of possible technological applications in which artificially time-compressed speech is used, *it is interesting* to know which type of fast speech listeners actually prefer.

3.5 模糊限制功能

简言之, 模糊限制功能用于表达“或然性”和“不确定”, 一般通过模糊限制语来实现。由于实验设计不可避免地存在缺陷、例外情况和不确定因素, 因此研究人员提供的理论分析永远都只是诸多可能性中的一种。鉴于此, 学术论文中常常使用大量模糊限制语来折射这个并不精确的客观现实。Hyland (1996a; 1996b) 提供了一个包容广阔的模糊限制语描述体系, 涵盖词、短语、小句等多种语言手段以及各种类型的交际目的。本节仅聚焦于模糊限制语的核心功能, 对表达“不确定性”的下列 TSS 做一个简要介绍:

(23) *it is* <possible> that

(24) *it is* <possible> to

(25) *it* <appears> that

(26) *it* <may> be that

(27) *there* <may> be

(28) *it seems* <likely> that

(29) *it* <would> seem that

实现模糊功能的关键成分大体分3类。第1类

为表达“可能性”的表语形容词(组),频率较高的有 *possible*、*likely*、*impossible*、*not possible*、*unlikely*、*likely*、*probable* 等;第2类为情态动词 *may*、*might* 和 *would* 等;第3类为半情态动词 *SEEM* 和 *APPEAR* 等。值得注意的是,(28)和(29)包含了至少两类模糊限制手段。它们是多个模糊限制语的“和谐组合”(Lyons, 1977: 808),并可能具备特殊的语用功能,如卫乃兴(2002)发现, *it would seem that* 的主要功能之一就是指出以往研究的不足之处。

模糊限制 TSS 共出现 1199 次,它们主要有 4 种功能:1) 尊重读者,避免绝对化表述;2) 缓和命题与现实世界的冲突;3) 减轻作者对命题内容所负的责任,反映作者的自信度;4) 更精确地表述命题。(Hyland, 1996b) 尽管 Hyland 进一步区分了“面向作者的”和“面向读者的”模糊限制语,但其最终目的都是让读者接受研究结论,是学术论文中至关重要的人际手段。

3.6 指示行为

指示行为通过指引和诱导,使读者接受作者的意图与视角。指示行为实质上属于 Lyons (1977) 说的道义情态 (deontic modality), 表达“许可”“必要”“义务”等意义。学术英语中,用以引导读者思路的语言手段不胜枚举,其中指示行为 TSS 无疑最具表达力。指示行为 TSS 在 JDEST 中共计出现 1210 次,最常见的除祈使句外,还包括以下 4 类:

(30) *it* <should> *be* <noted> *that*

(31) *we should* <note> *that*

(32) *it is* <important> *to*

(33) *it is* <important> *that*

指示行为 TSS 中,祈使句是最直接的手段。较为常见的词汇序列有 *note that*、*let us consider*、*suppose that*、*let us look at*、*refer to*、*let us assume* 等等。多数动词表达“认知行为”,其作用是向读者说明研究程序和步骤,如(34);或是吸引读者注意力、确保他们对命题内容理解无误,如(35):

(34) *Let us take* the term rural as an example.

(35) *Note that* this approach gives a comparison between different levels of Ro and DR.

(30) 的常见词汇序列为 *it should/must be noted/reminded/emphasized that*。其中, *should* 的出现频数远高于 *must*。究其原因,主要是 *must* 向读者施加了过于沉重的道义压力,而这显然不符合“尊重读者”这一要求。同样道理,(31)的词汇序列总频数远低于(30),这也说明指示行为最好

由更为客观的虚位主语 TSS 来实现。

(32) 和 (33) 的核心成分为表达“重要性”“必要性”的一组形容词,主要有 *important*、*necessary*、*essential*、*crucial* 等。其共同作用是提醒读者注意研究中的必要步骤。例如:

(36) *It is important to* compare the mobility patterns of. . .

(37) *It is important that* considered factors not be limited to those that. . .

3.7 文际互文行为

上述 6 种功能的 TSS 涉及的命题内容大多为作者的一般观点、具体研究结果以及试图让读者加以关注的研究行为和现象。这些命题实质上都是作者本人所提出的主张,是个人研究成果的一部分。学术文章中还有大量命题另有来源,比如他人的观点和发现。Sinclair (1986) 明确区分了这两种不同性质的命题,他使用“断言”(averral)和“归因”(attribution)来分别指代作者主张和他人主张。“归因”展现的是文本之间的互相关系,是互文性(textuality)的一种表现形式。根据命题的不同来源,它又可进一步分为文际互文(intertextuality)和文内互文(intratextuality)。文际互文行为即一般所说的“引用”(citation)。恰当地引用他人论述能大大增强论证的说服力。因此,引用的手段和技巧一直是评价语研究的重要话题,而文际互文 TSS 在 JDEST 中的出现频数也高达 3363 次。根据引用的不同形式,Swales (1990) 区分了融入式引用(integral citation)和非融入式引用(non-integral citation)。在融入式引用中,被引作者的姓名作为主语或被动句式的施动者出现在句子中,分别称为“施动型引用”(agentative citation)和“非施动型引用”(non-agentative citation)。非融入式引用又分两种情况。第一种是通过括号、上标等形式标注被引文献的相关内容,可称为“识别型引用”(identifying citation);第二种则完全隐藏被引作者的身份信息,可称为“非识别型引用”(non-identifying citation)。因此,文际互文行为可细分为 4 类,即施动型融入式引用、非施动型融入式引用、识别型非融入式引用以及非识别型非融入式引用。

施动型融入式引用可用以下 TSS 表达:

(38) <author> <ARGUE> *that*

其中,<author>代表被引作者的姓名以及文献出版信息,<ARGUE>代表报道动词。该 TSS 的人际功能可从时态选择、主语选择、动词选择等 3 方面来分析。首先,报道动词的时态选择常常反映作者

对引文的态度。通过对(38)中最高频的14个报道动词的统计分析发现,心理动词常使用一般现在时,特别是CLAIM(81%)、POINT OUT(78%)、SAY(70%)、ARGUE(69%)、BELIEVE(66%)和SUGGEST(60.4%);相反,推断动词有相当大的概率使用一般过去时,特别是FIND(90%)、DEMONSTRATE(62%)、REPORT(54%)和NOTE(54%)。这说明,我们倾向于把心理动词所报道的一般看法同该看法的“普遍性”联系在一起。相反,推断动词所报道的具体研究发现常被置于历史语境中,当他人和自己的结论或假设互相冲突时,这种趋势更加明显。(Oster, 1981)

其次,有<author>并非被引作者的实际姓名,而是泛指一个群体的名词词组(如many scholars)。Jacoby(1987)曾将这种引用形式称为“总结”(summary),意味着被引命题内容是作者对部分学者共同观点的综述。这些词汇序列将以往研究同作者的研究互相对照,通过验证假设,或提供进一步支持,或将其辩驳推翻。此时,谓语大多为心理动词的一般现在时形式,这进一步证明作者格外强调以往研究的现实意义。

最后,不同的报道动词常常表达作者对引文的不同态度。CONFUSE、IGNORE、MISUSE等词语无疑是最明显的例证。但是,大多数高频报道动词都不表达这样泾渭分明的态度意义。例如:

(39) *Gilchrist and Williams (2001) argue that this catch up is more consistent with ... However, as we saw in the second empirical example on technological locking in the previous section, this does not seem to have been the case.*

上例中, *argue* 本身并不表达明显的消极意义。但是,当该动词被用来报道他人观点时,后面常常接一个以 *but*、*however* 等词语为首的句子,表明作者对引文的反对态度。这种反对态度即 Sinclair(2004)所说的“语义韵”(semantic prosody),它隐藏在动词中,只有通过足够多的索引行和足够大的语境才能被揭示。除了ARGUE以外, BELIEVE、CLAIM、MAINTAIN等少数几个报道动词也在不同程度上具备消极语义韵特征。

非施动型融入式引用的典型TSS为:

(40) *as <described> by/in <author>*

其中,常见动词有 *described*、*noted*、*suggested*、*stated* 等。以 *it* 为形式主语的词汇序列(如 *it has been pointed out by Thompson that*) 在语料库中极为少见,这可能是因为此类表述和(40)相比显得过

于冗长。

识别型非融入式引用常利用以下3个TSS:

(41) *<most> <recent> <STUDY> <SHOW>*

that

(42) *it BE <suggested> that*

(43) *there is <evidence> (to <believe>) that*

(41) 词语变异范围较广, <most>、<recent> 和 <STUDY> 的具体实现形式分别有 *most/other/several*、*recent/previous* 和 *studies/a study/research*, 且谓语多为研究动词和推断动词。该TSS一般用来交代当前课题的研究背景以及同以往研究的联系,也可用来指出他人的局限性,是文献回顾的重要语言手段。(42)和(43)的主要词汇序列有 *it has been suggested/argued/estimated that* 和 *there is evidence that*。总的来看,尽管这3类TSS本身不是引用形式,但它们常常预示着被引作者以括号形式出现在命题后。

非识别型非融入式引用常利用以下4类TSS:

(44) *it is <well known> that*

(45) *one <might> <argue> that*

(46) *it might be <argued> that*

(47) *it is <true> that*

隐藏被引作者的身份通常有两个原因。第一,当一个观点能够代表社团成员的普遍看法或为默认事实时,就无需明确提出和秉持这一观点的人的身份。实现该功能的典型词汇序列有 *it is (well) known/generally believed that* 等。第二个原因则与让步有关。让步是一种特别的文际互文行为,这一点已被许多语言研究者所公认。(如Hunston, 2000)在(45)、(46)和(47)中,被引作者是实际上并不存在的读者。that从句中的命题由虚拟读者提出,紧接着,作者向该命题提出异议。例如:

(48) *One might argue that if the problems have not yet led to conflict between states, they are unlikely to do so in the future. However, some pressures may intensify considerably in the future.*

此例涉及两个层面的评价:1)作者预测读者的看法(读者评价);2)作者表达对这一看法的消极评价(作者评价)。这是一种常见、实用的论证技巧,目的是为了引导读者逐渐认同自己的观点。

3.8 文内互文行为

文内互文行为用来指示作者前文已提及的或下文即将谈论的内容。以下4类TSS最常用于指示上下文,在JDEST中共出现714次:

(49) *as <shown> in <FIGURE>*

(50) *as* <described> <above>

(51) *as will be* <seen>

(52) *as we* (<will>) <see>

这4个 TSS 的动词变异相差无几。<shown>、<described>、<seen> 和 <see> 的具体实现形式均为表达“讨论”“提及”的一组动词，频率较高的有 SHOW、DESCRIBE、MENTION、DISCUSS、NOTE、INDICATE、SEE 等。(49) 回指前面的图表或章节；(50) 使用 *above*、*earlier*、*previously* 等副词来指示前文但不涉及具体位置；(51) 介绍即将谈论的内容；(52) 使用第一人称代词 *we*，但频数较低。总的来说，使用文内互文 TSS 可避免重复命题内容，强化篇章衔接，从而方便读者获取或重新审视上下文信息。

综上所述，学术写作作为一种社会行为，深受学术社团的传统以及科学文化的制约和影响。从某种意义上说，科研论文唯有得到同行认可才真正具备学术价值。对 45 个 TSS 的分析从各个角度揭示了 TSS 最根本的功能实现方式，即如何向读者呈现命题以直接影响读者对命题的接受程度。这些话术手段将读者视为一个学科团体的重要成员和追求科学真理过程中的积极参与者，视为具备丰富的科学知识、为学科发展做出贡献的专业同人，视为有权利就论文的某些方面予以反驳的审稿人员。在指出其他研究和观点存在的局限性的同时，也必须承认其价值，并利用具有隐性消极含义的词语手段，从而保护同行面子。

4. 成语原则、共选与 TSS

TSS 是学术文本中表达态度意义，在句级结构中因循组合使用的半固定词语序列。它们的核心特征体现为一定程度的结构固定性、语义稳定性和词语变异性，其实质也就是 Sinclair 所说的“半构筑短语” (semi-preconstructed phrases)。(1991: 110) 这些单一选择单位的意义实现离不开词语、型式、语域等诸多因素的共选。

1) 词语与词语之共选。当 TSS 中的某个成分发生变异时，其他成分常常也随之变化。据统计，在 <I> <think> that 中，当主语为 *I* 和 *we* 时，谓语大多分别为 *think* 和 *believe*；在 <we> <can> <say> that 中，与 *we* 高频共现的词语为 *can see*，而与 *I* 共现的词语多为 *would argue*。这些例子反映出词语之间的共现行为。

2) 词语与型式之共选。大多数型式中经常出现的词语是有限的。这些词语意义趋同，但也可以进一步划分为不同语义场。心理动词、研究动词、

推断动词等动词类别的划分即是一个典型的例证。从另一个角度来说，当多义词出现在特定型式中时，产生歧义的可能性大大降低甚至不复存在。因此，型式“挑选词语”，其实质是“挑选词义”。

3) 型式与功能之共选。当评价性词语高频出现在某些型式中时（如 *it* v-link adj that 和 *it* v-link adj to-inf），这些型式本身就具备了评价功能。Hunston & Sinclair (2000) 称之为“评价性型式” (evaluative pattern)，揭示了型式与功能的密切关联。

4) TSS 与其他命题内容之共选。许多 TSS 的使用常常预示了邻近小句表达的命题及作者对命题的态度。比如，非识别型非融入式引用所使用的多个 TSS 均预示着作者对被引命题的反对。由此可见，TSS 不仅表达态度意义，而且实现上下文的衔接和连贯。

5) TSS 与修辞结构 (rhetorical structure) 之共选。学术论文大致遵循 IMRD (Introduction-Method-Results-Discussions) 的修辞结构模型。(Swales, 1990) 研究表明，许多 TSS 实施的语篇功能和它们在文章中的位置有关。比如，词汇序列 *it is suggested that* 在 Introduction、Results 和 Conclusion 中分别实施回顾文献、呈现实验结果、总结全文等不同功能。

6) TSS 与语域之共选。Sinclair 将语域视为语言选择的重要决定性因素。(1991: 110) 一旦选定某个语域，语言的变异范围便大大缩小。TSS 是专业领域里特有的表达态度意义的词语-语法序列。不论是频数、词语变异范围还是具体的语篇功能，学术英语文本中的 TSS 和普通英语文本中的常见句级序列必然有较大不同。

这些共选关系本身又是相互联结的。常见的情况是，当选定了语域，也就选定了型式的高频词汇序列；当选定了词汇序列在文章中的位置，也就随之选定了搭配词和邻近小句表达的命题。可见，形式变异和语境密不可分。大量而有一定规律可循的词语变异正是新弗斯语料库语言学所揭示的短语的重要特征。

5. 结语

本文基于语料库证据，提供了一个用以描述半固定句级态度表达，即篇章性句干的型式和功能框架，共涉及 19 个型式、45 个 TSS 和 1 万余个词汇序列实例。研究发现，学术论文中存在大量小句层级的人际手段，揭示作者对于命题内容和对于读者所持的态度。它们是小句型式、评价性词语、修辞结构、语域等相互共选后形成的半固定意义单位，体现了成语原则在语言组织中的主导地位。45 个

TSS 大致实施 8 种语言行为, 从不同视角说明学术论文绝非一种“冷冰冰”的写作形式。为了使研究成果得以发表并得到读者首肯, 需以合适的方法邀请读者参与研究, 允许读者提出质疑, 认可其贡献, 指出其不足。这些学术交际策略常常相辅相成, 缺一不可。研究结果对学术英语写作及教学有一定的启示。学术论文中读者与作者的对话互动贯穿于论证过程的始终。因此, 态度表达能力的训练应成为 ESP 课堂中的重要教学内容。具有评价意义的单个词固然重要, 句级态度表达手段也不可忽视。掌握一定数量的 TSS 及其词汇序列, 能够大大提高元话语手段的多样性和丰富度, 以此达到说服读者的目的。

注释:

- ① 大写单词表示该词有时、数等屈折变化。
- ② *pca* 指地点情形状语 (place circumstance adverbial)。它们是表示“文中位置”的词语, 如 *above*、*below*、*in the next chapter* 等。
- ③ “篇章性句干”的含义是说其发生于篇章层面, 但不一定仅仅实施 Halliday 的篇章功能, 也可能实施人际功能。
- ④ 实际上, 它们实施的功能极多, 这 8 种仅为其中极小的一部分, 只是尝试性的认定和划分。

参考文献:

- [1] 卫乃兴. 词语搭配的界定与研究体系 [M]. 上海: 上海交通大学出版社, 2002.
- [2] Austin, J. L. *How to Do Things with Words* [M]. Mass.: Harvard University Press, 1962.
- [3] Hockett, C. Logical consideration in the study of animal communication [C] // W. Lanyon & W. Tavalga. *Animal Sounds and Communication*. Washington, DC: American Institute of Biological Sciences, 1960: 392–430.
- [4] Hunston, S. Evaluation and the planes of discourse: Status and value in persuasive texts [C] // S. Hunston & G. Thompson. *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press, 2000: 176–207.
- [5] Hunston, S. & G. Francis. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English* [M]. Amsterdam & Philadelphia: John Benjamins, 2000.
- [6] Hunston, S. & J. Sinclair. A local grammar of evaluation [C] // S. Hunston & G. Thompson. *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press, 2000: 74–101.
- [7] Hyland, K. Talking to the academy: Forms of hedging in science research articles [J]. *Written Communication*, 1996a, 13 (2): 251–281.
- [8] Hyland, K. Writing without conviction? Hedging in science research articles [J]. *Applied Linguistics*, 1996b, 17 (4): 433–454.
- [9] Hyland, K. Bringing in the reader: Addressee features in academic articles [J]. *Written Communication*, 2001a, 18 (4): 549–574.
- [10] Hyland, K. Humble servants of the discipline? Self-mention in research articles [J]. *English for Specific Purposes*, 2001b, 20 (3): 207–226.
- [11] Jacoby, S. References to other researchers in literary research articles [J]. *ELR Journal*, 1987, 1: 33–78.
- [12] Lyons, J. *Semantics* [M]. Cambridge: Cambridge University Press, 1977.
- [13] Oster, S. The use of tenses in reporting past literature [C] // L. Selinker, E. Tarone & V. Hanzeli. *English for Academic and Technical Purposes: Studies in Honor of Louis Trimble*. Rowley, Mass.: Newburg House, 1981: 76–90.
- [14] Pawley, A. & F. H. Syder. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency [C] // J. C. Richards & R. W. Schmidt. *Language and Communication*. London: Longman, 1983: 191–227.
- [15] Sinclair, J. Fictional worlds [C] // M. Coulthard. *Talking about Text: Studies Presented to David Brazil on His Retirement*. University of Birmingham, English Language Research, 1986: 43–60.
- [16] Sinclair, J. *Corpus, Concordance, Collocation* [M]. London: Oxford University Press, 1991.
- [17] Sinclair, J. *Trust the Text* [M]. London: Routledge, 2004.
- [18] Swales, J. *Genre Analysis* [M]. Cambridge: Cambridge University Press, 1990.

(责任编辑 张立飞)

大学生英语现在完成时习得的语料库考察^{*}

——检验“突显度”与“形式优先”假说的新证据

章柏成

(重庆交通大学外国语学院, 重庆 400074)

摘要: 二语现在完成时的习得难度大, 研究结论尚不一致。本研究基于取自 SWECCCL, CLEC 和 COLSEC 的四个微型语料库, 从形式与意义维度考察基础阶段大学生习得英语现在完成时的特征。研究结果与“突显度”假说相符, 但与“形式优先”假说相悖。本文提出, 外语环境下的二语时体习得具有系统性和动态性, 其形式与意义的习得顺序并非恒定不变, 课堂教学与二语时体项目难度、学习者个体差异、学习阶段及母语二语差异等多种变量交互作用, 动态地影响二语时体习得。

关键词: 二语时体; 现在完成时; 突显度假说; 形式优先假说; 动态系统理论

中图分类号: H0 文献标识码: A 文章编号: 1004-6038(2013)04-0067-05

1. 引言

二语时体习得的难度已被众多研究证实。Bardovi-Harlig(2000)认为, 时体是二语学习中非常困难的一项内容; Hinkel(2004: 5)发现, 学习者在多年二语学习以后, 其二语学术论文中的时体使用依然存在困难; Izquierdo(2009: 587)提出, 在非正规语境下, 二语情状体习得对母语为英语的法语学习者是一大挑战; Clahsen 等(2010)指出, 即便是高水平学习者, 在希腊语二语过去完成时习得方面仍有困难。在各种二语时体中, 现在完成时的习得难度尤其突出。Labadi(1990)发现, 英语现在完成时是日本学习者最困难的项目; Hantrakuil(1990)指出, 二语学习者使用现在完成时比一般现在时和一般过去时困难更大。

本研究运用语料库方法, 在形式与意义两个维度上, 考察基础阶段(大学一、二年级)大学生习得英语现在完成时的情况, 检验“突显度”与“形式优先”假说, 进一步讨论二语时体习得与发展。

2. 相关研究回顾

已有研究的关注点可大致划归“母语效应”、“使用错误”及“形式与意义”三个主题, 因此我们从这三个方面回顾文献。

2.1 母语效应

大量研究发现, 母语时体系统影响二语时体的理解与使用(Flashner, 1989; Hinkel, 1997; Coppeters, 1987)。Ayoun 和 Salaberry(2008: 583)提出, 由于法语复合过去式与英语现在完成时构成相似, 母语为法语的英语学习者过度使用英语现在完成时; Izquierdo 和 Collins(2008: 350)指出, 西班牙语和法语的时体标记相似, 有效促进母语为西班牙语的法语学习者使用法语动词时体标记; 邓小涛(2004)认为, 中国学生缺乏支持英语现在完成时使用的内部语境, 汉语语境知识因而

介入补缺, 并影响英语现在完成时的习得; 鲍瑞和金力(2010: 111)发现, 蒙古语用过去时表示英语现在完成时的时体概念, 因而蒙古族大学生易将英语现在完成时误用为一般过去时; 张晶(2005)指出, 中国学生习得英语现在完成时时, 母语影响随着学习者英语水平的提高而减弱; 蔡金亭(2010: 46)考察了“着”、“了”、“过”三个汉语体标记, 发现它们有助于中国学生的英语时态习得, 学习者倾向于将带“了”标记的汉语句式英译为完成时或过去时。

2.2 使用错误

Montrul(2009)发现, 在美国长大的, 母语为西班牙语的双语者, 使用英语二语时体和语气形态变化的错误率显著高于本族语者。有关中国学生使用英语现在完成时的错误研究有五项: 樊长荣和林海(2002)指出, 初、中级水平学习者常将现在完成时表示过去的时间/地点状语连用, 将终结性动词与表示一段时间的状态连用; 顾凯和王同顺(2005)基于 CLEC 的研究发现, 在 1078 个现在完成时实例中, 91 例(8.44%)有构成形式错误, 认为动词突显度、时间状语等语言因素均带来影响; 韩存新(2006)与王素芳(2008)发现, 中学生对现在完成时的构成形式掌握较好, 但在意义上容易与一般过去时等混用; 胡江(2010)对比了 CLEC 与 COLSEC, 发现学习者易将现在完成时用在一般现在时、一般过去时和过去完成时语境。

2.3 形式与意义

有研究发现, 学习者对二语时体的形式习得早于意义, 即“形式优先”假说(Klein, 1986; Bardovi-Harlig, 1992; 熊建国, 2003; 张雪梅, 杨滢滢, 2009)。也有研究(Schumann, 1987; Dietrich & Perdue, 1995)认为, 学习者二语时体系统大致沿“语用-词汇-语法”路径发展, 二语时体概念表达是“意义-形式”的演变过程。

在英语现在完成时的习得研究方面, 熊建国(2003: 68)

^{*} 基金项目: 本文是教育部人文社会科学一般项目“中国大学生英语现在完成时习得研究: 动态系统理论视角”(项目编号: 12XJ4740014)的部分成果。

作者简介: 章柏成, 教授, 博士, 研究方向: 基于语料库的语言研究, 二语习得, 教师话语与身份

采用完形填空测试,发现大学生对现在完成时的形式习得先于意义,支持“形式优先假说”;胡江(2010:55-57-58)也获得类似结论,并认为外语环境下的“课堂教学没有改变现在完成时形式与意义习得的自然顺序”。然而,该研究在 CLEC 中的发现却与“形式优先”假说相悖,只是并未对其进行分析讨论。

概言之,开展本研究的理据有三:首先,相关研究发现仍有冲突,“突显度”和“形式优先”假说的适用性尚需进一步检验;其次,基础阶段大学英语教学通常更加注重语法知识,考察这一群体的现在完成时习得,有助于认识课堂教学与二语时体习得的关系;最后,专门针对基础阶段大学生展开的二语时体习得研究极少。

3. 研究设计

3.1 研究问题

1) 基础阶段大学生习得英语现在完成时的形式与意义有何特征? 2) 这些习得特征是否支持“突显度假说”和“形式优先假说”? 3) 基于对这两个假说的检验,如何从不同视角认识学习者的二语时体习得与发展?

3.2 语料抽取

本研究语料取自 CLEC、COLSEC 和 SWECCCL 三个学习者英语语料库。首先从 SWECCCL(1.0) 的 WECCL 子库中选取所有一二年级学生的作文语料,共 244 篇,77857 单词。然后从其余三个语料(子)库中抽取规模相当的语料:从 SECCL 的即席讲话(Task B)语料中随机抽取 257 个文本;从 CLEC 的大学英语四级考试语料(ST3)中随机抽取 404 篇作文;从 COLSEC 的“学生-学生型自由讨论”语料中随机抽取 29 个大学英语四级口语考试语料(去除了考官的话语),由此得到四个观测语料库(见表 1)。

表 2 英语现在完成时的形式习得

专业类型	英语专业				非英语专业			
语料类型	口 语(SECCL)		笔 语(WECCL)		口 语(COLSEC)		笔 语(CLEC)	
形式状态	形式正确	形式错误	形式正确	形式错误	形式正确	形式错误	形式正确	形式错误
规则动词	28 (68.3%)	13 (31.7%)	58 (84.1%)	11 (15.9%)	23 (60.5%)	15 (39.5%)	65 (76.5%)	20 (23.5%)
不规则动词	89 (84.8%)	16 (15.2%)	140 (95.9%)	6 (4.1%)	36 (75.0%)	12 (25.0%)	131 (89.1%)	16 (10.9%)
分项小计	117 (80.1%)	29 (19.9%)	198 (92.1%)	17 (7.9%)	59 (68.6%)	27 (31.4%)	196 (84.5%)	36 (15.5%)
专业合计	315/361(87.3%)				255/318(80.2%)			
总体合计	570/679(83.9%)							

4.1.1 英语专业

总体上,英语专业学生共使用现在完成时 361 例,形式正确 315 例(87.3%)。在 361 个实例中,规则动词 110 例,形式正确 86 例(78.2%);不规则动词 251 例,形式正确 229 例(91.2%),不规则动词的形式习得正确率显著高于规则动词($Z = 3.191 > 1.960$, $P < .05$)。

分项来看,在口语语料中,现在完成时出现 146 例,其中,规则动词 41 例,形式正确 28 例(68.3%);不规则动词 105 例,形式正确 89 例(84.8%),不规则动词的形式习得正确率显著高于规则动词($Z = 2.014$, $P < .05$)。笔语中,共有现在完成时 215 例,其中,规则动词 69 例(正确率 84.1%),

表 1 本研究所用语料信息

类型	专业口语	专业笔语	非专业口语	非专业笔语
来源	SECCL	SECCL	COLSEC	CLEC
形符	77680	77857	77530	77162
类符	4030	5894	3765	5144
篇数	257	244	29	404

3.3 研究工具

除运用 AntConc 进行语料检索以外,本研究沿用 Woods 等(1986:182-184)的方法,对来自不同组别的两个比例进行比较:先用公式计算出 Z 值,然后与临界值(1.960)进行比较,判断差异显著性。由于 SPSS 中尚未包含此类 Z 检验程序模块,我们根据 Z 值计算原理,在 Excel 中编制出函数计算公式,得到一个“Z 值计算器”,可方便快捷地计算 Z 值,进而检验差异显著性。

3.4 检索与分类

首先,用 AntConc 分别在四个观测库中检索包含 have/has 及其缩写形式've/'s 的用例,并人工剔除 have/has 用作实义动词等不合格用例。然后,借鉴胡江(2010)的方法,将所有实例中的实义动词区分为规则动词与不规则动词,并从构成形式与使用意义两个维度进行分类统计。构成形式分为“形式正确”与“形式错误”;使用意义区分“意义正确”和“意义错误”。本研究中,“形式正确”指“have/has + 过去分词”完全正确的时体构成形式,“意义正确”就是在应该使用现在完成时的语境中使用了这一时体,“意义错误”则指其他时体误用为现在完成时的情况。

4. 研究结果

4.1 形式习得

基础阶段大学生对英语现在完成时构成形式的习得情况见表 2,下面分项报告结果。

不规则动词 146 例(正确率 95.9%),不规则动词的形式习得正确率远远高于规则动词($Z = 2.724$, $P < .05$)。

4.1.2 非英语专业

非英语专业学生口语语料中,总共检索到英语现在完成时 318 例,其中,规则动词 123 例(正确率 71.5%),不规则动词 195 例(正确率 75.4%),不规则动词的形式习得正确率略高于规则动词($75.4\% > 71.5\%$, $Z = .640$, $P > .05$)。

在口语语料中,现在完成时实例出现 86 次,其中,规则动词 38 例(正确率 60.5%),不规则动词 48 例(正确率 68.6%),不规则动词的正确率高于规则动词,差异不显著($Z = 1.205$, $P > .05$);笔语中,现在完成时出现 232 次,其中,

规则动词 85 例(76.5% 正确),不规则动词 147 例(89.1% 正确) 不规则动词的习得正确率显著高于规则动词($89.1\% > 76.5\%$ $Z=2.367$ $P < .05$)。

以上结果表明,学习者对不规则动词的形式习得水平高于规则动词,与国内胡江(2010)在 COLSEC 中的发现一致,支持“突显度假说”(Dulay & Burt,1974; Wolfram,1985; Bayley,1994),但与顾凯和王同顺(2005)以及胡江(2010)对 CLEC 语料的考察结论不符。

4.2 意义习得

表 3 显示:英语现在完成时用例共出现 679 例,其中,意义正确 611 例(90.0%)。下面分专业报告意义习得情况。

表 3 英语现在完成时的意义习得

	专业类型	英语专业		非英语专业	
	语料类型	口 语	笔 语	口 语	笔 语
意义正确	分项小计	122(83.6%)	204(94.9%)	65(75.6%)	220(94.8%)
	专业合计	326/361(90.3%)		285/318(89.6%)	
	总体合计	611/679(90.0%)			
意义错误	一般过去时	14(9.6%)	1(0.5%)	4(4.7%)	5(2.2%)
	一般现在时	4 (2.7%)	7 (3.3%)	15(17.4%)	5 (2.2%)
	过去完成时	6 (4.1%)	3 (1.4%)	2 (2.3%)	2 (0.9%)

4.2.1 英语专业

英语专业语料中,共出现现在完成时 361 例,其中,口语 146 例(122/83.6%) 笔语 215 例(204/94.9%) 笔语中的意义习得水平显著高于口语($94.9\% > 83.6\%$ $Z=3.379$ $P < .05$)。

意义使用错误方面,口语语料中的错误显著高于笔语语料($16.4\% > 5.2\%$ $Z=3.379$ $P < .05$),主要表现为将一般过去时、一般现在时和过去完成时误用为现在完成时,但误用的比例有差异。

4.2.2 非英语专业

非英语专业语料中,共有现在完成时用例 318 例,其中,口语 86 例(正确率 75.6%),笔语 232 例(正确率 94.8%),笔语中的意义习得水平显著高于口语($Z=4.775$ $P < .05$)。

现在完成时的意义使用错误在口语中有 21 例(24.4%),笔语中 12 例(5.2%),口语错误率显著高于笔语($Z=4.775$ $P < .05$)。误用的情形基本一致,按出错比例由高到低依次为一般现在时、一般过去时和过去完成时误用为现在完成时。

4.3 形式与意义习得比较

表 2 与表 3 中“总体合计”数据显示:形式习得总体正确率(83.9%) 低于意义习得总体正确率(90.0%) Z 检验差异显著($Z=3.261$ $P < .05$)。

英语专业学生的形式与意义习得正确率分别为 87.3% 和 90.3%,表明意义习得水平高于形式习得水平,但差异不显著($Z=1.160$ $P > .05$)。口语语料中,形式习得正确率为 80.1% 意义习得正确率为 83.6%,意义习得水平略高于形式习得水平;笔语语料中,形式习得正确率为 92.1%,意义习

得正确率为 94.9%,意义习得水平也略高于形式习得水平。

非英语专业学生的形式习得总体正确率为 80.2%,意义习得总体正确率为 89.6%,意义习得显著高于形式习得($Z=3.20$ $P < .05$)。口语语料中,意义习得正确率略高于形式习得正确率($75.6\% > 68.6\%$ $Z=0.853$ $P > .05$);笔语语料中,意义习得水平显著高于形式习得水平($94.8\% > 84.5\%$ $Z=3.497$ $P < .05$)。

概括起来,以上结果与胡江(2010)不一致,学习者对现在完成时的意义习得水平高于形式习得水平,不支持“形式优先假说”。

5. 发现与讨论

我们结合三个研究问题,依次展开分析讨论。

1) 基础阶段大学生习得英语现在完成时的形式与意义有何特征?

本研究对象习得英语现在完成时具有如下特征:形式习得方面,对不规则动词的习得水平显著高于规则动词,在笔语中的习得水平高于口语,英语专业学生的习得水平高于非英语专业学生。意义习得方面,笔语中的习得水平高于口语。口语中,英语专业学生的习得水平高于非英语专业学生;笔语中,两个专业学生的习得水平基本无差异。形式与意义习得的比较,无论是口语还是笔语,英语专业还是非英语专业,对现在完成时的形式习得水平都低于意义习得水平。

以上特征中,规则与不规则动词习得比较以及形式与意义的习得比较,分别涉及“突显度”和“形式优先”假说,将在下一部分讨论。在此,我们仅讨论另外两方面的习得特征:笔语中的习得水平高于口语;英语专业学生的口语习得水平高于非英语专业学生,笔语习得水平基本无专业差异。

首先,笔语是计划性产出,口语基本上属无计划性产出。相比之下,计划性产出时,学生有更多时间思考,并能通过视觉监控表达形式的准确性;而无计划性产出本身的即时性特征,限制了对表达形式与意义的监控,因而导致在习得水平上低于笔语。

其次,英语/非英语专业学生的习得水平比较,口语方面前者高于后者,笔语方面两者无差异。这也不难理解:英语专业的培养目标、课程设置以及教学实践,使学生获得更多口语训练,其总体口语产出水平较高;而笔语产出方面,非英语专业学生面临大学英语四级考试,师生在课堂教学及课外学习中十分重视词汇、语法等基础语言项目的训练。可以说,大学英语四级考试导向、课堂教学强化,以及学生课外努力三方面的共同作用,导致非英语专业学生笔语中的现在完成时习得水平接近英语专业学生。

2) 这些习得特征是否支持“突显度假说”和“形式优先假说”?

本研究发现,学习者对不规则动词的形式习得水平高于规则动词,与“突显度”假说相符;形式与意义习得水平比较,形式习得低于意义习得,与“形式优先”假说相悖。

根据“突显度”假说,不规则动词的形态变化更复杂,突显度更高,更易被感知,因而会更早被习得。胡江(2010)和

顾凯和王同顺(2005)指出,动词形态变化的记忆难度、任务类型以及课堂教学对规则动词变化的训练,导致与“突显度”假说相悖。本研究则认为,不规则动词形态变化复杂,师生高度重视并经常强化操练,频率效应提高了学生对不规则动词的敏感性。因此,不规则动词的突显性和课堂教学强化共同诱发了“突显度”效应。

“形式优先”假说方面,Bardovi-Harlig(1992)提出,自然环境下的二语时体形式习得先于意义。熊建国(2003)发现,外语环境下的英语时体形式习得水平高于意义习得,支持“形式优先”假说。该研究将目标语境中所有语法正确的时体形式都视为“形式正确”,如在本该使用“went”的语境里,“is going”也算“形式正确”。由此,“意义正确”考察一般过去时,而“形式正确”却考察所有的时体形式,两个变量的考察对象之间本身就是一对多的关系,“形式正确”的比例自然高于“意义正确”。本研究聚焦现在完成时(当且仅当构成正确的“have/has + 过去分词”形式才属“形式正确”),保证了考察对象的同一性和数据的可比性。

与其意义和用法相比,现在完成时的构成形式更为简单,习得该时体理应“形式优先”,为何会出现“意义优先”呢?我们试从负迁移和课堂教学两方面予以解释:首先,汉语无动词形态变化和主谓一致要求,受该特征的负迁移影响,学习者容易出现形态及一致性方面的错误,这与本研究的错误数据一致:主要是过去分词以及 have/has 与主语的一致性错误。其次,英语现在完成时难度大、习得晚,致使师生高度重视,课堂教学中的精心归纳、反复强调及频繁操练,提高了该时体的意义习得水平。简言之,汉语负迁移和英语课堂教学的双重影响,促使学习者对现在完成时的意义习得优于形式习得。

显然,以上对支持“突显度”假说和背离“形式优先”假说的分析中,课堂教学是一个共同的因素。因此,我们进一步讨论课堂教学与二语时体习得的关系。

尽管有研究认为课堂教学无助于二语习得(Krashen, 1985; Prabhu, 1987),但更多研究结论与此相反:Larsen-Freeman 和 Long(1991)认为,课堂教学能促进二语习得进程、速度和成效;White(1991)发现,课堂语法教学能显著提高口笔头产出的准确性;Ellis(1994: 623-627)指出,对于难度大的语法项目,课堂教学仅能促进其计划性产出的准确性。

本研究发现契合以上结论:其一,笔语中的习得水平高于口语,表明课堂教学对计划性二语产出的促进作用;其二,对不规则动词的形式习得水平显著优于规则动词,乃动词突显性和课堂教学的共同作用;其三,对现在完成时的意义习得优于形式习得,仍与课堂教学紧密关联。这些发现进一步显示,外语环境下的课堂教学可能是导致本研究中“突显度”效应和“意义优先”特征的重要变量,外语教师和学习者理应充分重视这一变量,有效构建教学环境,营造学习氛围,推动学生二语发展。

3) 基于对这两个假说的检验,如何从不同视角认识学习者的二语时体习得与发展?

根据学者们(Larsen-Freeman, 2002; De Bot et al., 2007; Ellis, 2008)从动态系统理论视角对二语发展的研究,二语习

得是一个复杂的动态系统,其运作受多种因素的交互影响,各因素对系统的作用大小难以简单判定,且并非呈简单的线性因果关系。因此,首先要认识到二语时体习得系统的复杂性:因素众多,各因素受到系统制约,也对系统整体产生难以预测的影响;其次,要认识到系统的动态性,“突显度”效应和“形式优先”特征同样具有动态性,此时的“突显度”效应可能表现为彼时的“非突显度”效应,此阶段的“形式优先”可能转化为彼阶段的“意义优先”。如果忽略二语时体习得的复杂性和动态性,将其发展简单概括为特定的顺序,可能难以揭示二语时体习得真面目。

我们认为,二语时体习得具有动态性和系统性,其影响因素多样,可能主要涉及母语与二语的差异、学习者个体特征、学习阶段及其特征、二语时体项目难度以及课堂教学等,这些因素交互作用,共同影响二语时体习得进程。因此,可以说存在一定时段、语境和条件下的“形式优先”或“意义优先”,而不宜将二语时体习得特征笼统地定论为“形式优先”或“意义优先”。这有别于前人研究中的过度概括性结论,似乎更能促进我们对二语时体习得的理解。

6. 结语

本研究所获证据支持“突显度”假说,背离“形式优先”假说,表明二语时体习得具有复杂性、动态性。我们认为,二语时体形式和意义的习得顺序与水平绝非常量,时体项目难度、二语和母语的差异、学习者个体差异、学习阶段以及课堂教学等因素交互作用、共同影响。换言之,外语环境下,可能并不存在恒定不变的二语时体形式与意义的习得顺序与水平状况。这一主张符合二语习得与发展的动态系统观:众多因素的影响以及整个系统的运动复杂多变、难以预测,对某一阶段特征的静态概括不能反映整个习得过程的动态性特征。为此,更多研究亟待开展,以深化我们有限的认识和理解。

本研究局限有二。第一,观察语料容量有限;第二,由于时间、精力和操作性方面的限制,本研究仅考察了学习者使用现在完成时的实例,而忽略了本该使用但实际上并未使用的情况,这也可能影响对学习者的二语习得英语现在完成时的总体概括。要弥补此缺陷,需对语料中的所有动词逐一排查,此项工作极为耗时费力,是未来研究需要考虑并解决的问题。

参考文献:

- [1] Ayoun, D. & M. Salaberry. 2008. Acquisition of English tense-aspect morphology by advanced French instructed learners [J]. *Language Learning* 3: 555-595.
- [2] Bardovi-Harlig, K. 1992. The relationship of form and meaning: A cross-sectional study of tense and aspect in the interlanguage of learners of English as a second language [J]. *Applied Psycholinguistics* 3: 253-278.
- [3] Bardovi-Harlig, K. 2000. *Tense and Aspect in Second Language Acquisition: Form, Meaning and Use* [M]. Oxford: Blackwell.
- [4] Bayley, R. 1994. Interlanguage variation and the quantitative paradigm: Past tense marking in Chinese-English [A]. In E. Tarone, S. Gass & A. Cohen (eds.), *Research Methodology in Second Language Acquisition* [C]. Hillsdale: Lawrence Erlbaum.

- [5] Clahsen, H., M. Martzoukou & S. Stavrakaki. 2010. The perfective past tense in Greek as a second language [J]. *Second Language Research* 4 (4): 501–525.
- [6] Coppeters, R. 1987. Competence differences between native and near-native speakers [J]. *Language* 4 (3): 544–573.
- [7] De Bot, K., W. Lowie & M. Verspoor. 2007. A dynamic systems theory to second language acquisition [J]. *Bilingualism: Language and Cognition* 1 (1): 7–21.
- [8] Dietrich, R. & C. Perdue. 1995. Introduction [A]. In R. Dietrich, W. Klein & C. Noyau (eds.). *The Acquisition of Temporality in a Second Language* [C]. Amsterdam & Philadelphia: John Benjamins.
- [9] Dulay, H. & M. Burt. 1974. Natural sequences in child second language acquisition [J]. *Language Learning* 1 (1): 37–53.
- [10] Ellis, N. 2008. The dynamics of second language emergence: Cycles of language use, language change and language acquisition [J]. *The Modern Language Journal* 2 (2): 232–249.
- [11] Ellis, R. 1994. *The Study of Second Language Acquisition* [M]. Oxford: Oxford University Press.
- [12] Flashner, V. 1989. Transfer of aspect in the English oral narratives of native Russian speakers [A]. In H. Dechert & M. Raupach (eds.). *Transfer in Language Production* [C]. Norwood: Ablex.
- [13] Hantrakuil, C. 1990. *English Tense and Aspect Usage in Controlled Written Discourse by Non-native Speakers* [D]. Bloomington-Normal: Illinois State University.
- [14] Hinkel, E. 1997. The past tense and temporal verb meanings in a contextual frame [J]. *TESOL Quarterly* 2 (2): 289–313.
- [15] Hinkel, E. 2004. Tense, aspect and the passive voice in L1 and L2 academic texts [J]. *Language Teaching Research* 1 (1): 5–29.
- [16] Izquierdo, J. 2009. L'aspect lexical et le développement du passé composé et de l'imparfait en français L2: Une étude quantitative auprès d'apprenants Hispanophones [J]. *Canadian Modern Language Review* 4 (4): 587–614.
- [17] Izquierdo, J. & L. Collins. 2008. The facilitative role of L1 influence in tense-aspect marking: A comparison of Hispanophone and Anglophone learners of French [J]. *The Modern Language Journal* 3 (3): 350–368.
- [18] Klein, W. 1986. *Second Language Acquisition* [M]. Cambridge: Cambridge University Press.
- [19] Krashen, S. 1985. *The Input Hypothesis: Issues and Implications* [M]. London: Longman.
- [20] Labadi, Y. 1990. *A Contrastive Study of Tense and Aspect in English and Japanese: Pedagogical Implications* [D]. Austin: University of Texas.
- [21] Larsen-Freeman, D. 2002. Language acquisition and language use from a chaos/complexity theory perspective [A]. In C. Kramsch (ed.). *Language Acquisition and Language Socialization: Ecological Perspectives* [C]. London: Continuum.
- [22] Larsen-Freeman, D. & M. Long. 1991. *An Introduction to Second Language Acquisition Research* [M]. London: Longman.
- [23] Montrul, S. 2009. Knowledge of tense-aspect and mood in Spanish heritage speakers [J]. *International Journal of Bilingualism* 2 (2): 239–269.
- [24] Prabhu, N. 1987. *Second Language Pedagogy* [M]. Oxford: Oxford University Press.
- [25] Schumann, J. 1987. The expression of temporality in Basilean speech [J]. *Studies in Second Language Acquisition* 1 (1): 21–41.
- [26] White, L. 1991. Adverb placement in second language acquisition: Some effects of positive and negative evidence in the classroom [J]. *Second Language Research* 2 (2): 133–161.
- [27] Wolfram, W. 1985. Variability in tense marking: A case for the obvious [J]. *Language Learning* 2 (2): 229–253.
- [28] Woods, A., P. Fletcher & A. Hughes. 1986. *Statistics in Language Studies* [M]. Cambridge: Cambridge University Press.
- [29] 鲍瑞, 金力. 2010. 蒙古语时态对蒙古族大学生学习英语现在完成时的影响实证研究 [J]. 内蒙古师范大学学报(教育科学版), (11): 111–115.
- [30] 蔡金亭. 2010. The influence of Chinese aspect marker *le* on the simple past marking in English interlanguage: An empirical study of university students in China [J]. 中国应用语言学 2 (2): 35–47.
- [31] 邓小涛. 2004. 语境支持与二语形式学习——中国学生英语现在完成时习得研究 [D]. 广州: 广东外语外贸大学.
- [32] 樊长荣, 林海. 2002. 中国学生英语时体习得中的两大“误区” [J]. 外语教学与研究 6 (6): 414–420.
- [33] 顾凯, 王同顺. 2005. 语言因素对英语现在完成时使用的影响 [J]. 现代外语 1 (1): 53–60.
- [34] 韩存新. 2006. 中国学生英语写作中的现在完成时错误分析——基于 CLEC 的研究 [D]. 武汉: 华中科技大学.
- [35] 胡江. 2010. 对大学生习得英语现在完成时的语料库研究 [J]. 解放军外国语学院学报 6 (6): 55–60.
- [36] 王素芳. 2008. 基于语料库的情状体和语篇结构对现在完成时使用的影响 [D]. 武汉: 中国地质大学.
- [37] 熊建国. 2003. 英语专业学生英语时体的形式和意义习得研究 [J]. 解放军外国语学院学报 1 (1): 68–72.
- [38] 张晶. 2005. 论汉语时体特征对中国学生习得英语现在完成时的影响 [D]. 大连: 大连海事大学.
- [39] 张雪梅, 杨滢滢. 2009. 英语专业学习者的时态习得现状——一项基于中国英语专业写作语料库的研究 [J]. 外国语 3 (3): 133–138.

Abstract: L2 present perfect tense acquisition is notoriously difficult and the research on this issue has not got consistent findings. Based upon the four micro-corpora extracted from SWECC, CLEC and COLSEC, the present study explores the freshmen and sophomores' formal and semantic acquisition of English present perfect tense. Research findings support the Saliency Hypothesis but contradict the Form-acquisition-priority Hypothesis. Therefore the study tentatively proposes that L2 tense and aspect acquisition is systematic and dynamic in foreign language contexts, and there is not a stable/unchangeable order. The acquisition order and proficiency of L2 tense and aspect is dynamically influenced, shaped and even decided by the interaction of multidimensional factors including classroom teaching, the difficulty of certain L2 tense and aspect items, learner differences, learning stages, L1-L2 differences and some others.

Key Words: L2 tense and aspect; present perfect tense; Saliency Hypothesis; Form-acquisition-priority Hypothesis; Dynamic System Theory

中国英语学习者写作产出词汇的心理语言属性发展趋势 —— 一项基于 WECCL 和 MRC 数据库的研究

解放军外国语学院 李 海 张淑静

提要:本文以英国 MRC 数据库作为参照,借助中国学习者笔语语料库,调查中国二语学习者写作产出词汇的具体性、意象性、意义联系度和熟悉度四种心理语言属性的跨年级发展趋势。研究表明,从一年级到四年级,学习者产出词汇的具体性和意象性逐渐降低;意义联系度总体有所下降,但趋势不明显;词汇熟悉度先降低,但大二之后又逐渐提升。本研究表明,将 MRC 数据库引入二语词汇研究是可行的,且有利于二语词汇习得研究的深入。

关键词: WECCL、MRC 数据库、词汇的心理语言属性、二语词汇发展

[中图分类号] H313 **[文献标识码]** A **[文章编号]** 1003-6105(2013)02-0174-08

1. 引言

本文从词汇本身的心理语言属性出发,基于 WECCL 语料库调查中国英语学习者写作产出词汇的发展趋势。产出性词汇一直是二语词汇研究的热门话题,相关文献层出不穷,其中比较有代表性的如李志雪、李景泉(2005)、文秋芳(2006)、谭晓晨(2006)、杨世登(2007)、卢敏(2008)、童淑华、高金花(2010)、吴瑾(2011)等。这些研究都采用了各类测量工具考察了不同类型学习者的口语或笔语产出性词汇发展状况。以往研究对二语产出词汇的测量工具主要包括词汇精确度、词频和词汇多样性(Polio 2001),测量口语词汇的流利性、词汇多样性和词频广度(Ellis & Barkhuizen 2005; McCarthy & Jarvis 2010; 文秋芳 2006),以及词汇概貌(Laufer & Nation 1995; 林峻 2009)等。这些测量工具主要涵盖了词汇多样性、流利度和词频,在一定程度上也反映了词汇的难度和可学性。然而词频、丰富度等上文提及的属性只是衡量词汇难度和可学性的外围属性,仅仅是对学习者词汇使用状况的描述,未完全深入词汇内部更全面地调查词汇自身属性。词汇自身属

性更可能影响难度和可学性,如本文讨论的意象性。

本文考察的心理语言属性正是词汇本身的属性,直接影响词汇的心智加工和可学性(Salsbury 2011)。词汇的心理语言属性是影响难度和可学性的本质属性。以往研究,如熊永红(2011)等,可能涉及过类似概念,但未明确提出该属性,很可能因为未找到相应描述工具作参照。幸运的是,我们最近找到了描述词汇心理语言属性的工具——英国医学研究理事会心理语言学数据库。MRC 心理语言学数据库(Medical Research Center Psycholinguistic Database)是英国医学研究理事会研发的一个可用于心理学、语言学等多种领域研究的机读词表。该数据库可用作语言学或心理学研究者的实验刺激选词,也可作为人工智能和计算机科学等研究者提供所需词汇的语言学或心理学属性描述(Wilson 1988)。最新的 2.0 版本收录了 150,837 个英文单词、词根和词缀(下文统称为形态成份)且描述了这些形态成份 26 种常见的语言学和心理学属性(如重音模式、音标转写、词性、具体性、意象性等),并依据大量本族语者对这些成份一些属性的评分,确定了属性

值,方便定量研究。

本研究仅关注其中列出的四个心理语言属性:具体性(concreteness)、意象性(imageability)、意义联系度(meaningfulness)和熟悉度(familiarity)

2. 词汇的心理语言属性

词汇的心理语言属性在母语词汇习得中已得到广泛调查(如 Paivio *et al.* 1968; Toglia & Battig 1978; Gilhooly & Logie 1980),而用于二语词汇习得的研究较少。相当数量的母语习得研究考察过词汇的心理语言属性如何影响词汇习得和加工。例如,英语母语习得研究者一致认为词汇的具体性可分不同等级;例如“occasion”远不如“book”具体,且认为具体词比抽象词更易被加工和习得(Paivio *et al.* 1968; Toglia & Battig 1978)。其他心理语言属性,如意象性、意义联系度和熟悉度也有类似描述。笔者认为,虽然一语与二语的词汇习得存在一定差异,但制约母语者习得英语的词汇属性源于词汇本身,很可能对二语者的词汇习得产生类似影响,因此将词汇的心理语言属性用于二语词汇研究也是有必要的,然而我们很少能找到调查二语词汇心理语言属性的研究。有鉴于此,本文以MRC数据库提供的心理语言属性为参照,考察中国大学生英语写作词汇的心理语言属性发展特征。

2.1 具体性

词汇的具体性是指一个词的具体或抽象的程度,即一个单词指称的实体,或其蕴含的观点和概念的可达程度(Paivio *et al.* 1968; Toglia & Battig 1978; Gilhooly & Logie 1980)。MRC数据库中词汇的具体性分值源于对词汇信息先前的研究(Paivio *et al.* 1968; Toglia & Battig 1978; Gilhooly & Logie 1980)。先前研究要求本族语者对词汇的具体性评分,分1到7七个等级,意义越具体的单词,具体性分值越高。一个单词的具体

性分值是本族语受试评分的平均值。指称物体、材料和人的词汇的具体性分值更高(Toglia & Battig 1978)。词汇心理加工的文献中一个重要的研究成果就是:意义越具体的词汇越容易加工;在词汇回述、词汇辨别、词与非词判定、词汇理解等任务中,具体词比抽象词有更多优势(Gee *et al.* 1999)。同样在二语的词汇习得中,意义更具体的词比抽象词更早被习得(Crossley *et al.* 2009),也更容易习得(Ellis & Beaton 1993)。

2.2 意象性

意象性是指对某一词汇构建心智意象的难易程度。简单来说,就是看到或听到某词时,大脑是否容易产生涉及该词汇场景的意象。MRC数据库中描述意象性的分值也源于先前的研究(Paivio *et al.* 1968; Toglia & Battig 1978; Gilhooly & Logie 1980)。意象性的评分方法与具体性相似,根据大量本族语受试的判定,取平均值。构建心智意象越容易的词汇意象性分值就越高。如,“buffalo”比“relevant”更易构建心智意象,于是意象性分值更高。

值得注意的是,意象性和具体性两个属性容易混淆。一般来说,单词意义越具体,越容易构建心智意象,但容易构建心智意象的单词未必意义很具体。例如accident和marriage两个词,意义并不很具体,但意象性分值却很高,因为看到这两个词,受试很容易快速想到某经历过的情景。意象性分值之所以重要,是因为越易唤起学习者某特定意象或情景的词汇,回述出来越容易。Ellis和Beaton(1993)也发现,对于意象性强的单词,受试更容易联想到其使用的语境(即“语境可及性”强)。二语词汇习得时,可以以意象性强的单词为依托,将其作为“关键词”,辅助习得其他单词。故意象性对于词汇的加工和习得极为重要。

2.3 意义联系度

意义联系度是描述一个单词与其他单词发生语义联系多少的属性特征。意义联系度

高的单词,是构建心理词汇网络的关键(Ellis & Beaton 1993),为词汇网络的组织,和相关词汇的记忆提供可靠的中介。

意义联系度分值的确定参考了 Paivio *et al.* (1968) 与 Toglia 和 Battig (1978) 两个范式,但两者的关联度不高(相关系数只有 0.53)。此外,Paivio *et al.* (1968) 的意义联系范式在数据库中只对应了 760 个单词,而 Toglia 和 Battig (1978) 中的 Colorado 范式在数据库中对应了 2628 个单词。考虑到后者可以包含更多的单词,本文与 Salsbury *et al.* (2011) 的做法相同,采用 Colorado 意义联系范式。

2.4 熟悉度

熟悉度其实就是词频。高频词往往更熟悉,低频词则更陌生。MRC 数据库中熟悉度分值的确定依旧参照了 Paivio *et al.* (1968)、Toglia 和 Battig (1978)、Gilhooly 和 Logie (1980) 等研究。根据经验,词频越高的词汇越早习得。我们初步预测,随着学习阶段的提高,产出的低频词可能也会增多。

3. 研究设计

3.1 研究假设

本文考察二语者写作产出词汇的心理语言属性随年级变化的趋势,并提出如下假设: 1) 由于意义具体的词更易且更早习得,于是随学习阶段的提升,学习者会产出更多意义抽象的词,总体具体性分值会下降; 2) 意象性强的单词更易掌握且习得更早,因此随着学习阶段的提升,学习者会产出更多不易构建心智意象的词汇,总体意象性分值会下降; 3) 意义联系度高的词汇是习得其他词汇的中介,因此会更早习得;于是随着学习阶段的提高,产出词汇的意义联系度分值会总体下降; 4) 高频词往往先习得,因此学习者随学习阶段的提高会产出更多低频词,使写作产出词汇的熟悉度会总体降低。

3.2 实验材料

本研究的语料是“中国大学生笔语语料

库”(WECCL) 中一到四年级的议论文共 246 篇,其中一年级 64 篇,二年级 60 篇,三年级 62 篇,四年级 60 篇。这些议论文共分两个题目:“Education as a Lifelong Process”和“The Impact of Internet”,各年级的作文中这两个题目各占一半。为尽可能降低题目本身对产出词汇性质的影响,本文将两个题目的作文都列入调查之列。

3.3 实验工具

本文的实验工具除 MRC 数据库外,还有 Wordsmith 5.0、Excel 和 SPSS 13.0。Wordsmith 主要用于生成各年级的作文词表,并使用 MRC 中包含的四种心理语言属性对应的单词作为匹配词表,找出各年级的词表和 MRC 中四个心理语言属性词表互相重合的词汇。使用 Excel 则是利用 VLOOKUP 函数,找到重合词汇对应的四种心理语言属性值。最后把每种心理语言属性对应下的各年级的重合词汇及其属性值导入 SPSS 中,分四次进行单因素方差分析。

4. 数据分析与结果

4.1 WECCL 与 MRC 数据库重合的词汇

笔者首先利用 wordsmith 生成 WECCL 中各年级的议论文类符词表,随后将各年级的类符词表与 MRC 数据库中各心理语言属性对应的单词做匹配,匹配结果如表 1 所示。

最后一列“重合词占语料比例”是指,每种属性每个年级匹配的所有形符占该年级作文总形符数的百分比。例如,大一具体性的重合词占语料比例为 76.36%,是指大一所有(64 篇)作文中有 76.36% 的单词与 MRC 数据库的具体性属性对应词汇重合。总体来看,尽管 WECCL 和 MRC 数据库各心理语言属性对应单词的类符重合度较低,但形符重合度较高:65% 至 80%,因此用 MRC 数据库中的心理语言属性来描述学习者语料是可行的。

表1 WECCL 和 MRC 数据库重合的词汇

心理语言属性	MRC DB 中对应属性的单词频数	年级	WECCL中的类符数	MRC与 WECCL类符匹配数	匹配类符占 WECCL类符的比例(%)	重合词占语料的比重(%)
具体性	4294	大一	2253	782	34.71	76.36
		大二	2773	861	31.05	73.69
		大三	2544	760	29.87	72.60
		大四	2479	745	30.05	71.37
意象性	4827	大一	2253	899	39.90	78.97
		大二	2773	1009	36.39	76.50
		大三	2544	896	35.22	75.49
		大四	2479	870	35.09	74.19
意义联系度	2627	大一	2253	603	26.76	71.23
		大二	2773	621	22.39	67.71
		大三	2544	559	21.97	66.43
		大四	2479	554	22.35	64.64
熟悉度	4922	大一	2253	899	39.90	78.97
		大二	2773	1011	36.46	76.51
		大三	2544	897	35.26	75.50
		大四	2479	872	35.18	74.20

4.2 四种心理语言属性的变化趋势

4.2.1 具体性的变化趋势

各年级“具体性”均值走势如图1所示:一至四年级,此均值总体下降。据此可认为此结果证实了假设一:作文产出词汇的具体性随学习阶段提升而降低。单因素方差分析表明,年级间差异尚不显著($F=2.228, p=.083$),表2中跨年级间的多重比较也均无显著差异($p>.05$)。于是,我们可认为一至四年级的作文用词的具体性整体下降,但未发展到显著的程度。

4.2.2 意象性的变化趋势

四个年级“意象性”的均值走势如图2所示:从一年级到四年级,意象性的均值总体呈下降趋势,且单因素方差分析表明各年级间总体上有显著差异($F=3.467, p=.016$)。表3中跨年级间均值多重比较表明,一和四年级有显著差异($p=.036$),而其他各年级间则否($p>.05$)。据此,结果基本支持假设二:作文产

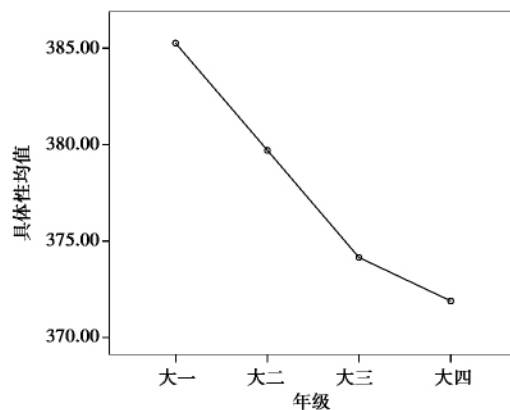


图1 各年级的具体性均值走势

出词汇的意象性会随学习阶段提升而降低。

4.2.3 意义联系度的变化趋势

四个年级的意义联系度均值如图3所示:从一年级至三年级,该均值总体下降,但三到四年级却有所回升。同时单因素方差分析表明,各年级间总体上无显著差异($F=$

表 2 各年级具体性均值多重比较

(I) 年级	(J) 年级	平均值差(I-J)	标准差	显著水平
大一	大二	5.54588	5.49402	.797
	大三	11.10583	5.66511	.279
	大四	13.36058	5.69396	.139
大二	大一	-5.54588	5.49402	.797
	大三	5.55995	5.53553	.799
	大四	7.8147	5.56505	.578
大三	大一	-11.10583	5.66511	.279
	大二	-5.55995	5.53553	.799
	大四	2.25475	5.73402	.985
大四	大一	-13.36058	5.69396	.139
	大二	-7.8147	5.56505	.578
	大三	-2.25475	5.73402	.985

表 3 各年级意象性均值多重比较

(I) 年级	(J) 年级	平均值差(I-J)	标准差	显著水平
大一	大二	6.56986	4.83058	.604
	大三	12.44318	4.97203	.100
	大四	14.64776*	5.00911	.036
大二	大一	-6.56986	4.83058	.604
	大三	5.87332	4.83486	.688
	大四	8.0779	4.87297	.432
大三	大一	-12.44318	4.97203	.100
	大二	-5.87332	4.83486	.688
	大四	2.20457	5.01323	.979
大四	大一	-14.64776*	5.00911	.036
	大二	-8.0779	4.87297	.432
	大三	-2.20457	5.01323	.979

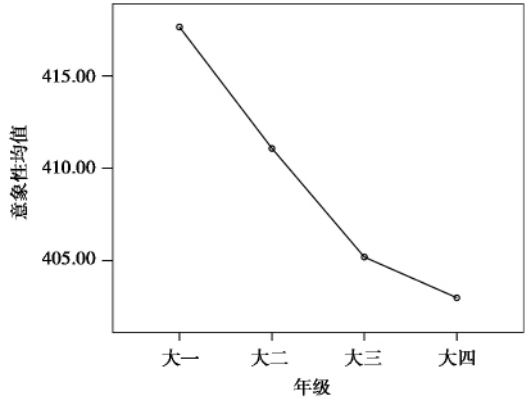


图 2 各年级的意象性均值走势

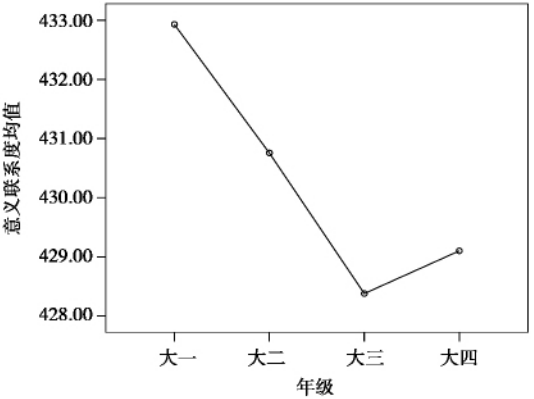


图 3 各年级的意义联系度均值走势

0.350, $p=.789$)。表 4 中跨年级均值多重比较表明,各年级间均无显著差异($p>.05$)。据此,考察结果不完全支持假设三:意义联系度会随学习阶段的提升而下降。

4.2.4 熟悉度的变化趋势

各年级用词的熟悉度均值走势如图 4 所示:从一年级到二年级,此均值显著下降,但二到四年级又有所回升。对各年级均值的单因素方差分析表明,总体上有显著差异($F=8.182, p=.000$)。表 5 中的均值多重比较表明,只有一年级和其他年级的均值有显著差异($p=.000$)。据此可认为此结果未验证假设四:

作文用词熟悉度会随学习阶段提升而下降。

四种心理语言属性的考察结果从总体上看,研究结果与假设不完全相符:具体性和意象性变化趋势基本符合预期假设,而意义联系度和熟悉度的变化趋势则否。

5. 讨论

5.1 关于具体性的发展趋势

对具体性发展趋势的调查表明,从一年级到四年级,产出词汇总体的具体性分值一直下降,即抽象程度越来越高。虽然方差分析

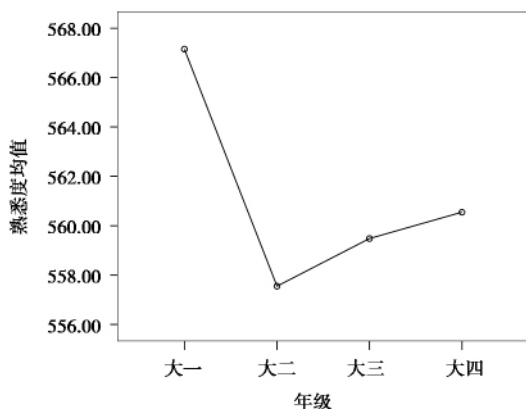


图 4 各年级的熟悉度均值走势

表 4 各年级意义联系度均值多重比较

(I) 年级	(J) 年级	平均值差(I-J)	标准差	显著水平
大一	大二	9.60182*	2.03335	.000
	大三	7.67412*	2.09328	.004
	大四	6.60756*	2.10825	.020
大二	大一	-9.60182*	2.03335	.000
	大三	-1.92770	2.03455	.826
	大四	-2.99426	2.04994	.545
大三	大一	-7.67412*	2.09328	.004
	大二	1.92770	2.03455	.826
	大四	-1.06656	2.10941	.968
大四	大一	-6.60756*	2.10825	.020
	大二	2.99426	2.04994	.545
	大三	1.06656	2.10941	.968

表明年级间均值差异并不显著, 但根据总体的下降趋势可以推测, 到了更高的学习阶段, 显著的变化趋势是可以预见的, 这种变化趋势在 Salsbury *et al.* (2011) 中也得到证实。本研究和 Salsbury *et al.* (2011) 都表明, 随着二语词汇水平的提升, 二语语篇越来越少地依赖于意义具体的词汇附带的“内在语境”(inherent context) (Salsbury 2011)。使用越来越多的抽象词汇而不拘泥于意义具体的词汇, 表明学习者随着学习阶段提升对抽象语

表 5 各年级熟悉度均值的多重比较

(I) 年级	(J) 年级	平均值差(I-J)	标准差	显著水平
大一	大二	2.17350	4.72027	.976
	大三	4.55110	4.84751	.830
	大四	3.82927	4.85885	.892
大二	大一	-2.17350	4.72027	.976
	大三	2.37760	4.81360	.970
	大四	1.65576	4.82501	.990
大三	大一	-4.55110	4.84751	.830
	大二	-2.37760	4.81360	.970
	大四	-.72183	4.94957	.999
大四	大一	-3.82927	4.85885	.892
	大二	-1.65576	4.82501	.990
	大三	.72183	4.94957	.999

言的掌握越来越好。在议论文中, 使用抽象的词汇将抽象思维表达清楚是语篇能力的基本要求, 因此产出更多的抽象词是学习者议论文语篇能力提升的重要标志。

Salsbury *et al.* (2011) 是对同一批受试产出词汇的具体性进行历时考察, 而本研究使用的是学习者语料库, 四个年级的受试不同, 该差异可能是造成年级间差异不显著的主要原因。此外, 本研究使用的材料是以说理为主的议论文, 与说明文、记叙文等文体相比, 意义具体的描述性词汇普遍偏少, 各年级都是如此, 这也是造成年级间差异不显著的重要原因。

5.2 关于意象性的发展趋势

对意象性的调查表明, 从低年级到高年级, 学习者产出词汇的意象性分值逐渐降低, 且这种趋势到四年级开始有显著意义。这说明, 随着学习阶段的提升, 二语语篇产出越来越少地依赖于意象性较强的词汇。产出词汇意象性总体分值的下降, 同时也说明了学习者对抽象语言的掌控能力不断提升, 进一步说明其抽象思维能力的增强。本文对词汇意象性发展趋势的考察结果与 Salsbury *et al.* (2011) 也是一致的。

5.3 关于意义联系度的发展趋势

本研究发现,从一年级至三年级,二语产出词汇的意义联系度均值在下降,但到四年级有所回升,且该均值在各年级间无显著差异。最初的假设是意义联系度弱的词更难掌握,于是高年级学生比低年级的掌握更多,反映在作文中就是词汇总体的意义联系度随学习阶段的提升而不断降低。该发展趋势在一年级至三年级与预期相同,但四年级学生产出词汇的意义联系度均值反而高于三年级。这在一定程度上是对假设的否定,也与 Salsbury *et al.* (2011) 的调查结果不同。

我们认为,二语习得中意义联系少的词未必更难,也未必更晚习得。以多义词为例,意义联系度强的词汇很多是多义词,往往可与多个词有近义或反义的联系,弄懂这些义项对低层次学习者来说挑战并不小。笔者又重新考察了各年级的词表,发现高年级作文中多义词更常见,低年级的则极为罕见。如: Internet is not only an instrument to shorten the distance(三年级学生作文中)。该结果与 Salsbury *et al.* (2011) 的不同可能源于语料:使用语料库事先无法了解各年级间初始二语水平的差异,且写作时的影响因素更无法控制,这些因素都可能影响结果。

5.4 关于熟悉度的发展趋势

对词汇熟悉度的调查显示,从一年级到二年级的熟悉度均值下降,而二年级到四年级则有所回升。一般来说熟悉度与词频正相关,越熟悉的词词频越高,反之亦然。从一年级到二年级,产出词汇的熟悉度显著下降($p=.000$),即高频词使用显著减少;从二年级到四年级,高频词的使用有所增加,但无显著意义($p>.05$)。从一年级到二年级熟悉度均值显著下降,即高频词显著减少或低频词显著增多,这在一定程度上表明了词汇量的扩大(一般来说,高频词习得早于低频词)。二年级之后,熟悉度略有下降,即低频词产出减少,

这可以用陈亚平(1997)和文秋芳(2006)的“高原期假设”来解释。陈亚平(1997)提出词汇学习的“高原期”出现在二年级。另一不可忽略的因素是备考。二年级恰是备考专四统测的阶段,密集的备考训练要求突击记忆大量的低频词,学生也会将其用到作文中。三年级以后,低频词使用减少,于是熟悉度回升¹。

本研究证实了前两个假设:写作产出词汇的具体性和意象性会随学习阶段的提升而降低。对于意义联系度和熟悉度的发展趋势,笔者认为还需进一步验证。此外还需指出,利用学习语料库研究年级间的词汇发展趋势固然方便,但一个难免的问题是对影响词汇历时发展的因素无法控制,如初始二语水平,语言学能等,这些不可控因素也会影响研究结果的效度。

6. 结论与教学启示

MRC 数据库为二语词汇研究提供了另一个视角。心理语言属性虽最早用于研究母语习得,但本研究表明其对二语研究也有很强的应用价值。本文对 MRC 数据库中四种心理语言属性发展趋势的调查发现,随着学习阶段的提升,学习者产出词汇的意义越来越抽象,意象性也越来越弱,这在一定程度上反映了学习者抽象思维能力的增强,以及用二语表达抽象思维能力的提高;而意义联系度和熟悉度则先下降后上升。这与以往的研究结果不完全一致。笔者认为这两种属性的变化趋势还需进一步验证,并且最好是通过历时跟踪同一批受试的方式验证。

我们发现二语写作词汇的具体性和意象性随学习阶段的提升而降低,具体词和意象性更强的词确实更早且更易习得。据此,可在词汇教学中进一步合理安排单词表:在低年级词表中加入更多意义具体、意象性强的词汇,而意义抽象、意象性弱的词汇可更多放到

¹ 我们并不否认一些四年级学生因专八备考会产出更多低频词的可能,但专四统测本身比专八更重视词汇知识的考察(比较试题要求即可),因此我们认为低频词记忆和产出的高峰出现在二年级是可以理解的。

高年级。意义联系度和熟悉度两属性的变化趋势仍有争议,以往认为意义联系度强的单词可为其他单词习得“搭建桥梁”,应尽早使学生掌握;但语义联系复杂的词会加重学生的记忆负荷,因此学生实际的接受能力也不可忽略。词汇教学中要掌握好两者的平衡才能优化教学效果。

参考文献

- Crossley, S., T. Salsbury & D. McNamara. 2009. Measuring L2 lexical growth using hypernymic relationships [J]. *Language Learning* 59, 2: 307-334.
- Ellis, N. & A. Beaton. 1993. Psycholinguistic determinants of foreign vocabulary acquisition [J]. *Language Learning* 43, 4: 559-617.
- Ellis, E. & G. Barkhuizen. 2005. *Analysing Learner Language* [M]. Oxford: OUP.
- Gee N., D. Nelson & D. Krawczyk. 1999. Is the concreteness effect a result of underlying network interconnectivity? [J]. *Journal of Memory and Language* 40, 4: 479-497.
- Gilhooly, K. & R. Logie. 1980. Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words [J]. *Behavior Research Methods & Instrumentation* 12, 2: 395-427.
- Laufer, B. & P. Nation. 1995. Vocabulary size and use: Lexical richness in L2 written production [J]. *Applied Linguistics* 16, 3: 307-322.
- McCarthy, P. & S. Jarvis. 2010. MTLT, vocd-D and HD-D: A validation study of sophisticated approaches to lexical diversity assessment [J]. *Behavior Research Methods* 42, 2: 381-392.
- Paivio, A., J. Yuille & S. Madigan. 1968. Concreteness, imagery, and meaningfulness values for 925 nouns [J]. *Journal of Experimental Psychology Monograph Supplement* 76, 1: 1-25.
- Polio, C. 2001. Research methodology in second language writing: The case of text-based studies [A]. In T. Silva & P. Matsuda (eds.). *On Second Language Writing* [C]. Mahwah, NJ: Erlbaum, 91-115.
- Salsbury, T., S. Crossley & D. McNamara. 2011. Psycholinguistic information in second language oral discourse [J]. *Second Language Research* 27, 3: 343-360.
- Toglia, M. & W. Battig. 1978. *Handbook of Semantic Word Norms* [M]. Hillsdale, N. J.: Erlbaum.
- Wilson, M. 1988. The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2 [J]. *Behavioral Research Methods, Instruments and Computers* 20: 6-11.
- 陈亚平, 1997, “二年级现象”的心理解释及对策[J]. *外语教学与研究*(4): 56-60.
- 董燕萍, 2005, 心理语言学与外语教学[M]. 北京: 外语教学与研究出版社.
- 李志雪、李景泉, 中国高水平英语学习者产出词汇使用情况研究——基于对中美大学生英语作文范文的对比分析[J]. *山东外语教学*(5): 56-59.
- 林峻, 2009, 基于笔语语料库的我国二语学习者产出性词汇习得状况研究[J]. *扬州大学学报(高教研究版)*(2): 91-93.
- 卢敏, 2008, 产出性词汇知识广度的发展特征——基于英语专业学生书面语的研究[J]. *外语教学理论与实践*(2): 10-15.
- 谭晓晨, 2006, 中国英语学习者产出性词汇发展研究[J]. *外语教学与研究*(3): 202-207.
- 童淑华、高金花, 2010, 第二语言学习者产出性词汇发展的理论模式[J]. *东北师大学报(哲学社会科学版)*(4): 93-96.
- 文秋芳, 2006, 英语专业学生口语词汇变化的趋势与特点[J]. *外语教学与研究*(3): 189-195.
- 吴瑾, 2011, 中国研究生产出性学术词汇知识深度的语料库研究[J]. *外语教学*(2): 52-55.
- 熊永红、黄元龙, 2011, 激活扩散模型与英语词汇教学中的常规关系建构[J]. *外语教学*(4): 61-64.
- 杨世登, 2007, 英语学习者产出词汇的发展模式[J]. *外国语言文学*(4): 202-207.

收稿日期: 2012-07-10;

作者修改稿, 2013-01-03;

本刊修订, 2013-03-01

通讯地址: 471003 河南省洛阳市 036 信箱 110-2 号

<lihai1017@163.com> (李)

471003 河南省洛阳市 036 信箱 30 号

<zhangshujing001@sina.com> (张)

译者的意象图式与合成概念化^{*}

——基于语料库方法的《红楼梦》“社会脸”翻译研究

谭业升

(上海外国语大学《外国语》编辑部,上海 200083)

摘要: 本文主要基于语料库的方法,以两个《红楼梦》英译本中“社会脸”的表达和概念化为例,考察意象图式、涉身体验与创造性翻译认知过程的关联。分析和考察显示,一方面,“社会脸”的源域意象图式在翻译中可以基于涉身体验在不同的维度上得到例示或拓展,并呈现两种语言文化在社会关系的隐喻性概念化方式上的偏好差异;另一方面,基于共享的意象图式表征和相互通约的涉身体验,两种语言偏好的概念化方式可以在翻译中合成,并产出新的译语概念化方式和表达形式。本文所采用的以“图式-例示关系”为基础的概念分析框架、非句级平行的语料分析以及结合多种语料库的方法,对于今后的认知翻译学 and 对比认知语言学研究有一定的参考价值。

关键词: 社会脸; 语料库方法; 意象图式; 概念化; 认知翻译学; 合成

中图分类号: H059 文献标识码: A 文章编号: 1004-6038(2013)03-0055-05

1. 引言

“脸面”问题传统上属于有关礼貌原则和交际策略的语用学研究范畴,与文化研究也有着密切的关系。在以往的研究中,如 Gu (1990: 237-57)、顾曰国 (1992)、Watts 等 (1992)、Yu (1999: 281-312) 等,对中英“脸面”的不同语用含义和社会含义的差异进行了考察。张建理 (2003: 54-58) 从词的多义性派生的角度对英汉的“face”和“脸”、“面”的共性和差异进行了研究,并指出:两者为对等的基本人体词,均存在一词多义现象,英汉词义派生在基本认知方面相类似,在涉及社会文化方面和名动转用方面存在差异。本文将将以“图式-例示关系”这一认知语言学的核心概念框架为基础,以更加简洁的方式概括英汉语言文化在“脸”的概念化方式上的共性与差异;并重点考察在跨语境场景下“共性”与“差异”之间的有机联系,及其在翻译认知策略和认知成果上的反映。

2. “脸”的意象图式及其例示 “社会脸”的认知阐释

“脸”的表达寄生和依赖于感知和交互行为等涉身体验的表征结构,其表征知识是基于与环境(包括人)的涉身交互体验而形成的。有关“脸”的涉身体验经选择提炼后基于反复出现的样式而发展出经验的完形,形成似意象的整体和拓扑结构,即“脸”的意象图式和范畴(Lakoff, 1987)。“脸”的意义是多样化的,但是“脸”的意象图式则是其多样意义中不变性的直接规定。它帮助我们构筑了进行社会关系概念化、范畴化以及抽象推理的基础(Ibarretxe-Antuñano, 1999)。这些结构可以通过隐喻性和转喻性的认知加工,而变得更细致、复杂,以用来理解更抽象、更复杂的社会经验,将零散的社会文化经验组织成有构造的整体。

Langacker (1987: 66-70) 指出,图式本身不负责实际的

语言表达,只是起到范畴化的功能,提供了一个表达被范畴化为一个型式的合格的例示所需要的最小规定特征,即图式允准了(sanction)例示。如果一个表达可以满足图式提供的最小规定特征,并且被判定为适合此图式,即使它的特征要比从图式及其组成要素所能预测到的结构更加精准和明确,我们仍可以说它们之间存在完全允准关系;而如果一个表达所实现的例示型式只是部分满足图式的规定性特征,那么它们之间存在部分允准关系。在原型理论中被称作原型-拓展关系(prototype-extension)。从总体上看,“脸”的范畴是靠“脸”的意象图式来协调统一的,并且在元认知层面呈现“中心-边缘”结构的原型性特征。如下文的分析所示,“脸”的图式可例示为“社会的脸”(简称“社会脸”),如“赏脸、给脸、丢脸”等)和“表情义的脸”(简称“表情脸”)。而“社会脸”可以看作是“脸”范畴中拓展的子范畴。它是基于身体部位“脸”的感知互动和运动等物理、生理体验,通过想象,隐喻性地投射到社会交互体验上而形成的。可以简单表示为:脸的物理生理特征(源域): 社会交互特征(目标域)。这种想象性的认知加工所获得的“社会脸”的表征在一些语言文化^①,尤其是汉语文化中,已经固化,从而具有图式的地位。它提供范畴化的规定特征,从而又会帮助我们组织具体的概念化过程。

“社会脸”范畴所拥有的隐喻关系图式在目标域上可以例示为更具体的“身份脸”、“尊严脸”、“交情脸”等;而在其源域上也可基于有关脸的涉身体验在不同的意象维度上例示。不同的例示都有可能在具体语境下实现为具体的语言表达,也可以说,具体的语言表达以适合语境的具体例示为基础。如下文所示,“有脸、没脸”既可以是“尊严脸”的表达实现,也可以是“身份脸”的表达实现。这两种用法都是对

^{*} 基金项目: 本文为上海外国语大学青年教师创新团队项目“汉英语言认知对比的理论、方法及其应用”(项目编号: QJTD11TYS01) 的阶段性研究成果

作者简介: 谭业升,副编审,博士,硕士生导师,研究方向: 认知翻译学,对比认知语言学

“社会脸”图式在“存在意象”维度上进行例示性识解的结果。

对于语言对比和翻译研究来说,在明确了与“脸”范畴有关的图式-例示关系和层级之后,我们就可考察两种语言文化在何种层级和例示维度上具有共性或互通,而在何种层级和例示维度上相互区别或发生分化(参见谭业升 2011)。不同语言文化的人都有基于对脸的涉身体验(包括物理、生理体验和社会文化体验)而形成的意象图式,是不同语言文化间互通的基础,也是翻译的基础。如以往的研究所示,在不同的语言文化中,“脸面”具有的社会含义不同。这是因不同的社会文化体验或因突显了不同的涉身体验而产生的差异,但这种社会含义的不同仍是以共享的涉身体验和“脸”的意象图式为基础,也就是说,在相似的意象图式基础上,两种语言文化可以作不同的拓展例示和实现。语言文化之间的分化也会表现为在表征上与理想或典型成员的距离的差异或例示方式和使用频率上的差异。以此为基础,我们就可以深度认识“基本认知方面”的共性和“社会文化方面”的个性(张建理, 2003)间的有机联系,并认识翻译“语言对”的“同中有异”和“异中有同”(潘文国 谭慧敏 2006),从而把握译者认知策略的系统性和动态性。

3.《红楼梦》翻译中“社会脸”图式的多维度例示与“社会关系”的概念化

接下来,我们将对《红楼梦》原文和两个英语译文,即 Yang 和 Yang(1978)译文以及 Hawkes 和 Minford(1977)译文(以下简称杨译文、霍译文)中出现的脸面和 face 表达展开具体分析,侧重考察“社会脸”隐喻关系图式(包括源域和目标域)的多样化例示及其实现形式,考察与“社会脸”的翻译有关的认知策略,及其所体现的英汉语言文化在社会关系概念化方式上的共性和差异,并探究有关表达的认知来源。

3.1 对于研究方法的说明

我们利用语料库分析工具 Concordance 3.2 对原文和两个译文中的脸面/face 表达进行索引分析和认知对比分析,然后基于两个本族语语料库 BNC 和 COCA 进行译语有关表达的考证分析。探讨认知层面上的对应关系时,“原文没有脸而译文有 face”的情况往往具有重要的价值,如果仅仅做句级对齐平行语料的分析,就会遗漏这种情况;因此,必须突破

表2 《红楼梦》汉语原文“社会脸”表达的分布统计表

“社会脸”意象 维度(源域)	词型(type)搭配和词例(token)(131例)(重复实例在总计数中剔除;篇幅所限,仅举部分例证,下同)	突显的社会义 (目标域)
整体感知 (16例)	嘴脸、脸(2例):但只你我这样个嘴脸	身份+地位
	脸、脸面(4例):……仗着老子娘的脸面	交情+尊严
	老+脸(7例):……倒还是舍着我这付老脸去碰一碰	尊严+身份+交情
	脸皮+厚、老着+脸(3例):……急忙出来老着脸和他母亲道:……我把脸皮厚些	羞耻心+尊严
视觉和计量 (6例)	脸+好看、不好看(5例):到底府上脸面不好看	尊严
	脸+大(1例):前儿亏你还有那么大脸	身份+尊严
味觉(1例)	脸+酸(1例):脸酸心硬,一时恼了,不认人的	交情

“局部对等”框架和有关方法论限制,在整个文本层面上寻找脸的意象图式的对应实现情况,以获得有关概念化方式的更全面、更深层的认识。而如下文所示,翻译对应情况也可在一定程度上通过表达出现的频次、词型搭配的数量和意象维度的分布反映出来。也就是说,我们的分析是平行的,但并不一定在句级上平行。总之,因为研究的出发点、问题、对象和目标均不同于传统对等理论框架下的翻译研究,我们的语料库分析方法也存在差异。

3.2 “社会脸”图式在三个文本中的多维度例示和认知对比分析

我们首先分别以“脸”、“面子”和“face”为中心词(head-word)对《红楼梦》原文和两个经典译文进行了索引分析,并得到了以下分布情况表:

表1 “脸、面子”表达和“face”表达在《红楼梦》原文和译文中的分布情况

文本 表达类型	《红楼梦》原文 (脸、面子表达)	杨译文 (face)	霍译文 (face)
各文本中脸\面(face)表达	458(脸)+7(面子)=464	350	566
各文本中社会义“脸、面子”(face)表达	131	58	33
各文本中其它脸、面(face)表达(包括表情义表达)	333	292	533

需要说明的是:在“社会义”表达的统计中,我们对原文和译文中与“人脸”无关的脸面/face 表达,如“鞋面子”,及其本义表达和表情义表达,如“鸭蛋脸面、他一翻脸”^②、金荣气黄了脸”、“face should light up first, he threw them in her face”等,我们都作了人工剔除。我们对三个文本的脸面/face 表达出现的总频次和“社会脸”表达频次分别两两进行了构成比的卡方检验,得到的结果分别为:原文和杨译文: $\chi^2 = 9.63a$, $p = .002$;原文和霍译文: $\chi^2 = 68.199a$, $p = .000$;杨译文和霍译文: $\chi^2 = 22.380a$, $p = .000$ 。卡方检验获得的 p 值均小于 0.05,因此,我们可以说,三者之间都存在显著性差异。

表1大体反映了各类型“脸”表达的分布情况。接下来,我们集中对“社会脸”表达展开更加具体的定性和定量认知分析,考察“社会脸”表达的分布、所例示的意象维度及其所突显的不同的社会义。

触觉(5例)	脸+软(5例):只是脸软怕人恼	品性+交情
处置 (57例)	赏、给(留)、逞+脸(11例):老祖宗又不肯赏脸。……逞了凤丫头的脸……如今还求奶奶给留脸	尊严+恩惠
	给+没脸(3例):……也不知谁给谁没脸?	尊严+羞耻心
	要、搁、照顾、得、丢、失、剩+脸(面)(16例):这性命脸面要也不要?……如今我的脸搁在那里呢!……连外头的脸面都不顾了……我若不得脸败了时,……脸面也不能剩……就连你们素日的老脸也都丢了(与老脸重复)	尊严+身份
	伤、抓破、碍着+脸(面)(7例):又伤不了脸面……一个个的不象抓破了脸的,……又碍着赖尚荣的脸面	尊严+身份+交情
	保、盖+脸(3例):他臊了,没的盖脸……只怕保不住脸了	尊严+羞耻心
	上、开+脸(4例):倒叫你进来瞧瞧,就上脸了……进门就开了脸,就封你姨娘	身份+尊严
	(老)脸上+下不来、过得去、过不去、碍不过去(13例):他嫂子脸上下不来,……他也不管人脸上过得去过不去	尊严+羞耻心
存在 (51例)	有、无、没(头/大)+脸(面)(11例):……有脸者不服铃束,……宝钗等姊妹并外头的几个执事有头脸的媳妇,……无脸者不能上进。……前儿亏你还有那么大脸(与“大脸”重复)	身份+地位
	有、没(什么)+脸(面)(40例):又可分为两小类:1)表示社会行为的依仗:……什么脸出去见人!……亏你还有脸来告诉! 2)表示价值判断,……况且都作的是有什么有脸的好事	尊严+羞耻心

此表显示,在原文中,“社会脸”的源域意象图式基于整体感知、视觉、触觉、味觉、处置、存在等涉身体验在不同维度上得到了广泛的多样化例示,并且在具体意象上呈现出对立的样态,如“软一硬”、“给/赏/逞一得/要”、“有一无/没”、“留一丢/失/剩”、“保、照顾、盖一伤、抓破、碍”、“上一下”、“好看一不好看”、“过得去一过不去/(碍)过不去”等,当然还有“老、大、搁、开、厚”等独立意象。“社会脸”的目标域语义图式在原文中有以下几种例示形式:突显“社会身份和地位”的脸,如“有脸者不服铃束,无脸者不能上进”;突显交情

的脸,如“老脸”、“脸酸心硬”;关乎“羞耻心”的脸,比如:“况且都作的是有什么有脸的好事”。当然,某一种“社会脸”表达所具有的意义往往是多重的,因为关乎社会交互关系的“社会身份、地位、交情、尊严、羞耻心”等都是相互作用和交叠的。只不过在某些具体情景的表达中突显了某一方面的意义。比如“有脸、没脸”在不同的语境中突显的意义就不同。

接下来我们来考察“社会脸”图式在译文中的例示情况。社会义的 face 表达在两个译文中的分布情况请见以下表3和表4:

表3 杨译文“社会脸”表达分布情况表

“社会脸”意象 维度(源域)	face 词型搭配和词例(33例,重复计数已经剔除)	突显的社会义 (目标域)
整体感知 (2例)	face(1例) This isn't a question of competing for status, it has nothing to do with face. old face(1例): slap my old face and pull down my temple	身份+尊严 年龄+尊严
面对和呈现 行为(15例)	be able to face, how could/can/shall + face, How am I to face _{vt.} (11例): How can you face people after this? show + face; look sb in the face.(3例): not to show his face here any more... after turning everything upside-down how can I look you in the face again?	尊严+羞耻心
	afraid to face sb(1例): being actually afraid to face her	勇气+尊严
存在(6例)	have(how much no) + face, have the face to(6例): no matter how much face the offender may have... so it seems I really have face	尊严+ (身份地位)
处置(35例)	give + face(4例): give us a little face... Then if I get given face you can bully people outside	交情+尊严+ 身份地位
	get/gain + face(3例): got quite enough face... I could gain face by distributing as presents.	
	leave/save + face(5例): wanted to save your face... So please leave her some face.	尊严+羞耻心
	lose + face; loss of face(23例): lose all the face... I must find some way of stopping their mouths without any loss of face.	
计量和视觉	a little/enough face/ how much + face(重复例):... give us a little face... got quite enough face... no matter how much face the offender may have. big + face(与“存在、处置”意象重复例):... yet you feel you have big face... only send presents to those who have power and big face.	交情+尊严+ 身份地位的 社会度量

表4 霍译文“社会脸”表达分布情况表

“社会脸”意象 维度(源域)	face 词型搭配和词例(33例,重复计数已经剔除)	突显的社会义 (目标域)
整体感知 (7例)	face(3例):... As regards face, it makes no difference to me.	尊严
	old + face(2例): But as my old face is tough enough to stand a slap or two	尊严+羞耻心+(年龄)
	bold + face(1例): I'll just have to put a bold face on it(对应:老脸皮子厚的)	有胆量+羞耻心
	holy + face(1例): I would never be granted another chance to behold your holy face.	神圣+身份+ 尊严

面对和呈现行为(7例)	how will I/ am I going to/shall I be able to + face _{vt} Sb. (5例): How am I going to face Mrs Lian when I next see her? look sb in the + face (3例): How can I look you and Auntie in the face again? show + face (2例): How are you going to show your face anywhere?	尊严 + 羞耻心
	not daring/hard to face sb. (2例): He took to his heels not daring to face his master and mistress.	勇气 + 尊严 (羞耻心)
存在(4例)	have the face to (4例): ... How can you have the face to ask such a question?	尊严 + 羞耻心
处置(10例)	save/leave + face (6例): ... saved yourselves a bit of face ... leave us much face ... What sort of face does it leave you with?	尊严 + 羞耻心
	loss of face (4例): ... That would be a real loss of face .	尊严 + 羞耻心
计量(重复例)	a bit of/ no/ much face (均为与“处置意象”有关的重复例): ... saved yourselves a bit of face ... you have no face left to save ... leave us much face .	尊严 + 羞耻心 + 身份的度量

表3和表4两个统计显示了两个译文在词型搭配上的共性和差异:

(1) 两个译文共有的词型搭配有: show + face, look sb. in the face, old face, save/leave + face, have no face, loss of face, be able to/be to + face, have the face to; 而类似的词型搭配有: how shall I / will I/ am I going to + face_{vt}, afraid/daring/hard to face sb, how could/can/ + face_{vt}。

(2) 杨译文有而霍译文无的词型搭配: have (how much) + face, a little/enough face, big face, get + face, give + face, lose + face

(3) 霍译文有而杨译文无的词型搭配: bold/ holy face, a bit of/much face

在两个译文中,“社会脸”的图式在例示性识解的频次上较原文大幅缩减,其基于源域涉身体验而得以例示的意象,除整体感知意象外,还包括:处置意象,如“给一得、留一失”等,存在意象(“有无”)和视觉意象(“大小”),但少了“保、照顾、盖一伤、抓破、碍”、“上一下”,“好看一不好看”、“过得去一过不去/(碍)不过去”、“搁、开”、“软一硬”等意象。“old face”从两个译文中的具体语境看,更倾向于突显年龄,而不具有原文中9个例子(尤其是“老着脸”)所突显的典型社会义。

两个译文都比原文多了“多少”的意象。此外,尽管原文中也存在“有脸+见”的表达,但是两个译文突显了具有强烈社会义的“呈现”和“面对”行为意象^③。

“社会脸”图式在霍译文中作例示性识解的频次较原文和杨译文都少,霍译文除比原文少了上述意象特征外,还比杨译文少了“给、得”的处置意象和“大小”的视觉意象,但是却增加了“有胆子(bold)”和“神圣(holy)”意象。当然,这些意象已经偏离了原文“社会脸”的典型意象。另外,其目标域的社会关系意义少了交情义,“存在(有无)”意象中少了身份地位义。需要注意的是,在霍译文中出现了1例“Her face hardened (as she pronounced)”,其中的“hardened”取表情义不取社会义,换句话说,“软硬”意象没有直接进入社会关系概念化的过程。

3.3 英汉两种语言文化在社会关系概念化上的共性与分化

通过上述分析,我们可以说,在两种语言文化中,“脸”的基本意象图式和有关涉身体验在社会关系概念化中都发挥着作用;基于此,“脸”的社会义才能够在翻译中实现会通。

在两种语言文化中,“脸”都可能是社会概念或社会范畴,也可能是表情范畴,这是共性的。汉语中的“脸”不仅是社会性的(尊严和交情“面子”),也与个人身份相关、与个人愿望相关,比如:汉语中“拼了这张老脸”,既与个人身份相关,也与个人愿望的实现有关。而英语既有体现个人愿望和勇气的“be able to face sb”,也有体现社会的接受和认可意愿的“save face”、“lose face”。在认识到这些共性的同时,我们也须看到,两个译文在“社会脸”图式的例示方式和语言实现频次上与原文表现出较大的差异,即出现了分化:一方面,原文在分布频次和突显的意象维度上都要多于杨译文和霍译文,尤其是霍译文,其“社会脸”表达频次仅有原文的不三分之一;另一方面,在原文中较为突显的意象维度,在译文中没有对应的体现,而译文的有些拓展例示已经偏离了典型的社会义。这说明英汉两种语言文化在将“脸”识解为社会性质的概念时,存在着较大的差异;而汉语文化尤其突显了“脸”在概念化社会关系中的作用,并表现出比较特殊的社会文化概念化结构。

4.《红楼梦》译文中“社会脸”表达的认知来源

接下来的问题是,两个译文中的“社会脸”表达来自于何处?是英语本身的规约性表达还是创新性表达呢?为了较为客观地反映和认识两个译文对“社会脸”图式进行例示的作用力来源和具体过程,我们利用两个本族语料库 BNC (1991年后兴建,2003、2007年做过小的调整)和 COCA (2008年建成,每年更新两次)对两个译文中出现的“社会脸”表达的分布情况进行了考察(详见谭业升,2012: 317-325)。

统计显示,在比两个语料库的创建都早的杨译文和霍译文中,一些“社会脸”表达不是译语本身已有的表达形式,即不是来单纯来自于译语现成的表达库存;而同时,相对于原文来说,两译者也都基于共享的“脸”图式在不同的关涉社会关系概念化的意象维度上作了进一步的创造性认知拓展,比如两译文增加了“多少”维度,并且在“呈现”和“面对行为”意象上附加了社会义,而霍译文则增加了“勇气”、“神圣”等具体意象。也就是说,在表达生成过程的背后,译者在认知上对两种语言文化偏好的概念化方式进行了合成。

在意象的拓展上,杨译文相对更加大胆,对原语中典型的关涉社会关系概念化的“给一得”、“大小”意象进行了移

植,其中具有社会义的创造性表达形式“give face”、“gain a little face”、“enough face”等虽没有在BNC中出现,却在较晚的COCA中出现了,即已经被英语读者所接受。不过,显然是源自于汉语“面子大”的具有社会义的“big face”却始终没有进入两个英语本族语语料库,这也反证了译者在进行概念化合成时所存在的风险。总之,“社会脸”图式在两个译文中的例示过程,是基于共享的涉身体验,对原语和译语有关“脸”的偏好概念化方式进行平衡考量,并进行合成概念化的过程,而译者在再符号化过程中可以有限地突破规约,为目标语带来新的表达形式。此合成的过程充分展示了译者在认知上的创造性和个性化特征。

5. 结语

本文在认知翻译研究上做了三方面的尝试。首先,展示了以“图式-例示关系”为基础的认知分析框架(另可参考谭业升,2011);其次,采用了非句级平行的语料分析方法;最后是结合多种语料库的方法。研究表明,“脸”的图式-例示的级阶可以揭示和解释语言对比和翻译所涉及的语言文化之间的“同中有异”和“异中有同”关系,以及以此为基础的翻译认知策略。本文通过翻译文本的认知对比分析和本族语语料库的对比验证,获得了有关语言特有概念化方式及其交互作用的更具体认识。以翻译语料所揭示的问题(主要是对系统差异性的期望或假设)为引子,充分利用其它种类的语料库加以综合对比和验证的方法,有望在今后的认知翻译学和对认知语言学研究,带来更多有关语言认知共性和个性的更丰富、更深入的认识。

注释:

- ①固化的“社会脸”图式和范畴是否存在于所有或大多数语言文化中,有待于进一步考察。
- ②表情脸和社会脸的界限并非泾渭分明,存在一些边缘性的、模棱两可的范畴成员。一些表达如“多少年来从来没有红过脸”、“他们俩翻脸了”等更倾向于直接表现社会关系,但因为建立在“表情义”基础上,所以仍划归“表情脸”范畴。
- ③在原文中,也存在“当面”这种表示“面对”行为的表达,有13例,但是没有明显的社会义,如:等张材家的来要,当面称给他瞧了等。

参考文献:

- [1]Cao,Xueqin(清).1977.Hawkes,D.&J.Minford(trans.).*The Story of the Stone*[M].Bloomington:Indiana University Press.
- [2]Cao,Xueqin(清).1978.H.Yang&Yang,G.(trans.).*A Dream of Red Mansions*[M].Beijing:Foreign Languages Press.
- [3]Gu,Y.1990.Politeness phenomena in modern Chinese[J].*Journal of Pragmatics* 2:237-57.
- [4]Ibarretxe-Antuñano,J.1999.*Polysemy and Metaphor in Perception*

Verbs: A Cross-linguistic Study[D].Edinburgh:University of Edinburgh.

- [5]Lakoff,G.1987.*Women,Fire and Dangerous Things:What Categories Reveal about the Mind*[M].Chicago:The University of Chicago Press.
- [6]Langacker,R.1987.*Foundations of Cognitive Grammar(Vol 1):Theoretical Prerequisites*[M].Stanford:Stanford University Press.
- [7]Watts,R.,S.Ide&K.Ehlich.1992.*Politeness in Language:Studies in Its History,Theory and Practice*[M].Berlin:Mouton de Gruyter.
- [8]Yu,M.1999.Universalistic and culture-specific perspectives on variation in the acquisition of pragmatic competence in a second language[J].*Pragmatics* 2:281-312.
- [9]曹雪芹(清).1989.红楼梦[M].长沙:岳麓出版社.
- [10]顾曰国.1992.礼貌、语用和文化[J].*外语教学与研究* 4:10-17.
- [11]潘文国,谭慧敏.2006.对比语言学:历史与哲学思考[M].上海:上海教育出版社.
- [12]谭业升.2011.转喻的图式及其例示的语言差异:以英汉名词动用为例[J].*外国语文* 3:57-62.
- [13]谭业升.2012.认知翻译学探索:创造性翻译的认知路径与认知制约[M].上海:上海外语教育出版社.
- [14]张建理.2003.英汉多义词异同研讨:以“脸、面”为例[J].*外国语文* 4:54-58.

Abstract: Via a corpus-based analysis of how “social face” is presented and conceptualized in two English versions of *Hong Lou Meng*, the paper intends to explore the connections between the creative cognitive processes of translation and the underlying image schema and embodied experiences. The analysis reveals that in the process of translation the source-domain image schema of “social face” can be instantiated or extended along various dimensions based on embodied experiences, involving different preferred modes of metaphorical conceptualization of social relations in the two language-cultures. And on the other hand, given the common underlying image-schema and the commensurable embodied experiences, language-specific modes of conceptualization can be integrated, and thus produce new ways of conceptualization and new linguistic expressions in the target language. The tentative conceptual framework of “schema-in-stance” hierarchy, the non-sentence-level parallel analysis and multi-corpus analysis adopted in the paper may shed light on the research of cognitive translology and contrastive cognitive linguistics in future.

Key Words: social face; corpus-based analysis; image schema; cognitive translology; integration

基于 R-gram 的语料库分析软件 PowerConc 的设计与开发

许家金, 贾云龙

(北京外国语大学 中国外语教育研究中心, 北京 100089)

摘 要: 在继承以往语料库分析软件优点的基础上, 本研究开发了具有独立知识产权的 PowerConc 语料库分析工具。PowerConc 对传统的词汇索引、词表生成、主题词计算等功能进行了重构、扩展和优化。整个软件以基于正则表达式(regular expressions)的 N 元组(N-gram)为基础。二者的有机结合即本文所提出的 R-gram。R-gram 这一概念大大增强了检索和匹配的灵活性。同时我们设计了兼容正则表达式的简易输入语法——Smart Input, 降低了用户使用的难度, 提高了软件的易用性。PowerConc 软件基于面向对象的思想开发, 核心功能被封装在不同的类中, 与界面分离, 具有很好的扩展性和可维护性。PowerConc 的开发将有效促进语料库语言学研究的开展。

关键词: 语料库分析工具; PowerConc 软件; R-gram; 语料库

中图分类号: H319.3

文献标识码: A

文章编号: 1001-5795(2013)01-0057-0006

1 语料库分析软件的开发背景

1.1 引子

语料库研究需要对大量文本进行计算机分析, 其中语料库分析软件的作用十分关键, 且很大程度上决定着研究数据的准确性和可靠性。没有良好的语料库工具支持, 语料库研究便难以有效开展。目前常用的语料库分析工具有: Mike Scott 设计的 WordSmith Tools (以下简称 WordSmith)、Laurence Anthony 设计的 AntConc、Michael Barlow 设计的 MonoConc Pro 和 R. Watt 设计的 Concordance 等。其中 WordSmith 功能最全, 学界认可度最高。其他软件有的是 WordSmith 的(部分)重写, 有的只能实现 WordSmith 的少量功能。综合来看, 这些软件通常包含词汇索引(concordancing)、词表生成(word list)、主题词计算(keywords)等功能, 但在统计和搭配计算等方面, 对正则表达式

(regular expressions) 和 N 元组(N-gram)的支持, 易用性和计算效率方面还有待提高。

近些年来, 国内外语料库建设蓬勃发展, 但语料库分析软件的开发却相对滞后, 一方面新工具开发较少, 同时, 原有语料库分析工具升级缓慢, 在核心功能上改进不大。本研究希望结合语料库语言学近年的发展, 开发出与之相适应的分析工具。

1.2 语料库分析软件发展概述

语料库是指按一定原则取样获得的大规模电子文本汇集(Sinclair, 1991; Hunston, 2002; Baker, 2006)。语料库规模通常很大, 因此需要借助计算机软件来辅助分析。近半个世纪以来, 语料库分析工具层出不穷, 数量、种类不断增加。

语料库软件包括: 词汇索引工具(concordancer)、自动和手工标注工具(词性标注、句法标注、语义标注、语用标注等)、文本整理工具(文本格式转换、文本编

作者简介: 许家金: 男, 博士, 副教授。研究方向: 话语分析、语料库语言学。

贾云龙: 男, 硕士。研究方向: 语料库语言学、教育技术。

收稿日期: 2012-08-18

基金项目: 本文的撰写得到国家社科基金项目“基于双语语料库的汉语复杂动词结构英译研究”(项目编号: 12CYY060)和教育部分“新世纪优秀人才支持计划”(项目编号: NCET-12-0790)的资助。

码转换)、口语转写工具、统计分析工具等。语料库分析工具中最常用的是索引工具。一般所谓的通用型语料库分析工具即指索引工具。最早的计算机索引工具由 Roberto Busa 于 1951 年开发 (McEnery & Hardie, 2012:37)。当时的索引工具只能提供索引行语言实例。后来索引工具的功能得到很大扩展,但名称仍然叫做索引工具。现在的通用型索引工具,通常至少包括生成索引行和词表两大功能。

根据 McEnery & Hardie (2012:37-48) 对语料库分析工具的时代划分,我们将相关工具开发情况汇总如表 1。

表 1 四代语料库分析工具

代次	年代	作者	索引工具名称
第一代	1951	Roberto Busa	不详
	1978	A. Reed	CLOC
	1992	Roger Garside	XANADU
第二代	1989	J. Bradley, et al.	TACT
	1991	Higgins	MiniConcordancer
	1993	Mike Scott & Tim Johns	MicroConcord
第三代	1997	Mike Scott	WordSmith Tools V2.0
	2012	Mike Scott	WordSmith Tools V6.0
	2002	Laurence Anthony	AntConc
	早于 1999	Michael Barlow	MonoConc Pro
	早于 2005	Lou Burnard & Tony Dodd	Xaira
第四代	2000	S. Hoffmann, S. Evert & A. Hardie	CQPweb
	2002	Mark Davies	BYU corpora
	2002	Adam Kilgariff	SketchEngine

其中,第一代和第二代索引工具主要是在 DOS 环境下运行。第一代工具更受硬件限制,运行速度缓慢。第二代索引工具已能初步实现今天索引工具的基本功能,如:索引行的生成、词频表、短语表的生成,甚至是词语搭配的计算(如 TACT)。

第三代语料库工具以 WordSmith 为代表,这些软件主要在 Windows 或其他图形界面操作系统中运行。WordSmith 是商业软件,AntConc 为功能相近的免费替代软件。这两款软件最能代表第三代语料库分析工具,两者都拥有广泛的用户群体。WordSmith 各版本主要功能划分为三大模块,即:词汇索引(Concord)、主题词(KeyWords)、词频表(WordList)。到 5.0 版本

(2008 年)以后,WordSmith 增加了框合结构(ConcGram)功能,但从界面功能划分看,主要还是维持三大核心模块。大模块下还有词簇提取(cluster)和词语搭配等子功能模块。WordSmith 的三大模块成为了通用

语料库软件开发领域的事实标准。

第四代语料库工具主要指基于互联网的语料库网络应用(web application)。这类工具通过浏览器与服务器的交互,将语料库与检索工具融为一体。这些工具基于数据库和索引技术,检索响应时间快,用户体验好,一般用来处理大型语料库,如 BNC。但这些工具的灵活性不够,用户通常无法(或很难)处理本地语料库,同时,因受索引格式和数据量的限制,检索语法一般较为简单,不支持复杂检索。

因此,当前及今后很长一段时间,第三代和第四代语料库分析工具将会并存。从研究者的角度看,第三代语料库软件更能满足个人的实际研究需要。我们所开发的工具即属于第三代语料库工具。本研究主要关注单语语料库,双语和多语语料库的分析处理并不涉及。

1.3 第三代语料库工具的不足

WordSmith 和 AntConc 是目前较有代表性的第三代语料库分析工具,但前者不支持正则表达式,且界面复杂,不易操作;后者较为易用,但功能较少、计算效率较低,处理语料时容易死机或意外退出。

因此,在借鉴 WordSmith 和 AntConc 优点的基础上,本研究开发了 PowerConc 软件,对语料库软件的功能进行了梳理和扩充,并在软件易用性和计算效率上进行了优化。PowerConc 由许家金、梁茂成、贾云龙设计,贾云龙负责程序开发。PowerConc 软件的开发是对此前北外语料库语言学团队开发的 Keywords plus、Collator、Colligator(许家金、熊文新,2009)等系列软件的整合。

2 基于 R-gram 的 PowerConc 软件的设计与开发

PowerConc 软件使用 Delphi 语言开发,支持 Windows 操作系统。与以往语料库分析软件相比,PowerConc 具有以下特点:

2.1 支持多种格式和不同语言的语料

PowerConc 在底层是基于 ANSI 编码的,除了支持英文,ANSI 编码在不同语言的操作系统中对应不同的字符集,以支持相应的操作系统默认语言,如:在简体中文系统中,ANSI 代表 GB2312 编码;在日文操作系统

中,ANSI 代表 Shift-JIS 编码。因此,PowerConc 在不同语言操作系统中,可同时支持英文和该操作系统所对应的默认语言。对简体中文操作系统而言,PowerConc 至少能同时处理英文和中文两种语言的文本。

PowerConc 未来会以独立版本的形式支持 Unicode 编码的文本,以便能处理多种语言。因字符编码和存储长度等原因,Unicode 版软件效率会远低于 ANSI 版。若仅是处理某一种语言(操作系统默认语言)或英语语料,使用 ANSI 编码不失为两全其美的选择。

PowerConc 按语言特征将语料分为两大类:

(1) 无空格的连续文本:这类语料以字符为基本单位,字符间无空格分隔。未分词的汉语和日语等语言构成的语料,都属于这一类。

(2) 以空格分词的文本:这类语料以词为基本单位,词与词之间以空格分隔。英语语料属于这一类,分词后的汉语、日语等语言构成的语料也属于这一类。

PowerConc 按语料加工的程度将语料分为三大类:

生语料:主要指未经加工处理的原始语料,如:

面对严重灾情,广东省各级党组织和广大党员,带领广大群众万众一心抗击冰雪灾害。

Corpus Linguistics is empirical by nature.

切分语料:分词后的语料,如:

会议听取了有关阿坝藏族羌族自治州汶川县等地灾情的汇报。

标注语料:对语料进行词性赋码后产生的语料,如:

Miss_NNB Green_NP1 came_VVD in_II secretly_RR

国家_n 体育馆_n 瞬间_nt 成为_v 欢乐_a 的_u 海洋_n

基于以上分类标准,在简体中文操作系统环境下,ANSI 版的 PowerConc 软件共支持五类语料:英文生语料、英文标注语料、中文生语料、中文切分语料、中文标注语料。

PowerConc 目前只支持“词汇_码”格式的标注语料,即:词汇在前,标注码在后,词码间以下划线“_”连接。对于英文生语料,PowerConc 将自动在标点前添加空格,以去除标点符号对检索的干扰。

本文将主要以英文语料为主,对 PowerConc 的各项功能进行介绍。

2.2 支持 R-gram

基于 N 元组的检索分析是对单词的扩充,它突破了研究单位长度的限制。PowerConc 除了检索及统计功能完全支持 N 元组外,其搭配计算、词表生成和关键词计算也完全支持 N 元组,PowerConc 将单词视作一元组(uni-gram)处理。

正则表达式是语料库分析的利器,它的最大特点

之一在于其高度的概括性,使研究者可以按字符种类、字符数量和字符位置三个维度描述字符特征,使研究的范围得到了极大的丰富。

PowerConc 在 N 元组中加入了对正则表达式的支持,使 N 元组得到了扩展。本研究把基于正则表达式的 N 元组称为 R-gram。

在描述能力方面,R-gram 非常强大,它既可以描述具体的语言单位(如单词或短语),也可以对抽象的语言现象(如,动词+名词、a+名词+of)进行描述;R-gram 不像单词和 N 元组那样受长度的限制。总之,R-gram 继承了正则表达式的优点,具有高度灵活性和抽象度,可以对长度和种类都不确定的语言单位进行描述,词汇和 N 元组都可以看作是 R-gram 的子集。对于赋码语料,R-gram 的优势表现得尤为突出,可以对不同数据类型的语言单位进行描述。

R-gram 除了用于检索外,也可用于搭配计算。PowerConc 支持针对 R-gram 的搭配计算,用户可自定义一个 R-gram 列表,PowerConc 将在指定跨距内计算这些 R-gram 和节点(node term)的搭配强度,这扩大了搭配研究的范围,使研究者可以对更为抽象和复杂的搭配现象进行研究。

正则表达式的缺点是语法抽象复杂,可读性差,不便掌握,对于初学者而言,出错的几率很高。为使 R-gram 便于操作和容易理解,我们设计了一种全面兼容正则表达式的简易语法——Smart Input,以降低用户的检索难度。

2.3 支持 Smart Input 语法

Smart Input 语法主要包含以下内容,软件会将用户输入的 Smart Input 语法自动转换成相应的正则表达式:

(1) “@”(@ 符):@ 放在单词原形前;表示将匹配该单词的所有屈折形式,如:@ be 将匹配 am、are、be、been、being、is、was 和 were 等 8 种形式;

(2) “#”(# 号):# 放在词性类别码(如:n、v、adj 等)之前,表示将匹配该词性大类对应的各词性码子类,对 CLAWS (C7) 赋码语料,#n 将匹配 NN、NN1、NN2、NP 等 22 类表示名词(n)的词性码;

(3) “*”(星号):对于以空格分割的文本代表任意一个词,如:be * at 将匹配 be good at、be amazed at 等;* 对于无空格分割的汉语文本代表一个字,如:国 * 将匹配“国家”、“国际”等;

(4) 词性码检索:对于词性赋码语料,可直接输入词性码或码串进行检索,如:对 CLAWS (C7) 赋码语

料,NN1 将匹配 man_NN1、world_NN1 等;

(5) 词形检索:对于词性赋码语料,可直接输入单词或短语进行检索,如:对 CLAWS(C7)赋码语料,look 将匹配 look_NN1、look_VV0、look_VVI 等;

(6) 正则表达式检索:用户可直接编写正则表达式进行检索。

以上(1)~(5)可混合使用在同一表达式中,如: @be * #prep 将匹配 is capable of、was involved in、be responsible for 等内容。

综上所述,Smart Input 语法的引入使 R-gram 变成了一种可读性高、描述能力强的表述体系。表 2 列出了一些 R-gram 应用实例。

表 2 Smart Input 语法

类型	R-gram	表达式含义
词性(部分)	#n	匹配名词
	#v	匹配动词
	#adj	匹配形容词
词汇原型	@be	匹配 be 的不同屈折形式
	@go	匹配 go 的不同屈折形式
POS 码	NN1	匹配 POS 码为 NN1 的语言单位,如:part_NN1
	JJ	匹配 POS 码为 JJ 的语言单位,如:local_JJ
单词	books	匹配词汇部分为 books 的语言单位,如:books_NN2、books_VVZ 等
	can	匹配词汇部分为 can 的语言单位,如:can_VM、can_NN1 等
正则表达式	所有正则表达式	略
叠加表达式	JJ NN1	匹配 increasing_JJ number_NN1、integral_JJ part_NN1 等
	@be @good	匹配 are_VBR best_RRT、be_VBI better_RRR 等
混合表达式	#v #adv	匹配 are_VBR still_RR、be_VBI more_RGR 等
	@be #adj #prep	匹配 be_VBI capable_JJ of_IO、being_VBG involved_JJ in_II 等
	NN1 @do not #v	匹配 agriculture_NN1 does_VDZ not_XX enable_VVI 等等
	do * #prep	匹配 do_VD0 agree_VV1 to_II、do_VD0 attend_VVI with_IW 等

2.4 支持基于文件或文件分组的分布统计

语料库研究经常会涉及对比分析,如按时间、地域、性别、话语者类型(母语还是学习者)、学习者阶段等维度进行对比。分布数据可以直观地显示出语料之间的差异,对于对比性研究具有非常重要的意义。

PowerConc 支持两种分布:

(1) 基于语料库文件的分布:以语料库文件进行分布统计,如图 1。

(2) 基于语料库文件分组的分布:按用户指定的分组条件,将语料库文件分成若干个组,然后以文件分组为单位进行统计。分组条件由针对文件名的正则表达式构成。比如,在配置文件中用正则表达式“0[1-9]”将相关文件归为一组,组名为“AmE”;文件名满足

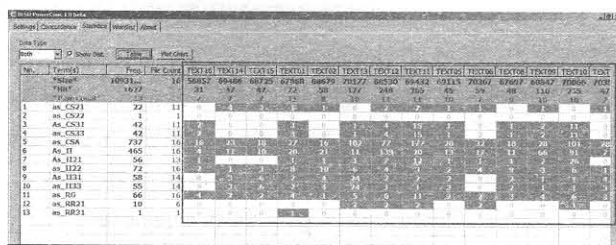


图 1 按文件进行分布统计

正则表达式“1[0-9]”的语料库文件归为另一种,组名为“BrE”。基于正则表达式的分组方式,具有高度的灵活性,同一个语料库文件可以出现在不同的分组之中。图 2 中,各语料库文件已按指定分组原则,被归入了“AmE”和“BrE”两组中。

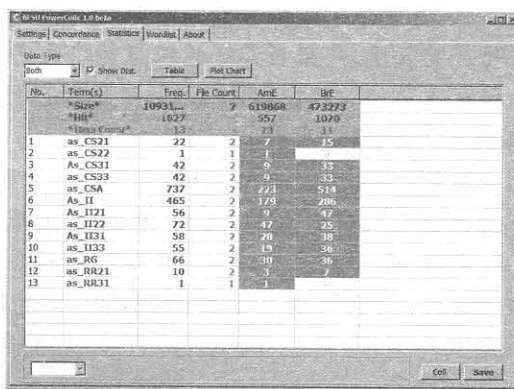


图 2 按用户指定分组条件呈现统计结果

2.5 功能结构清晰

应用软件不仅要实现特定的功能,还需要根据用户习惯对这些功能合理规划,以提高易用性。鉴于此,PowerConc 在设计过程中,对语料库分析和研究常用的各种功能进行了梳理和分类,淘汰了不常用的功能,加入了多项以往语料库软件没有的功能,同时归并到统一的框架下,使所有功能一目了然,当用户熟悉了一个模块后,可以很快掌握其他模块,有效降低软件的学习

表 3 PowerConc 各模块功能描述

	I 检索相关分析(微观分析)			II 词表相关分析(宏观分析)	
	1. 词汇索引	2. 结果统计	3. 搭配计算	1. 词表生成	2. 主题词计算
1. 参数设置	√	√	√	√	√
2. 数据计算	√	√	√	√	√
3. 分布统计	x	√	x	√	x
4. 结果显示	√	√	√	√	√
5. 结果排序	√	√	√	√	√
6. 结果过滤	√	x	√	√	√
7. 结果抽样	√	x	x	x	x
8. 结果保存	√	√	√	√	√

成本和操作难度。

PowerConc 将语料库分析归为微观和宏观两大类,词汇索引主要是对某个或某类语言现象进行的有针对性的分析,属微观分析;词表是对语料库整体情况的考察,属宏观分析。

PowerConc 共五个模块,检索相关的功能包括三个子模块,词表相关的分析包含两个子模块:

(I) 检索相关功能(微观分析)

(1) 词汇索引(Concordance):根据输入内容进行检索并返回索引行;

(2) 结果统计(Statistics):对检索命中的内容进行统计;

(3) 搭配/类联接计算(Collocation & Colligation):根据索引行计算检索结果的搭配信息。

(II) 词表相关功能(宏观分析)

(1) 词表生成(N-gram list):根据指定的数据类型和长度生成词表;

(2) 主题性计算(Keyness):根据参考词表计算主题词,或主题短语。

每个模块的操作又可细分为参数设置、数据计算、分布统计、结果显示、结果排序、结果过滤、结果抽样、结果保存八个类别。

3 PowerConc、WordSmith、AntConc 的对比

PowerConc 不是对 WordSmith 等软件的重写或复制,PowerConc 有其独立的设计原则。

继承发扬:吸收以往语料库软件中那些广为接受的功能(如:词汇索引、词表等),在充分考虑用户体验的基础上,删繁就简,对功能的设计、展示和实现方式进行优化,以降低学习成本和操作难度,并将它们纳入 PowerConc 的统一设计框架中,使软件设计具有高度一致性;其次,对这些功能进行扩展,丰富已有功能;再次,对算法进行优化,提高程序处理数据的能力和效率。

将 PowerConc 与 AntConc 和 WordSmith 进行对比,可以归纳出以下几方面差异:

学习成本和操作难度:三款软件中 PowerConc 的学习成本和操作难度最低,具有一般软件操作的常识和语料库的基本知识的研究者,都可以很快上手。WordSmith 的学习成本和操作难度最高,它的功能非常繁杂,即使经验丰富的研究者也很难快速掌握它的全部功能,WordSmith 6.0(最新版)的说明书已达 415 页。AntConc 学习难度居中,但配置复杂,不便操作。

功能划分和界面布局:PowerConc 完全以研究者的视角来进行功能划分和界面布局,全部功能的设计遵守统一规范,用户可举一反三。WordSmith 的功能规划和界面设计不合理,不符合常规软件的设计原理,若不借助说明书,仅靠界面本身提供的信息,很难进行操作。AntConc 在 WordSmith 基础上进行了一定的优化,但在局部设计上,缺乏连贯性。

核心功能:PowerConc 和 AntConc 对 WordSmith 的功能进行了取舍,分别实现了 WordSmith 的词汇索引、词表、关键词计算三大核心功能。AntConc 几乎是对 WordSmith 三大功能的简化和重写,相对 WordSmith 而言没有本质的变化。PowerConc 以 R-gram 为基础重新设计,使语料库软件的功能得到了扩展。

功能创新:WordSmith 不支持正则表达式,这使它的功能受到了极大的限制。WordSmith 历史较久,版本众多,但每一个新版本的变化并不大,通常只是加入了个别新功能或修正个别小错误,而核心功能几乎没有多少改变。AntConc 加入了对正则表达式的支持,但相对 WordSmith 而言,其功能并无创新。PowerConc 最大的创新是对 R-gram 和 Smart Input 的支持,它使语料库研究的范围得到了扩展。同时,PowerConc 对数据分布统计也具有很好的实现。

算法效率:因为功能设计的差距,三个软件很难直接对算法效率进行对比。整体而言,AntConc 的算法效率最低,对数据量的大小较敏感,容易死机或意外退出。WordSmith 的效率一般,算法没有进行优化,个别计算要耗费大量的时间。PowerConc 在算法优化上做了大量尝试,最大程度上避免了数据拷贝带来的资源浪费,同时一些模块使用了缓存方式以避免信息的重复计算,使计算效率大大提升。

可扩展性:PowerConc 基于面向对象的方法开发,核心功能被封装在不同的类(Class)中,实现了界面和功能的分离。这使 PowerConc 具有非常好的扩展性,一方面,可以不断对现有功能进行升级和维护,也可以加入新功能;另一方面,可以利用这些核心功能类,开发出衍生产品。这些优势是 AntConc 和 WordSmith 等软件不具备的。WordSmith 的升级和维护几乎是用打补丁的方式进行的,这使 WordSmith 的安装包越来越大,操作越来越复杂。WordSmith 6.0 的安装文件有 54MB,安装后有 108 个文件,而 AntConc 和 PowerConc 都是绿色软件,无需安装,AntConc 3.2.4(最新版)的大小是 4.4MB,PowerConc 只有 1.5MB。

相信 PowerConc 的开发将有效促进语料库语言学

研究的开展。

感谢梁茂成教授对本文提出的宝贵的修改意见。



参 考 文 献

- [1] Baker, P., Hardie, A. & McEnery, T. (eds.). *A Glossary of Corpus Linguistics*[M]. Edinburgh: Edinburgh University Press, 2006.
- [2] Hunston, S. *Corpora in Applied Linguistics* [M]. Cam-

bridge: Cambridge University Press, 2002.

- [3] McEnery, T. & Hardie, A. *Corpus Linguistics: Method, Theory and Practice*[M]. Cambridge: Cambridge University Press, 2012.
- [4] Sinclair, J. *Corpus, Concordance, Collocation*[M]. Oxford: Oxford University Press, 1991.
- [5] 许家金,熊文新. 基于学习者英语语料的类联接研究:概念、方法及例析[J]. 外语电化教学, 2009(3).

The Design and Development of the R-gram Based Corpus Analysis Tool ‘PowerConc’

XU Jia-jin, JIA Yun-long

(National Research Center for Foreign Language Education, Beijing Foreign Studies University, Beijing 100089, China)

Abstract: This paper describes the innovative corpus tool PowerConc developed by the authors themselves. In its implementation, such key functionalities of corpus tools as concordancing, wordlist generation and keyword analysis, were redesigned, expanded and optimized. What underlies the whole design of PowerConc is the inventive synergy of regular expressions and N-gram, namely, R-gram proposed in this study. The R-gram feature allows for flexibility in concordancing, exhaustive listing of linguistic units, and key terms of varying length, and more likely than not, enables analyses of linguistic structures with uncertain words or categories. To minimize the inconvenience of operation, Smart Input has been introduced to facilitate easy search with enhanced returned hits. The user-friendly PowerConc is object-oriented software. The key functions have been packaged in different classes as a dynamic link library (*.dll) file, independent from the user interface, which warrants easy maintenance and expandability. It is hoped that PowerConc will be conducive to corpus-based research in its own way.

Key words: Corpus Tools; PowerConc; R-gram; Corpus

摘要转载（20 篇）

编号：AB2013A01

题名：**基于多语体语料的英语双及物结构研究——以 give 为例**

作者：王琳琳

单位：上海外国语大学

期刊：外国语(上海外国语大学学报)

年份：2013

期号：01

主题词：双及物结构；多语体语料；信息因素；give

摘要：文章以最典型的双及物动词 give 为例，通过调用英国国家语料库中口语、小说、报纸和学术期刊语料分库，分析和统计双宾语结构和与格结构的例句中接受者和客体的定指性、代词性和成分长度，从结构内部、结构之间和语体之间三个层面对比呈现了接受者和客体的上述特征。研究发现，就双宾语结构而言，诸语体中接受者和客体的语序基本体现了信息流对语序的制约。然而，就与格结构而言，接受者和客体的信息特征在上述语体中表现各异，信息因素对其语序及双及物结构句式使用的制约力度在不同语体间存在差异，部分语体中甚至存在与前人的发现不一致的现象。在此基础上，文章计算了联系各个因素预测上述不同语体中双及物结构的句式使用的准确度。结论是，从信息角度解释和预测双及物结构的句式使用还应考虑语体差异。

编号：AB2013A02

题名：**不同过程类型双及物小句的语类变异研究**

作者：张跃伟

单位：渤海大学外国语学院

期刊：外语与外语教学

年份：2013

期号：03

主题词：双及物小句；语料库；语类变异；过程类型

摘要：本文以系统功能语法为指导，采纳基于用法的语言观，结合真实语料中双及物小句的实际用例，围过程类型对其功能语义特征进行分析。研究表明双及物经验意义表征潜势巨大，双及物过程包含能产性最强的物质过程以及心理、言语、行为和关系过程等 5 种类型。实证考察揭示了不同语类语料库中双及物小句的过程类型、典型动词及语义配置类型的分布特征，即物质过程的强势分布、“GIVE”双及物小句的“一家独大”地位及“人-人给予”语义类别的优势分布数量。分析表明：这些特征主要是由语义动因引起的，语类特征也是影响双及物小句实际使用模式的重要因素。

编号：AB2013A03

题名：**从“意义即使用”哲学观到语料库的“意义单位”探究**

作者：何安平

单位：华南师范大学外国语言文学学院

期刊：外语与外语教学

年份: 2013

期号: 03

主题词: “意义即使用”; “语境意义论”; 语料库语言学; 意义单位; 语境意义模式

摘要: 本文以文献述评的方式探讨从 Wittgenstein“意义即使用”的语言哲学观、到 Firth 的“语境意义论”、再到 Sinclair 基于语料库短语对意义单位进行理论和实践探究的历史沿革。首先分别简述三者的核心理念及相互关联; 然后述评国内部分学者基于语料库对多种形态短语意义单位的探究成果; 最后归纳三个重要观点。

编号: AB2013A04

题名: 英汉形容词拷贝式的语料库与类型学研究

作者: 罗思明

单位: 复旦大学外文学院; 宁波大学科学技术学院

期刊: 外语教学

年份: 2013

期号: 02

主题词: 形容词; 拷贝式; 可能性; 句法功能; 语言类型

摘要: 本文基于权威语料聚焦考察英汉形容词拷贝式的可能性、句法功能和语言类型学特征。研究发现, 两者是共性与个性并存。共性体现在: 拷贝式都受制于音节、语义和形态; 均可作前置定语、后置定语(偶尔)、表语, 都不受程度词修饰, 都不用于祈使句、评价句和进行体, 都具有肯定极性倾向; 都属“有形容词拷贝式语言”; 个性体现在: 汉语多音节形容词一般无拷贝式, 少数英语多音节形容词有拷贝式; 英语形容词拷贝式偶尔有比较级或最高级; 英语形容词拷贝式不直接作谓语, 汉语可以; 英语形容词拷贝式可作宾补, 汉语英语形容词拷贝式可作“得”字补语; 英语形容词拷贝式不用于结果构式, 汉语方言偶尔可以; 英语形容词拷贝式偶尔可作状语, 汉语形容词拷贝式常作状语; 英语是“单一形容词拷贝式语言”, 汉语是“复杂形容词拷贝式语言”。

编号: AB2013A05

题名: “下来”、“下去”作补语时的句法语义限制及其认知解释——一项基于语料库的研究

作者: 刘正光、孙一弦

单位: 湖南大学

期刊: 外语学刊

年份: 2013

期号: 01

主题词: 下来、下去; 句法语义限制; 认知

摘要: 趋向动词“下来”、“下去”在功能和意义上有很多相似之处, 但各自也有不同的特点。根据语料库统计, “下来”主要表示空间意义, 占语料的 72.56%; “下去”主要表示时间意义, 占语料的 49.5%。二者不能互换使用时, “下来”凸显终点, “下去”强调终点不可见或难以确定。二者不宜互换使用时, “下来”空间上立足终点, 时间上表示已然; “下去”空间上立足起点, 时间上表示未然。二者可以互换使用时, 语义上具有弱空间性, 或者是与“了、被、以后”等句法标记连用。认知上, “下来”与“有界”相关, “下去”与“无界”相联系, 同时也反映出主观识解和人类经验在认知参照点构建过程中的重要作用。

编号: AB2013A06

题名: **学术文本中短语序列的语篇行为**

作者: 李晶洁、卫乃兴

单位: 东华大学; 北京航空航天大学

期刊: 外语教学与研究

年份: 2013

期号: 02

主题词: 短语序列; 语篇行为; 语篇组织功能; 共选; 篇章指示

摘要: 学术文本通常对语篇模式和词汇表达有严格要求, 因此学术作者需要在写作中使用特定的语言形式来实施某些具体的研究行为, 即学术语篇行为。本文基于 NEW-JDEST 语料库证据, 描述与概括学术文本中三类凸显语篇行为: 陈述观点或事实、报道和篇章指示。数据显示, 学术文本的语篇行为有其各自的典型词语实现方式, 某些序列被频繁用来实施具体的语篇组织功能。某种意义上, 典型的复现序列成为了具体功能的标志。

编号: AB2013A07

题名: **否定系统与语类变体——一项基于英语学术语篇的否定概率研究**

作者: 刘英

单位: 广西大学外国语学院

期刊: 外语与外语教学

年份: 2013

期号: 01

主题词: 否定系统; 概率; 语类变体; 英语学术语篇

摘要: 本文基于系统功能语言学系统和概率语法理论, 首先探讨英语否定系统的构建和扩展, 然后运用语料库研究方法, 以英语学术语篇为例, 发现否定系统中词汇语法选项的使用概率, 解读其否定意义的构建, 以此揭示否定系统概率与语类变体之间的关联性。研究表明, 在英语学术语篇中, 英语否定系统的不同选项呈不均匀分布, 词汇否定出现频率远远高于语法否定和半语法否定, 说明英语学术语篇否定意义的构建呈词汇化倾向, 这一特征是识别英语学术语篇的重要依据。

编号: AB2013A08

题名: **基于布朗家族语料库的英语现在进行体的历时考察**

作者: 章柏成、许家金

单位: 重庆交通大学外国语学院; 北京外国语大学中国外语教育研究中心

期刊: 外语教学

年份: 2013

期号: 01

主题词: 现在进行体; 布朗家族语料库; 口语化; 语法化; 英语语法

摘要: 本研究基于 600 万词次, 时间跨度近 50 年的三代英、美英语语料库, 考察英语现在进行体在形式分布及意义演变方面的历时变化。数据显示, 20 世纪 60 年代以来, 英语现在进行体的使用总体持续增长, 增速在 21 世纪初显著加快; 在语体、语态方面, 除了表现出总体增长趋势外, 也存在一些变异。英语现在进行体表达将来义和主观态度的情况也持续增

长。研究还表明,书面语中的英语现在进行体存在明显口语化趋势;同时,英语现在进行体也表现出语义虚化的语法化倾向,这与英语语法的整体演变趋势一致。

编号: AB2013A09

题名: **二语学术写作介入标记语的使用与发展特征: 语料库视角**

作者: 徐昉

单位: 南京大学外国语学院

期刊: 外语与外语教学

年份: 2013

期号: 02

主题词: 学术英语; 二语学术写作; 介入标记语; 语料库

摘要: 本次研究基于语料库视角,对中国学习者本硕博阶段 120 篇英语学位论文和 240 篇国际核心期刊论文进行对比分析,力图描述二语学术写作介入标记语的使用与发展特征。研究发现,总体上学生论文比期刊论文显著少用介入标记语,但是比期刊论文显著多用指令语标记。从发展走势看,三种介入标记的使用呈现三种不同的发展特点,明显的差别出现在本科与硕士论文之间。这些研究结果对二语学术写作教学有一定的启示。

编号: AB2013A10

题名: **中国英语专业学生英语议论文句法复杂性研究**

作者: 徐晓燕、王维民、熊燕宇、蒋婧、潘小燕、孙念红

单位: 西南交通大学

期刊: 外语教学与研究

年份: 2013

期号: 02

主题词: 句法复杂性; 单位长度; 单位密度; 句式类型; 紧缩子句

摘要: 本研究对比分析中国英语专业学生与英语本族语者在议论文中句法复杂性的差异。指标有单位长度(W/T 和 W/C)、单位密度(C/T 和 DC/C)、四类句子、从句、紧缩子句和限定与非限定动词的被动形式。结果表明,中国英语专业学生的单位长度和密度与语言能力变化方向一致。长度指标显著低于本族语者,密度在多数情况下也显著偏小。简单句在 4 年级和写作高分组下降显著,状语、宾语和表语从句随年级的升高只呈下降趋势。中国学生对上述句型的使用显著多于本族语者。紧缩子句和被动句只在写作高分组中显著增多,但低于本族语者。

编号: AB2013A11

题名: **基于笔语语料库的大学英语学生词汇发展研究**

作者: 栾岚

单位: 上海外国语大学; 哈尔滨工程大学

期刊: 外语学刊

年份: 2013

期号: 02

主题词: 纵向发展; 词汇多样性; 词汇复杂性; 词汇密度; 非英语专业学生

摘要：作为二语习得研究的重要组成部分，词汇评估不仅有助于了解学习者第二语言的掌握程度和学习者的共有特点，而且对促进二语教学具有重要意义。本文以非英语专业(大学英语)学生的定时作文为语料，通过自建小型历时语料库和 Complete Lexical Tutor 提供的 Vocabprofile 功能来分析语料库中的数据，结果表明：词汇发展总体来说是随着年级的增加呈现上升趋势，但上升幅度不显著。高频词中学术词汇(AWL)的变化基本上呈线性增长；词汇的多样性不断提高，学习者用词范围变宽；比起学习者词汇多样性和复杂性的变化幅度，词汇密度的变化呈现出缓慢上升趋势。

编号：AB2013A12

题名：**情态序列作为言语行为套语的习得研究——用基于语料库的方法**

作者：刘陈艳

单位：中北大学人文社会科学学院

期刊：解放军外国语学院学报

年份：2013

期号：01

主题词：情态序列；请求；套语；言语行为

摘要：本文通过现场录音的方法收集不同水平学习者的口语语料，以考察他们在使用情态序列实施请求过程中表现出来的特征。结果表明：中国英语学习者在大学阶段偏好使用规约类间接言语行为中的间接询问类来实施请求言语行为，具体表现为“情态动词+代词+动词”的情态序列；不同水平学习者在情态序列选择的范围上没有变化，没有体现出语用语言能力习得的差异；社会维度的变化对学习者情态序列的选择没有产生明显影响。

编号：AB2013A13

题名：**中国学习者英语口语产出中程式化序列使用型式研究**

作者：邓耀臣

单位：大连海事大学外国语学院

期刊：外语与外语教学

年份：2013

期号：03

主题词：程式化序列；使用型式；心理短语库；口语产出策略

摘要：本文以 SECOPETS 语料库为基础，采用中介语对比分析的研究方法，从对程式化序列的依赖程度和高频程式化序列功能特征两方面探讨了中国英语学习者口语产出中常用程式化序列的使用型式。结果表明，中国英语学习者口语产出呈现程式化序列多用、少用和误用的混合型式。中国英语学习者明显过多依赖英语中高频程式化序列而过少使用英语中频数较低但对英语本族语来说具有较高心理凸显性的惯例化表达方式。就功能而言，中国英语学习者明显过多地依赖立场标记语和语篇标记语，而过少使用回应标记语、修正标记语和话轮调节语。研究结果在一定程度上反映了中国英语学习者和英语本族语者心理短语库构成模式的差异，为我们理解中国英语学习者口语产出策略提供了新视角。

编号：AB2013A14

题名：**基于平行语料库的《红楼梦》意义显化翻译考察——以霍译本林黛玉人物特征为例**

作者：姚琴

单位：江苏大学

期刊：外语教学与研究

年份：2013

期号：03

主题词：平行语料库；《红楼梦》霍译本；林黛玉；意义显化

摘要：本文通过对《红楼梦》的霍译本与其所据原著程乙本进行平行语料库的源语/目标语语际对比和分析，考察《红楼梦》霍译本中有关林黛玉人物描写的翻译特征。本研究的观察点是平行语料库中 Dai-yu's 后面的人物特征名词。为了增强代表性，本研究选取出现频次极高的 said Dai-yu 作为佐证，重点关注与 said 搭配的情状副词词义。语料库研究表明：对林黛玉的人物特征，霍译本以释意、增添等方式进行意义显化翻译。本研究指出，霍译本进行意义显化翻译的动因主要是出于对英汉两种语言文化差异的敏感和对西方译语受众的考量。

编号：AB2013A15

题名：基于语料库的汉英会议口译中 some 的应用研究

作者：潘峰、胡开宝

单位：上海交通大学外国语学院

期刊：解放军外国语学院学报

年份：2013

期号：02

主题词：some；语料库；汉英会议口译；应用；动因

摘要：Some 一词在汉英会议口译中常作为限定词及不定代词被译员频繁应用。相较于书面翻译而言，some 在汉英会议口译中每万词的标准频次几乎为前者的 3 倍。译员在口译中应用 some 除基本对应原文语义外，还常用来显化原文隐含意义、弱化及少量的强化原文语义。译员频繁地使用 some 既有英语语言规范上的动因也有译员自身因素上的动因。

编号：AB2013A16

题名：口译语料库研究的原则与方法

作者：张威

单位：北京语言大学外国语学院

期刊：外语电化教学

年份：2013

期号：01

主题词：语料库；口译研究；原则；方法

摘要：口译语料库是口译教学与研究的一个新领域，也是提高口译研究质量的重要途径与方法。口译语料库研究的主要原则包括：定量与定性相平衡、描述与解释相配合、微观与宏观相结合、共时与历时相协调、证实与证伪相促进、基础性与应用性并重等。口译语料库研究主要包括基于语料库(corpus-based)、语料库驱动(corpus-driven)、语料库指导(corpus-informed)等三种设计方案，具体采用语料库内嵌式统计分析、社会统计学分析、多库对比参照等三种方法。同时，应该客观认识。译语料库建设与研究方法的不足，考虑与其他口译研究方法综合应用，以多类型客观数据，最大限度地反映口译现象的本质。

编号: AB2013A17

题名: **基于平行语料库和可比语料库的英汉预期标记对比研究**

作者: 王艳伟

单位: 上海交通大学; 新乡学院

期刊: 现代外语

年份: 2013

期号: 02

主题词: 预期标记; 对应关系; 翻译对等项; 互译度

摘要: 预期标记不论在广义的语用标记还是在特定的言据范畴中都有其重要作用。本文通过平行语料库与可比语料库相结合的方法探讨了英语预期标记 *actually*、*really* 和 *in fact* 与其汉语对应标记实际上、真正和事实上的跨语言对应关系及其互译度双向不均衡的成因。研究表明: 1) 英汉预期标记的语义场相互重叠的程度不同, 双向对应关系呈现出不同的图景; 2) 英汉预期标记的互译强度不同, 双向对应关系并不均衡; 3) 就预期标记及其翻译对等项在源语和译语中的分布来说, 从英语出发和从汉语出发所得出的差异显著性检验的结果并不相同。我们发现, 搭配特征、语义映射和语用含意有可能是导致英汉预期标记互译度失衡亦即英汉双向对应关系不均衡的因素。

编号: AB2013A18

题名: **基于平行语料库的“大家”的对应研究**

作者: 易焱、王克非

单位: 北京外国语大学英语学院; 北京外国语大学中国外语教育研究中心

期刊: 外语与外语教学

年份: 2013

期号: 03

主题词: 平行语料库; 对应研究; 人称代词; 大家

摘要: 本文基于英汉、汉英双向平行语料库对现代汉语人称代词“大家”和它在英语中的对应项进行描写和分析。研究发现, “大家”在英语中的功能对等项有 *everybody* 等全称代词、第一、二、三人称复数代词、*people*、*the crowd* 等指人的集体名词、*other* 等他称代词、*both* 等指双数的代词、无施事被动式、*there be* 构式和零形式。我们运用对比语言学、翻译学、语义学和功能类型学的概念和理论解释了这些对应现象。本文也证明了双向平行语料库在上述几个领域的研究中的作用。

编号: AB2013A19

题名: **学术论文作者自称与身份构建——一项基于语料库的英汉对比研究**

作者: 吴格奇

单位: 杭州师范大学外国语学院

期刊: 解放军外国语学院学报

年份: 2013

期号: 03

主题词: 学术论文; 自我指称; 作者身份

摘要: 本研究基于自建语料库, 对比分析了英汉学术论文作者通过自称构建身份的方式, 发

现：研究者和话语构建者身份构成了英语论文作者身份的主要方面；汉语论文作者注重于构建自己的研究者身份。英汉身份构建的不同源于英汉学术文化、对话语产生和理解的原则以及社会准则的差异。研究结果对跨文化背景下的论文写作及其教学有指导意义。

编号：AB2013A20

题名：**基于语料库的高校英语教师课堂元话语多维功能机制研究**

作者：闫涛、张丽云

单位：天津医科大学医学英语与健康传媒学院；齐齐哈尔大学外国语学院

期刊：解放军外国语学院学报

年份：2013

期号：02

主题词：语料库；元话语；课堂话语分析；学术口语语篇；多维功能机制

摘要：元话语不仅仅是一种语言现象，而且也是一种修辞和语用手段。外语教师课堂元话语在课堂教学中发挥着重要作用。本研究通过自建语料库，结合定量研究和定性分析考察元话语在高校英语教师课堂话语中的应用情况，揭示元话语手段在相对广阔的语言环境中通过其相互协调和相互作用实现其多维功能性，并总结其运行机制，为外语教学和教师发展提供启示，扩展元话语的研究领域。

2013 上半年国内刊发的语料库相关论文题录信息

- [1] 安雪花, 2013, 语料库在外语写作教学中的应用研究评述及展望,《西安外国语大学学报》(2): 54-58。
- [2] 鲍苏宁, 2013, 用语料库检索工具分析文学作品——分析短篇小说《一小时的故事》,《天津市经理学院学报》(1): 93-94。
- [3] 毕冉、陈桦, 2013, 中国英语学习者音调模式的纵深研究,《外语与外语教学》(1): 50-54。
- [4] 布占廷, 2013,《评价研究的语料库路径——短语学与评价语言》评介,《现代外语》(1): 105-107。
- [5] 布占廷, 2013, 基于评价理论的语言学书评标题研究,《外语与外语教学》(4): 53-57。
- [6] 蔡美兰, 2013, 医学英语动词名化研究——基于美国当代英语语料库的研究,《中国高等医学教育》(4): 13-14。
- [7] 蔡强, 2013,《红楼梦》红色文化英译分析——基于汉英平行语料库的对比研究,《名作欣赏》(17): 170-171。
- [8] 蔡强, 2013, 网络在线语料库资源在英语教学中的应用探讨,《中国报业》(6): 139-140。
- [9] 蔡强、张建平, 2013, 从语料库主题词角度分析中西方高等教育理念的差异——基于中国与英语国家高校网页“学校简介”内容的比较,《现代教育科学》(5): 16-19。
- [10] 曹静芬, 2013, 日语接尾辞“-的”的使用状况——基于中日对译语料库的研究,《兰州教育学院学报》(6): 121-122。
- [11] 曹贤文, 2013, 留学生汉语中介语纵向语料库建设的若干问题,《语言文字应用》(2): 127-134。
- [12] 曾琳, 2013, 语料库在翻译研究中的适用性探讨,《英语广场(学术研究)》(5): 38-39。
- [13] 曾庆敏, 2013, 基于语料库的美国总统就职演讲主题词研究,《重庆第二师范学院学报》(1): 61-64。
- [14] 曾晓霞, 2013, 浅析日语教学中语料库的运用,《吉林省教育学院学报(上旬)》(1): 47-48。
- [15] 陈浩, 2013, 第二语言口语非流利产出的重复现象研究,《解放军外国语学院学报》(1): 72-77。
- [16] 陈佳, 2013, 以语料库为基础的译者痕迹研究,《辽宁教育学院学报》(3): 97-99。
- [17] 陈杰, 2013, 英语专业四级词汇语料库在基础阶段教学中的应用,《吉林省经济管理干部学院学报》(1): 96-98。
- [18] 陈晋秋, 2013, 基于语料库的应用语言学学术语篇发展趋势研究,《海南师范大学学报(社会科学版)》(1): 107-111。
- [19] 陈丽娜, 2013, 英语口语语料库中引语动词的分析,《海外英语》(8): 271-272。
- [20] 陈林、戴玉群, 2013, 论汉译英教学中参考语料库与分析语料库的应用,《长春工程学院学报(社会科学版)》(2): 72-75。
- [21] 陈宁, 2013,“good”话语标记语结构、位置、共现特征和功能的四维探析——基于自建小型口语语料库的研究,《疯狂英语(教师版)》(2): 173-178。
- [22] 陈宁、张晓枚、陈苏, 2013, 中医院校公共外语的医学特色语料库教学方法,《文教资料》(6): 180-182。
- [23] 陈平, 2013, 基于语料库的高级英语数据驱动教学模式探讨,《陕西教育(高教版)》(21): 27-28。
- [24] 陈清, 2013, 基于语料库的大学生口语词块研究,《沧州师范学院学报》(1): 81-83。
- [25] 陈伟、刘丹亮, 2013, AntConc 在话语标记语教学研究中的应用,《语文学刊(外语教育

教学)》(4): 119-120。

- [26] 陈小莹, 2013, 藏语拉萨话语音合成语料库的研究与建立,《科技信息》(9): 13-14。
- [27] 陈妍玲, 2013, 论数据驱动学习方式对高中英语词汇教学效果的影响,《英语教师》(4): 31-36。
- [28] 陈悦, 2013, 基于自建英语新闻报道语料库对 change 一词的语义韵分析,《绥化学院学报》(6): 65-68。
- [29] 程璐璐、白志刚、冯德煜, 2013, 语料库和 BNC 在英语专业翻译教学中的应用,《沈阳师范大学学报(社会科学版)》(3): 181-183。
- [30] 程钰琼, 2013, 基于语料库的中国英语学习者期刊论文中的弱化修饰词研究,《武汉冶金管理干部学院学报》(1): 57-60。
- [31] 程子璐, 2013, 基于语料库的译者性别身份研究——以高慧勤《雪国》译本为例,《柳州职业技术学院学报》(1): 79-83。
- [32] 迟秋玲, 2013, 美国当代英语语料库在英语词汇学习中的应用,《科技信息》(4): 241-242。
- [33] 崔建周, 2013, 基于语料库的旅游景区公示语翻译研究,《海外英语》(2): 108-109。
- [34] 崔维霞、王均松, 2013, 国内学科专业语料库研究现状及发展趋势,《西安外国语大学学报》(1): 55-58。
- [35] 德吉梅朵, 2013, 面向藏语语料库建设的数量词的分类研究,《剑南文学(经典教苑)》(1): 228-229。
- [36] 邓耀臣, 2013, SECOPETS 语料库的设计原则及初始研究,《语言教育》(1): 35-39。
- [37] 邓耀臣, 2013, 中国学习者英语口语产出中程式化序列使用型式研究,《外语与外语教学》(3): 60-65。
- [38] 丁玲玲, 2013, 基于语料库 BNC 和 COCA 对 SORRY 的对比研究,《青年文学家》(13): 156。
- [39] 丁俏蕾, 2013, 基于英语语料库的高职英语互动式教学模式应用探析,《职教通讯》(15): 40-42。
- [40] 董爱华, 2013, 基于语料库的广告英语语言特色分析,《北京印刷学院学报》(1): 32-35。
- [41] 杜爱玲, 2013, 文学作品的语料库研究范式综述,《怀化学院学报》(3): 92-94。
- [42] 杜海紫, 2013, 基于语料库的商务英语口语教学方法研究,《语文学刊(外语教育教学)》(3): 124-126。
- [43] 杜金榜、罗红秀, 2013, 信息型法律语料库及其在法律语篇分析中的作用,《云梦学刊》(1): 135-140。
- [44] 樊斌、韩存新, 2013, Wolfgang 的语料库语言学思想及其认识价值,《大连海事大学学报(社会科学版)》(2): 112-116。
- [45] 樊晓萍, 2013, 基于 BCCWJ 语料库的数据驱动学习在基础日语教学中的实践,《哈尔滨职业技术学院学报》(3): 82-83。
- [46] 范春香, 2013, 中英语言文学类学术论文摘要语类及语言特点对比分析——一项基于自建语料库的研究,《河北旅游职业学院学报》(2): 72-76。
- [47] 范俊军, 2013, 汉语方言自然口语语料库建设的几个基本问题,《学术研究》(2): 153-158。
- [48] 范心乐、郭惠、刘立鹏, 2013, 语料库语言学研究方法在司法鉴定中的应用,《黑龙江科技信息》(10): 38。
- [49] 方喜、詹宏伟, 2013, 从新闻语篇的副词使用看中日媒体意识形态的差异性——基于语料库的解读,《伊犁师范学院学报(社会科学版)》(2): 95-98。
- [50] 飞龙、高光来、鲍玉来, 2013, 蒙古语电话语音语料库的建立,《内蒙古大学学报(自然科学版)》(3): 320-323。
- [51] 冯萌, 2013, 基于语料库的汉译文学作品的显化实证研究——以《名利场》两个中译本

为例,《海外英语》(7): 227-228。

- [52] 冯全功, 2013, 花开一朵表两枝——《(红楼梦)中英文语料库的创建及应用研究》评介,《外语艺术教育研究》(1): 32-35。
- [53] 付海燕、顾毅, 2013, 语料库驱动的英汉语篇人称系统研究,《南昌教育学院学报》(4): 139-146。
- [54] 付瑶, 2013, 商务英语媒体语言概念隐喻分析——基于语料库的研究,《江苏外语教学研究》(1): 74-77。
- [55] 傅慧英, 2013, 非英语专业大学生母语对英语词汇搭配的负迁移现象——基于中国学习者英语语料库的研究,《考试周刊》(31): 61-63。
- [56] 傅顺、罗永胜、杨劲松, 2013, 教学型医学英语语料库的应用研究,《广东医学院学报》(1): 101-103。
- [57] 高定国、索郎桑姆, 2013, 大型藏语平衡语料库建设中样本类别号的研究,《西藏大学学报(自然科学版)》(1): 54-58。
- [58] 高枫舒, 2013, 浅谈语料库语言学及其教学应用,《科技信息》(10): 183。
- [59] 高俊伟、孔芳、朱巧明, 等, 2013, 语料对中文名词短语指代消解影响研究,《中文信息学报》(3): 61-68。
- [60] 高文成、张丽芳, 2013, 英汉语双名词内向结构认知对比研究——基于 BNC 和 CCL 语料库,《西安外国语大学学报》(2): 1-6。
- [61] 耿延, 2013, 语料库在大学生英语写作中的应用研究,《山西煤炭管理干部学院学报》(2): 162-163。
- [62] 龚频, 2013, 如何通过建立学习者英语口语语料库加强英语语音教学,《西北成人教育学报》(2): 80-82。
- [63] 郭嘉, 2013, 外语信息化教学新探——以数据驱动学习模式为例,《闽西职业技术学院学报》(1): 79-83。
- [64] 郭云云, 2013, 基于语料库的高考英语命题分析——短语动词视角,《英语教师》(4): 42-49。
- [65] 郭振伟, 2013, 我国隐喻语料库研究现状与展望,《浙江师范大学学报(社会科学版)》(3): 38-43。
- [66] 国防, 2013, 基于语料库的话语标记语“well”研究,《集美大学学报(哲学社会科学版)》(2): 76-82。
- [67] 何安平, 2013, 从“意义即使用”哲学观到语料库的“意义单位”探究,《外语与外语教学》(3): 44-48。
- [68] 何国锦、施晗, 2013, 基于语料库的汉语“语块”识别初步探索,《语文学刊》(12): 7-8。
- [69] 贺文照, 2013, 平行语料库辅助翻译实践实证研究,《嘉兴学院学报》(2): 64-69。
- [70] 侯羽, 2013,《红楼梦》英译研究的新阶段——读《<红楼梦>中英文语料库的创建及应用研究》有感,《英语广场(学术研究)》(1): 28-29。
- [71] 胡丹, 2013, 基于公安英语语料库的心理词库功能研究,《外国语文》(1): 75-79。
- [72] 胡健、胡鸿, 2013,《语料库、认知与致使构式》述评,《山东外语教学》(1): 109-112。
- [73] 胡丽, 2013, 基于语料库的中国非英语专业学习者议论文写作中转折连接词的研究,《巢湖学院学报》(1): 142-148。
- [74] 胡丽珍, 2013, 语料库视野下汉语大型辞书的常用动词释义——基于“啃”的个案研究,《辞书研究》(1): 13-22。
- [75] 华金、杨柳, 2013, 基于语料库的不同水平中国学习者情态动词使用研究,《大学英语(学术版)》(1): 127-130。

- [76] 黄丹婵, 2013, 英语专业大学生英语作文中的主题词结构及语义网络——基于 SWECCCL 语料库的分析,《桂林师范高等专科学校学报》(1): 93-99。
- [77] 黄海军、王昌米, 2013, 语料库和应用翻译——以蚕学馆旧址纪念碑碑文的翻译为例,《上海翻译》(1): 66-68。
- [78] 黄立波、朱志瑜, 2013, 国内英汉双语平行语料库建构与研究现状及展望,《当代外语研究》(1): 45-49。
- [79] 黄缅, 2013, 英国学术语料库与学术英语写作研究,《黑河学刊》(1): 123-124。
- [80] 黄沐云, 2013, 语料库驱动下的《21 世纪大学英语》词表分析,《语文学刊(外语教育教学)》(3): 10-12。
- [81] 黄晓梅、黄勇, 2013, 语料库语言学研究发展现状浅析,《新西部(理论版)》(8): 150-151。
- [82] 黄亚萍, 2013, 基于语料库研究中国英语专业学生对双宾动词的使用情况,《南昌教育学院学报》(3): 157-158。
- [83] 黄莺、陈光先, 2013, 中医英语语料库建设初探,《科技信息》(16): 65-67。
- [84] 吉红, 2013, 高职日语写作课程语料库的建设初探,《语文学刊(外语教育教学)》(1): 158-160。
- [85] 吉红, 2013, 基于语料库的高职日语词汇教学改革探索——以复合动词「一だす」为例,《中国校外教育》(12): 112。
- [86] 江利, 2013, 基于语料库研究的语言教学探索,《理论界》(5): 174-176。
- [87] 江莉、王荃、张四红, 等, 2013, 基于语料库的中医术语翻译中 essence 和 spirit 的差异性研究,《科教文汇(上旬刊)》(3): 109-111。
- [88] 姜文东、任娟, 2013, 基于语料库的《格列佛游记》文本词汇特征分析,《山东广播电视大学学报》(2): 55-56。
- [89] 蒋丽平, 2013, 基于语料库的 IT 英语语体特征及教学研究,《湖南工业职业技术学院学报》(3): 93-95。
- [90] 金婧, 2013, 基于语料库的译者主体性研究——以《雾都孤儿》十一个译本为例,《文教资料》(8): 30-32。
- [91] 靳松, 2013, 基于口译教学的语料库建设初探,《社科纵横》(3): 176-178。
- [92] 孔来信, 2013, 如何利用基于语料库进行大学互动式英语教学,《教育教学论坛》(23): 164-165。
- [93] 孔来信、陈文翻, 2013, 基于语料库的网络辅助英语教学,《课程教育研究》(11): 111。
- [94] 孔雁, 2013, 后殖民语境下基于语料库的汉语新词翻译标准与策略,《作家》(8): 189-190。
- [95] 邝计嘉、蒋庆胜, 2013, 自建小型学生作文语料库以及运用,《大学英语(学术版)》(1): 78-80。
- [96] 蓝瞻瞻, 2013, 计算机辅助翻译双语语料库的研建,《电子测试》(9): 203-204。
- [97] 雷冬平, 2013, 语料库视野下汉语大型辞书的疑难词语释义——基于“诡随”的个案研究,《辞书研究》(1): 9-12。
- [98] 雷雨, 2013, 基于语料库的“把 NP+V 下”与“把 NP+V 下来”对比研究,《华文教学与研究》(1): 80-86。
- [99] 李冰, 2013, 日本留学生生造词偏误分析——基于中介语语料库的调查研究,《语文学刊》(4): 17-19。
- [100] 李德超, 2013,《英汉翻译中的汉语译文语料库研究》介评,《中国翻译》(2): 43-45。
- [101] 李芳兰, 2013, 语料库数据驱动的英语语言意识建构,《北京化工大学学报(社会科学版)》(2): 61-65。
- [102] 李海、张淑静, 2013, 中国英语学习者写作产出词汇的心理语言属性发展趋势——

一项基于 WECCL 和 MRC 数据库的研究,《现代外语》(2): 174-181。

- [103] 李恒、吴铃, 2013, 手语语料库建设基本方法,《中国特殊教育》(3): 37-41。
- [104] 李金满、吴芙芸, 2013, 类型学概括与二语学习者汉语关系从句产出研究,《外语教学与研究》(1): 80-92。
- [105] 李晶洁、胡文杰, 2013, 学术文本中的语言形式共选——基于语料库的实证研究,《外语与外语教学》(4): 62-66。
- [106] 李晶洁、卫乃兴, 2013, 学术文本中短语序列的语篇行为,《外语教学与研究》(2): 200-213。
- [107] 李景、李国鹏、汪滨, 2013, 标准文献语料库的构建与应用研究,《标准科学》(5): 6-11。
- [108] 李美霞、焦瑗琿, 2013, 基于语料库的英语逻辑结果程式语语义韵研究,《外语教学》(2): 21-26。
- [109] 李梦骁、刘永兵, 2013, 基于语料库的中学生英语写作中话语标记语的使用研究,《现代中小学教育》(5): 55-57。
- [110] 李敏慧、冯健, 2013, 基于语料库的大学生英语情态动词 MUST 使用情况探析,《黑龙江教育(高教研究与评估)》(2): 30-31。
- [111] 李奇志, 2013, 论水利英语语料库的创建及其应用,《河南科技学院学报》(1): 81-83。
- [112] 李炎芳、安圆圆、胡艳超, 2013, 基于 CLEC 语料库的大学英语词汇教学研究,《吉林省经济管理干部学院学报》(2): 123-125。
- [113] 李振营, 2013, 中国大学生英语口语词汇特征研究,《辞书研究》(1): 35-41。
- [114] 李政文, 2013, 基于语料库的名词“doubt”的对比研究,《语数外学习(英语教育)》(3): 12-13。
- [115] 李梓, 2013,《语料库数据驱动的专业文本语义韵研究》评介,《读与写(教育教学刊)》(1): 33-34。
- [116] 梁红梅, 2013, 语料库视角下的语言教学观,《英语研究》(2): 73-78。
- [117] 梁健丽, 2013, 多模态教学语料库在教师教育中的应用,《大学教育》(6): 62-63。
- [118] 梁健丽, 2013, 中学英语网络书面语料的采集,《教学月刊(中学版)》(1): 38-40。
- [119] 林巧文, 2013, 网络语料库背景下词块习得对英语写作的实证研究,《洛阳师范学院学报》(1): 96-98。
- [120] 刘彬、戈玲玲、李广伟, 2013, 基于语料库的大学英语写作教学平台建设及其模式研究,《湖南第一师范学院学报》(1): 51-55。
- [121] 刘陈艳, 2013, 情态序列作为言语行为套语的习得研究——用基于语料库的方法,《解放军外国语学院学报》(1): 62-66。
- [122] 刘春芬, 2013, 基于语料库关于应用语言学方向学术论文中四字词丛对比研究,《安徽文学(下半月)》(2): 100-102。
- [123] 刘蕊, 2013, COCA 语料库对 EFL 学习者写作影响的研究,《宝鸡文理学院学报(社会科学版)》(2): 190-194。
- [124] 刘寒冰, 2013, 对 COCA 和 COHA 的比较研究,《陕西教育(高教版)》(Z1): 89-90。
- [125] 刘汉武, 2012, 基于越南学生中介语语料库的趋向补语偏误分析,《国际汉语学报》(2): 168-173。
- [126] 刘火群, 2013, 运用语料库分析方式解读海明威作品的艺术特色,《长春理工大学学报》(2): 72-73。
- [127] 刘佳、韩丽娜, 2013, 低碳经济英语语料库建设与应用,《天津大学学报(社会科学版)》(2): 178-182。

- [128] 刘剑, 2013, 汉英翻译中疑问标记的缺失与转换——基于《呐喊》平行语料库的分析,《西南科技大学高教研究》(2): 45-48。
- [129] 刘金玲, 2013, 基于语料库的英语作文自动在线批改方式评析,《中国科教创新导刊》(10): 35-36。
- [130] 刘娟, 2013, 语料库的英语学术语篇转述动词研究,《赤峰学院学报(汉文哲学社会科学版)》(3): 168-169。
- [131] 刘克强、沈映梅, 2013,《三国演义》中英文平行语料库的研制与创建,《红河学院学报》(2): 34-39。
- [132] 刘丽娟, 2013, 基于语料库的论文摘要显化翻译研究,《中北大学学报(社会科学版)》(3): 53-56。
- [133] 刘露露、屈琼, 2013, 基于语料库的《红楼梦》两译本词汇特征的对比分析,《三峡大学学报(人文社会科学版)》(S1): 81-84。
- [134] 刘娟路, 2013, 基于语料库的美国总统就职演说之人际意义研究,《保定学院学报》(3)。
- [135] 刘小燕, 2013, 学习者翻译语料库中英语被动句汉译特点及其成因研究,《中北大学学报(社会科学版)》(3): 34-38。
- [136] 刘英, 2013, 否定系统与语类变体——一项基于英语学术语篇的否定概率研究,《外语与外语教学》(1): 27-30。
- [137] 刘盈盈、罗森林、冯扬, 等, 2013, BFS-CTC 汉语句义结构标注语料库,《中文信息学报》(1): 72-80。
- [138] 刘宇婷、李文中, 2013, “适者生存”在《白牙》中的阐释——基于语料库的《白牙》文本检索分析,《海外英语》(2): 181-182。
- [139] 刘正光、孙一弦, 2013, “下来”、“下去”作补语时的句法语义限制及其认知解释——一项基于语料库的研究,《外语学刊》(1): 60-64。
- [140] 龙绍赞, 2013, 英语专业学生议论文中情态序列使用频率的年级表征,《中国外语教育》(1): 3-14。
- [141] 龙桃先、陈丽云、张德敬, 等, 2013, 在少数民族地区建英语学习者语料库进行高频动词研究探析,《长春理工大学学报》(3): 157-158。
- [142] 娄宝翠, 2013, 基于语料库的研究生学术英语语篇中外壳名词使用分析,《外语教学》(3): 46-49。
- [143] 卢静, 2013, 基于语料库的译者风格综合研究模式探索——以《聊斋志异》译本为例,《外语电化教学》(2): 53-58。
- [144] 栾岚, 2013, 基于笔语语料库的大学英语学生词汇发展研究,《外语学刊》(2): 126-130。
- [145] 罗景凤, 2013, 基于语料库的《了不起的盖茨比》文体分析,《乐山师范学院学报》(3): 50-53。
- [146] 罗思明, 2013, 英汉形容词拷贝式的语料库与类型学研究,《外语教学》(2): 6-10。
- [147] 吕倩, 2013, 自建小型口译语料库用于口译教学,《青年文学家》(7): 123。
- [148] 马彩梅、朱益平, 2013, 陕西省旅游景区公示语汉英平行语料库的设计与建设,《西安外国语大学学报》(1): 113-116。
- [149] 马创新、陈小荷, 2013, 基于XML的《论语》与其注疏文献对齐语料库的知识表示,《图书情报知识》(1): 107-113。
- [150] 马武林、蒋艳, 2013, 混合式教学模式下非外语专业大学生英语听力存在问题及其成因分析——一项基于词频分析的研究,《外国语文》(1): 138-143。

- [151] 马莹, 2013, 基于 Range 的《新概念英语》词汇分析,《黔南民族师范学院学报》(1): 60-62。
- [152] 马媛, 2013, 三本大学学生英语听力语料库建库刍议,《科教文汇(中旬刊)》(3): 90-91。
- [153] 毛文伟, 2013, 现代日语书面语均衡语料库应用研究,《日语学习与研究》(2): 17-24。
- [154] 毛艳文, 2013, 科技英语词汇习得方法探析——基于语料库的科技英语高频词汇的研究,《常州工学院学报(社科版)》(1): 113-117。
- [155] 莫如愚, 2013, 语料库在翻译教学中的应用研究,《佳木斯教育学院学报》(5): 388-389。
- [156] 莫秀兰, 2013, 基于语料库的英语专业毕业论文写作教学,《湖北科技学院学报》(1): 74-75。
- [157] 慕媛媛、朱纯深, 2013, 在线翻译与双语写作教学/自学平台建设中的语料精加工,《当代外语研究》(1): 56-62。
- [158] 倪晓黎, 2013, 时间状语 now 置于句首不一定倒装——基于利用网络语料库对英语教学的研究,《考试周刊》(3): 91-92。
- [159] 倪云, 2013, 社会科学论文英文摘要中的模糊限制语——一项基于平行语料库的研究,《郑州航空工业管理学院学报(社会科学版)》(3): 73-76。
- [160] 牛桂玲, 2013, 基于语料库的中国学术期刊论文摘要汉英翻译显化特点分析,《西安外国语大学学报》(2): 112-116。
- [161] 潘峰、胡开宝, 2013, 基于语料库的汉英会议口译中 some 的应用研究,《解放军外国语学院学报》(2): 95-100。
- [162] 彭宣红, 2013, 基于语料库的《老人与海》译本情感忠实性比较探究,《外国语文》(1): 126-129。
- [163] 钱恒、钮萌, 2013, 基于语料库的英语存在句研究,《剑南文学(经典教苑)》(1): 328-329。
- [164] 钱源, 2013, 基于香港法律语料库的纽马克翻译理论应用研究,《剑南文学(经典教苑)》(2): 227。
- [165] 秦建华, 2013, 基于语料库的 CONTRIBUTE 的语义韵对比研究,《海外英语》(1): 234-236。
- [166] 秦建华, 2013, 基于语料库的程度副词类联接和语义韵对比研究,《牡丹江大学学报》(2): 73-77。
- [167] 秦建华, 2013, 基于语料库的形容词增强语语义韵对比研究——以 totally 和 terribly 为个案,《黄冈师范学院学报》(1): 94-97。
- [168] 秦建华, 2013, 基于专门用途英语(ESP)语料库的词汇研究——探索大学英语教师向 ESP 教师转型的途径,《内蒙古民族大学学报(社会科学版)》(2): 89-93。
- [169] 秦建华, 2013, 语料库数据驱动对二语词汇类联接和语义韵习得的促进作用,《鄂州大学学报》(2): 55-57。
- [170] 秦静, 2013, 基于语料库的《红楼梦》“石头叙事”及其英译研究,《江苏科技大学学报(社会科学版)》(1): 59-62。
- [171] 权晓辉、吴江, 2013, 语料库及数据驱动学习辅助高中英语词汇教学的优势,《甘肃科技》(7): 72-75。
- [172] 冉玉体, 2013, 语料库辅助下的词块习得与写作能力的相关性研究,《英语研究》(2): 68-72。
- [173] 任德国, 2013, 基于语料库的“由”字句习得偏误分析,《安徽文学(下半月)》(1):

110-111。

- [174] 任吉, 2013, 关于语料库写作教学的实践,《江苏教育》(10): 51-52。
- [175] 任艳、陈建生, 2013, 语料库文体学在文学批评中的应用,《中国轻工教育》(1): 28-31。
- [176] 尚鲁冰、李世超, 2013, 浅谈高职英语语料库的建设,《教育教学论坛》(23): 272-274。
- [177] 邵霞、孙雪娥, 2013, 基于语料库的大学英语写作教学初探,《山西煤炭管理干部学院学报》(2): 160-161。
- [178] 邵雨清, 2013, 基于语料库对 experience 作动词时语义韵的研究,《英语广场(学术研究)》(3): 23-24。
- [179] 盛文渊, 2013, 语料库在日语专业毕业论文指导中的应用,《读与写(教育教学刊)》(1): 22-23。
- [180] 盛亚南、吴芙芸, 2013, 指量结构与汉语关系从句共现时的不对称分布及其原因——一项基于真实口语语料库的研究,《现代外语》(2): 150-157。
- [181] 施春宏、张瑞朋, 2013, 论中介语语料库的平衡性问题,《语言文字应用》(2): 117-126。
- [182] 石静、吴云芳、邱立坤, 等, 2013, 基于大规模语料库的汉语词义相似度计算方法,《中文信息学报》(1): 1-6。
- [183] 石璞璞, 2013, 基于语料库的高职行业英语教学探究,《课程教育研究》(10): 107-108。
- [184] 石志亮, 2013, 正则表达式的获取及在语料库研究中的应用,《西安外国语大学学报》(2): 33-36。
- [185] 史敏, 2013, 大学英语课堂教学语料库相关问题研究,《教书育人》(12): 80-81。
- [186] 史文霞, 2013, 中国博士生科技论文中转述的意向结构,《外语教学》(3): 27-31。
- [187] 宋改荣、孙淑芹, 2013, 科技论文摘要中情态动词的语境顺应性解读——一项基于语料库的研究,《阜阳师范学院学报(社会科学版)》(1): 65-68。
- [188] 宋庆伟、匡华、吴建平, 2013, 国内语料库翻译学 20 年述评(1993—2012),《上海翻译》(2): 25-29。
- [189] 宋晓燕, 2013, 基于语料库的英语词汇教学研究,《琼州学院学报》(1): 135-136。
- [190] 宋亚军, 2013, 语料库数据驱动的学生自主学习能力培养模式探究,《河南机电高等专科学校学报》(1): 83-86。
- [191] 苏琦, 2013, 基于语料库的学习者汉语四字格口译质量研究,《黑龙江史志》(9): 321-322。
- [192] 苏向丽、胡晓清, 2013, 韩语背景 CSL 学习者量度形容词的混用分布特征分析——基于国别化中介语语料库的调查,《鲁东大学学报(哲学社会科学版)》(2): 40-45。
- [193] 孙巍, 2013, 基于语料库的中国英语学习者常用动词搭配行为分析——以动词 develop 为例,《大连海事大学学报(社会科学版)》(2): 117-120。
- [194] 谭业升, 2013, 译者的意象图式与合成概念化——基于语料库方法的《红楼梦》“社会脸”翻译研究,《外语与外语教学》(3): 55-59。
- [195] 唐丽玲, 2013, 中国英语专业学习者情态动词语义发展的语料库考察,《当代外语研究》(6): 32-36。
- [196] 汪茂林, 2013, 国内语料库中介语研究现状分析,《文教资料》(12): 19-20。
- [197] 王大鹏, 2013, 车载语音导航系统的语音语料库构建探究,《电子测试》(6): 94-95。
- [198] 王大鹏, 2013, 基于 COCA 语料库的英语同义词辨析——以 obtain 等同义词为例,《渤海大学学报(哲学社会科学版)》(1): 72-74。
- [199] 王丰元, 2013, 基于 ESP 教学的高职英语语料库建设研究——以湖州职业技术学院为例,《职业技术教育》(5): 38-40。
- [200] 王芙蓉、王敏, 2013, 基于语料库的应用语言学学术英语词块研究及词块表的创建,

《英语研究》(2): 79-84。

- [201] 王冠, 2013, DDL 在二语词汇教学中的应用,《湖北函授大学学报》(6): 125-126。
- [202] 王冠, 2013, 基于 COCA 在线语料库的大学英语词汇教学探析,《大学英语(学术版)》(1): 120-122。
- [203] 王冠, 2013, 基于语料库的中国大学生情态动词使用情况研究(英文),《语文学刊(外语教育教学)》(3): 24-25。
- [204] 王惠霞, 2013, 基于笔语语料库的 GET 使役用法词汇型式研究,《乐山师范学院学报》(2): 67-70。
- [205] 王鉴棋, 2013, 系统透视英语分裂不定式,《外国语文》(2): 87-90。
- [206] 王静, 2013, 基于语料库与云计算的高职英语写作错误分析,《南通纺织职业技术学院学报》(1): 88-92。
- [207] 王均松、田建国, 2013, 词典编纂的语料库方法,《郑州航空工业管理学院学报(社会科学版)》(1): 71-73。
- [208] 王乐娜, 2013, 基于语料库辅助的鲁迅小说《孔乙己》两英译本的译者风格研究,《剑南文学(经典教苑)》(1): 46-47。
- [209] 王磊、郑书云, 2013, 基于语料库对 2012 年美国总统大选辩论语篇的人际意义解读,《湖北广播电视大学学报》(2): 93-94。
- [210] 王琳琳, 2013, 基于多语体语料的英语双及物结构研究——以 give 为例,《外国语(上海外国语大学学报)》(1): 45-54。
- [211] 王敏、刘丁, 2013, 中国学习者英语学术论文手稿中立场标记词块使用研究,《现代外语》(2): 190-197。
- [212] 王敏、郑权, 2013, 手语语料库建设的价值、内容及策略,《中国教育信息化》(5): 24-26。
- [213] 王楠、李光赫, 2013, 基于语料库的中日因果复句对照研究——以倒装「カラ」句为中心,《现代语文(语言研究版)》(4): 136-137。
- [214] 王佩, 2013, 语料库驱动下的自主学习模式应用初探,《湖北科技学院学报》(2): 87-88。
- [215] 王平, 2013, 数据库汉字学刍议——以魏晋南北朝石刻用字数据库与断代汉字发展史研究为例,《中国文字研究》(1): 159-165。
- [216] 王瑞, 2013, 中国英语学习者转折类话语标记语的使用调查——一项基于语料库的研究,《陇东学院学报》(1): 120-121。
- [217] 王少爽、高乾, 2013, 语料库翻译学的建构与拓展——王克非《语料库翻译学探索》评述,《中国翻译》(2): 39-42。
- [218] 王苏, 2013, 语料库视角下《匆匆》英译文风格赏析,《兰州教育学院学报》(4): 127-128。
- [219] 王恬, 2013, 语篇主旨与模糊限制语的使用——一项基于语料库的跨语域对比研究,《疯狂英语(教师版)》(2): 160-168。
- [220] 王伟, 2013, 新闻英语语料库研制与应用的探索性研究,《长春理工大学学报(社会科学版)》(2): 159-160。
- [221] 王晓娟, 2013, 基于语料库的现代汉语派生词类型研究,《现代语文(语言研究版)》(4): 35-36。
- [222] 王雪梅, 2013, 机关文书语料库的开发和应用,《现代语文(语言研究版)》(4): 104-105。
- [223] 王艳伟, 2013, 基于平行语料库和可比语料库的英汉预期标记对比研究,《现代外

- 语》(2): 182-189。
- [224] 王宗迎, 2013, 语料库辅助高中英语阅读教学案例分析,《山东师范大学外国语学院学报(基础英语教育)》(1): 9-13。
- [225] 魏际兰, 2013, 基于语料库的大学生语码转换的认知研究——以汉英语码转换为例,《江苏外语教学研究》(1): 90-92。
- [226] 魏金凤、陈丽勤, 2013, 基于语料库的中国英语学习者书面语中减弱语使用情况调查,《广西师范学院学报(哲学社会科学版)》(2): 137-141。
- [227] 吴格奇, 2013, 学术论文作者自称与身份构建——一项基于语料库的英汉对比研究,《解放军外国语学院学报》(3): 6-11。
- [228] 吴军莉, 2013, 基于语料库英语写作与精读教学——寻求提高医学专业类本科学生英语词汇搭配及写作能力的新途径,《佳木斯教育学院学报》(6): 449-451。
- [229] 吴娴, 2013, 国内口译语料库的发展现状及存在的问题,《海外英语》(6): 154-155。
- [230] 吴晓龙、顾毅, 2013, 基于语料库的博物馆展览概述英译的文体研究,《长春理工大学学报(社会科学版)》(3): 148-150。
- [231] 吴永荣, 2013, 基于语料库的“亲自”“亲手”对比研究,《语言与翻译》(1): 24-28。
- [232] 奚军秀, 2013, 语料库在大学英语教学中的运用,《传奇·传记文学选刊(教学研究)》(4): 52。
- [233] 奚丽云、高泉、赵常友, 2013, 语料库方法在大学英语词汇教学中的应用,《云南农业大学学报(社会科学版)》(2): 102-107。
- [234] 习晓庆, 2013, 基于语料库的学术英语综述性动词对比研究,《中州大学学报》(1): 69-70。
- [235] 夏云, 2013, 英译汉文学语言形合特征的历时研究,《外语教学与研究》(1): 105-116。
- [236] 项炜、金澎, 2013, 大规模语料库上的 Stanford 和 Berkeley 句法分析器性能对比分析,《电脑知识与技术》(8): 1984-1986。
- [237] 邢慎宝, 2013, 原本《玉篇》的俗写字分类研究——基于魏晋南北朝石刻语料库的调查,《湖北民族学院学报(哲学社会科学版)》(2): 140-143。
- [238] 熊保霞, 2013, 语料库在大学英语教师教学中的应用,《科教文汇(上旬刊)》(2): 95-102。
- [239] 熊莉, 2013, 基于语料库的“各 V 各的 N”的考察,《现代语文(语言研究版)》(4): 47-48。
- [240] 徐昉, 2013, 二语学术写作介入标记语的使用与发展特征:语料库视角,《外语与外语教学》(2): 5-10。
- [241] 徐华、刘丹丹、钱龙华, 等, 2013, 基于双语依存关系映射的中英文词表构建研究,《中文信息学报》(1): 15-20。
- [242] 徐鹏, 2013, 基于语料库方法在同义词辨析上的应用,《湖北经济学院学报(人文社会科学版)》(4): 117-119。
- [243] 徐山、杜卫锋, 2013, 不可靠语料库的提纯及词权度量指标 IDF 的改进,《微型机与应用》(4): 61-63。
- [244] 徐晓燕、王维民、熊燕宇、蒋婧、潘小燕、孙念红, 2013, 中国英语专业学生英语议论文句法复杂性研究,《外语教学与研究》(2): 264-275。
- [245] 许家金, 2013, 中国学习者英语口语叙事中的话语评价研究,《外语教学与研究》(1): 69-79。
- [246] 许家金、贾云龙, 2013, 基于 R-gram 的语料库分析软件 PowerConc 的设计与开发,《外语电化教学》(1): 57-62。
- [247] 许炜、周俊英, 2013, 话语标记语 I mean 的对比研究,《重庆第二师范学院学报》

(1): 49-52。

- [248] 闫洁, 2013, 基于语料库的国家社科基金外国文学立项热点分析,《河南科技学院学报》(5): 62-64。
- [249] 闫涛、张丽云, 2013, 基于语料库的高校英语教师课堂元话语多维功能机制研究,《解放军外国语学院学报》(2): 59-64。
- [250] 闫卫建, 2013, 基于语料库的同义词语义韵特征研究,《海外英语》(4): 252-253。
- [251] 严苡丹, 2013, 从特色词看目的语文本与源语文本之间的关系——基于语料库的翻译研究,《外语与外语教学》(4): 72-77。
- [252] 杨红, 2013, 小学语文教学语料库的设计与开发,《课外语文(初中)》(2): 56。
- [253] 姚琴, 2013, 基于平行语料库的《红楼梦》意义显化翻译考察——以霍译本林黛玉人物特征为例,《外语教学与研究》(3): 453-463。
- [254] 易焱、王克非, 2013, 基于平行语料库的“大家”的对应研究,《外语与外语教学》(3): 49-54。
- [255] 雍文明, 2013, 医学英语语料库的构建及医学文献英文文本的对比研究,《河南科技学院学报》(3): 73-75。
- [256] 雍文明, 2013, 语料库辅助医学英语教材编写,《新乡医学院学报》(3): 238-240。
- [257] 俞珏, 2013, 中国学习者英语短语动词教学研究——基于 Keep 的语料库调查,《长春师范学院学报》(3): 180-183。
- [258] 郁伟伟, 2013, 基于语料库的中外大学简介话语体裁对比分析,《绥化学院学报》(6): 72-77。
- [259] 袁远, 2013, 高校英语教学中语料库应用情况调查与分析——以安徽省为例,《湖北经济学院学报(人文社会科学版)》(3): 196-197。
- [260] 张宝林, 2013, 关于通用型汉语中介语语料库标注模式的再认识,《世界汉语教学》(1): 128-140。
- [261] 张碧琼、余毓国, 2013, 基于语料库的"Heritage"和"Legacy"的语义韵比较研究,《英语广场(学术研究)》(2): 32-33。
- [262] 张丹, 2013, 基于语料库探索汉语集体量词的英译——以“一群”为例,《长春理工大学学报(社会科学版)》(2): 149-151。
- [263] 张冬艳, 2013, 语料库辅助大学英语词汇教学的探讨——以个人语料库为视角,《武汉船舶职业技术学院学报》(2): 87-89。
- [264] 张帆, 2013, 基于语料库的高职聋生程度副词使用偏误分析——以“很、更、最、极”为例,《现代特殊教育》(5): 35-37。
- [265] 张帆, 2013, 浙江中高职聋生汉字偏误语料库的建设,《教育评论》(2): 120-122。
- [266] 张芳、马广惠, 2013, 对 when-状语从句汉译的个案研究——基于英汉对应语料库,《江苏教育学院学报(社会科学)》(1): 124-128。
- [267] 张付花, 2013, 基于语料库的中外水利论文英文摘要主题词分析,《河南科技学院学报》(5): 55-58。
- [268] 张会平、刘永兵, 2013, 添加关系话语标记语的句法特征分析——一项基于学习者语料库的对比研究,《当代外语研究》(1): 14-19。
- [269] 张乐、卫乃兴, 2013, 学术论文中篇章性句干的型式和功能研究,《解放军外国语学院学报》(2): 8-15。
- [270] 张蕾, 2013, 新闻语篇中经济隐喻的建构功能——一项基于小型自建语料库的研究,《天津外国语学院学报》(1): 7-12。
- [271] 张黎黎、黄永新, 2013,《论语》两译本之主题词分析,《兰州教育学院学报》(2):

144-145。

- [272] 张黎黎、黄永新, 2013, 基于语料库的《论语》二译本对比研究,《河北广播电视大学学报》(2): 57-59。
- [273] 张黎黎、黄永新, 2013, 欧·亨利短篇小说的语料库文体学分析,《短篇小说(原创版)》(2): 47-48。
- [274] 张黎黎、黄永新, 2013, 中国景点英文导游词主题词分析,《长春理工大学学报(社会科学版)》(1): 140-141。
- [275] 张立飞、严辰松, 2013, 汉语复杂名词短语的独“的”偏好——整合语料库和认知语言学的证据,《外语教学》(1): 30-36。
- [276] 张丽萍, 2013, 基于语料库的就“heart”一词的语义分析,《语文学刊》(6): 42-60。
- [277] 张丽萍, 2013, 基于语料库语言学的立体化外语教材编写探索,《语文学刊(外语教育教学)》(4): 127-128。
- [278] 张丽萍, 2013, 中国英语学习者口语中模糊标签语使用情况分析——基于语料库的对比研究,《中国石油大学学报(社会科学版)》(2): 95-98。
- [279] 张莉, 2013, 语料库辅助的英语阅读教学设计,《中国校外教育》(9): 77。
- [280] 张莉、王惠瑜, 2013, 基于语料库的大学英语教材词汇深度知识研究——以动词为例,《赤峰学院学报(自然科学版)》(8): 227-230。
- [281] 张绵, 2013,《傲慢与偏见》的语料库词汇特征分析,《大学英语(学术版)》(1): 246-248。
- [282] 张绵, 2013, 基于语料库的词汇特征分析——以《爱玛》为例,《安徽文学(下半月)》(4): 37-38。
- [283] 张明权, 2013, 论阐释翻译学的实践维度——“道德”一词英汉互译的语料库分析,《现代语文(语言研究版)》(3): 130-134。
- [284] 张琪, 2013, 基于语料库的简单化、明晰化和规范化研究——以《克兰福德镇》中文译本为例,《吉林广播电视大学学报》(6): 88-89。
- [285] 张茜, 2013, 基于语料库的数据驱动模式在大学英语教学中的运用,《海外英语》(6): 40-41。
- [286] 张仁霞, 2013,《新标准大学英语》教材中国文化元素呈现特征分析——一项基于语料库的调查,《英语教师》(5): 53-58。
- [287] 张威, 2013, 口译语料库研究的原则与方法,《外语电化教学》(1): 63-68。
- [288] 张英、秦洪武、孔蕾, 2013, 英语不定式使用中的问题:基于语料库的研究,《成都师范学院学报》(4): 81-84。
- [289] 张跃伟, 2013, 不同过程类型双及物小句的语类变异研究,《外语与外语教学》(3): 11-15。
- [290] 章柏成, 2013, 大学生英语现在完成时习得的语料库考察——检验“突显度”与“形式优先”假说的新证据,《外语与外语教学》(4): 67-71。
- [291] 章柏成、许家金, 2013, 基于布朗家族语料库的英语现在进行体的历时考察,《外语教学》(1): 42-47。
- [292] 赵华, 2013, 微型文本应用于英语教学的思考,《新乡学院学报(社会科学版)》(2): 137-139。
- [293] 赵继光, 2013, 论语境化、数字化语法教学与研究的课程资源——评基于剑桥国际语料库的《当代英语语法》,《中学生英语(高中版)》(Z1): 125-128。
- [294] 赵建东、高光来、飞龙, 2013, 蒙古语语音合成语料库标注规则的设计,《内蒙古大学学报(自然科学版)》(3): 324-328。

- [295] 赵嫚, 2013, 关于定位词根的一些思考——以“绩效”的词典学和语料库考察为例, 《合肥师范学院学报》(2): 83-85。
- [296] 赵巍, 2013, 语料库辅助大学英语写作教学的研究, 《经济师》(1): 140-141。
- [297] 郑军锋, 2013, play the guitar 还是 play guitar——一项基于语料库的调查, 《湖北函授大学学报》(2): 159-160。
- [298] 郑摄, 2013, 浅析语料库 BNC 在英语介词教学中的应用, 《中国科技信息》(5): 106-107。
- [299] 郑艳群, 2013, 语料库技术在汉语教学中的应用透视, 《语言文字应用》(1): 131-138。
- [300] 郑占国, 2013, 基于语料库的大学生英语辅音错误研究, 《教学研究》(2): 32-35。
- [301] 钟婧嵘, 2013, 浅析语料库视角下的大学基础英语和行业英语的可行性研究, 《考试周刊》(19): 80。
- [302] 周芬, 2013, 基于 COCA 语料库的 V + to + V-ing 结构中主动词特征探究, 《台州学院学报》(2): 44-47。
- [303] 周杰、陈娟, 2013, 贵州对外宣传平行语料库的建设和应用, 《贵州大学学报(社会科学版)》(2): 147-152。
- [304] 周娟、王澜, 2013, 平行语料库在翻译教学中的应用——以《简·爱》译者文体分析为例, 《教育理论与实践》(12): 57-58。
- [305] 周俊英、贺霞, 2013, 基于语料库的英语专业学习者动名搭配错误研究, 《北京教育学院学报》(1): 53-58。
- [306] 周韵, 2013, 基于美国当代英语语料库的英语词汇练习设计, 《海外英语》(5): 267-268。
- [307] 朱纯深、慕媛媛, 2013, 以文本解释力为导向的语料库翻译教学——香港城大翻译与双语写作在线教学/自学平台的设计与试用分析, 《中国翻译》(2): 56-62。
- [308] 朱丽巍, 2013, 语料库检索在大学英语语块教学中应用的可行性研究, 《现代交际》(3): 198-199。
- [309] 朱薇、李敏杰, 2013, 基于语料库的《红楼梦》中“孽障”英译研究, 《盐城师范学院学报(人文社会科学版)》(3): 109-113。

AILA Matters

Academic Publishing and Presenting in a Global Context

Coordinator: Mary Jane Curry

The AILA Academic Publishing and Presenting Research Network held an informal get-together at the American Association for Applied Linguistics conference in Dallas, Texas in mid-March. We discussed plans for putting together our colloquium for AILA, as well as holding an online seminar in the months leading up to AILA and possibly an edited book to result from the colloquium. At present we have about 85 members on a Google Groups list server (AILA-Pub-Ren@googlegroups.com) which is open to anyone.

Interlocutor/Instructor Individual Differences in Cognition and Second Language Acquisition

Convenor: Laura Gurzynski-Weiss
Indiana University
lgurzyns@indiana.edu
www.IndividualDifferencesinSLA.com

The AILA Research Network on Interlocutor and Instructor Individual Differences in Cognition and Second Language Acquisition (SLA) provides a central forum for the systematic investigation into the role of interlocutor and instructor individual differences (IDs) in cognition and behaviour believed to mediate SLA. Specifically, the ReN aims to (1) create awareness of existing and ongoing research on this topic, (2) encourage additional study and collaboration, and (3) support the development and exchange of ideas on projects that explore how interlocutor and instructor IDs influence cognition and behavior related to SLA processes, and what this means for both SLA theory and practice.

The inaugural ReN meeting was held at the American Association for Applied Linguistics in Dallas on 17 March. We are currently seeking

additional abstracts for our AILA ReN symposium, which will focus on the relationship between interlocutor or instructor ID factors (including native language, gender, experience, educational and research background) in relation to cognition and/or behavior regarding factors believed to affect language acquisition in instructed settings (e.g. input/instruction, feedback, tasks, elicitation of student output, etc.). Please send submissions – a title, abstract up to 300 words, 50-word summary and author affiliation – to the ReN convenor, Laura Gurzynski-Weiss, at lgurzyns@indiana.edu by 31 May 2013. Authors will be notified of acceptance in June 2013. Additional information about this and other ReN events can be seen on the website, www.individualdifferencesinsla.com.

Completing its first year, ReN members include seasoned and junior faculty from various countries and theoretical backgrounds, as well as advanced- and beginning-level graduate students. We are always looking to add to our growing community, both in terms of members and research awareness. Please contact the ReN Convenor if you have or are currently conducting studies on interlocutor or instructor IDs, or are interested in learning more about or joining this burgeoning area of research.

Corpus Linguistics and Technology (CLAT)

Convenor: Maocheng Liang

Several important activities have been organized to enhance research in corpus linguistics and corpus applications in Applied Linguistics.

Conferences and symposia

1. Corpus Linguistics in China 2011 (CLIC 2011) convened on 19–20 November, 2011 in Beijing Foreign Studies University (Conference website: <http://www.fleric.org.cn/clic2011/>). This is the first national conference on corpus linguistics in China, jointly organized by the Corpus Linguistics Society of China and members of the ReN. The conference attracted more than 160 researchers from various parts of the country, including Hong Kong and Taiwan. Most of the conveners and participants of the ReN spoke at the conference.
2. Corpus Technologies and Applied Linguistics 2012 convened on 28–30 June, 2012 in Xian Jiaotong Liverpool University (Conference website: <http://www.xjtlu.edu.cn/corpus/>). The conference was jointly organized by the Corpus Linguistics of China and the host university. More than 120 researchers from several countries attended the conference. Most of the conveners and members of the ReN gave presentations at the conference.

3. 2013 Symposium on Corpus Linguistics and Discourse Analysis, to be held on 1–2 June in Beijing Foreign Studies University. The symposium organizer has invited some 50 corpus researchers in China, many of whom are leading scholars in the area in China.

Annual workshops

Every year, members of the ReN organize two workshops on corpus applications, with each workshop attracting some 200 university teachers from a wide range of areas in the country. This year's workshops will take place in early August.

Media Linguistics

Convener: Daniel Perrin

Writing as routine and emergence – tracking creativity and change in the newsroom

At the 2014 International Conference on Writing Research in Paris (<http://conference2014.fr>), the AILA Research Network on Media Linguistics will discuss recent research projects focusing on the interplay of routine and emergence in the dynamic system of newswriting. The research network was formed in 2011 with the mission to “investigate the highly complex and dynamic interplay of language use in public spheres, newsrooms, and news source domains such as business or politics”. It aims at developing “appropriate inter- and trans-disciplinary theories, methodologies, and solutions to practical problems in this emerging field of applied linguistics” (www.aila.info/networks/list-of-research-networks/media-linguistics.html).

A tale of two C's: Comparing English varieties with Crown and CLOB (the 2009 Brown family corpora)

Jiajin Xu and Maocheng Liang

*National Research Centre for Foreign Language Education,
Beijing Foreign Studies University*

1 Background

Since 7 August 2010, the Corpus Linguistics team at Beijing Foreign Studies University had led the construction of two corpora of written contemporary English – Crown and CLOB. With the collaborative efforts of over 147 college English teachers and postgraduate students from over 107 universities across China, the corpus building project was nearly completed in December 2011, after which minor revisions had also been made till February 2012.

One of the underlying assumptions behind the development of the 2009 Brown family corpora is that corpora following a similar sampling frame at a certain interval can serve as good resources for describing language change. That is to say, Crown and CLOB can be viewed as good reference corpora for contrastive studies in terms of historical change (e.g. comparison among Brown, Frown, AmE06 and Crown) and regional variation (e.g. Brown vs LOB, Frown vs FLOB, BE06 vs AmE06, and Crown vs CLOB). Moreover, a bigger balanced corpus of contemporary written English can be created when the Brown corpora are merged.

The collection of English texts aimed at a balanced, in a modest sense, corpus of written contemporary English modelled after the sampling frame of the Brown Corpus. The name *Crown* is the fusion of the initial of *China* and the hind part of *Brown*, which means a new Brown family corpus collaboratively built by Chinese scholars. Likewise, *CLOB* gets its name with an initial *C*. We closely followed Kučera and Francis' sampling frame of the Brown Corpus.

The first standard release of Crown and CLOB was in June 2012. The corpora are available in raw texts (with and without metadata), PoS-tagged format (with CLAWS C7 tagset), and parsed format (with BFSU Stanford Parser). An online query interface of the corpora is also available for public access at <http://124.193.83.252/cqp/>.

2 Descriptive statistics of Crown and CLOB

2.1 The size of Crown and CLOB

The total numbers of tokens as well as the tokens of four major genres of Crown and CLOB are provided in Table 1, together with those of Brown, LOB, Frown, and FLOB. (Note: four broad genres refer to newspaper texts (A–C, 88 texts), miscellaneous informative prose or general prose (D–H, 206 texts), learned and scientific English (J, 80 texts), and fiction (K–R, 126 texts).¹

Table 1: The number of tokens of three generations of Brown family corpora²

	Genre	Sub-corpus tokens	Total tokens		Sub-corpus tokens	Total tokens
Brown 1961	Fiction	259,467	1,027,021	LOB 1961	258,722	1,018,785
	General prose	423,160			418,137	
	Learned	163,309			162,322	
	Press	181,085			179,604	
Frown 1992	Fiction	260,414	1,027,323	FLOB 1991	260,664	1,024,643
	General prose	421,933			419,990	
	Learned	163,228			163,286	
	Press	181,748			180,703	
Crown 2009	Fiction	259,250	1,026,226	CLOB 2009	259,484	1,023,466
	General prose	422,799			421,163	
	Learned	163,197			163,139	
	Press	180,980			179,680	

2.2 Years of publication of the texts in Crown and CLOB

The years of publication of the texts included in Crown and CLOB are summarised in Figures 1 and 2:

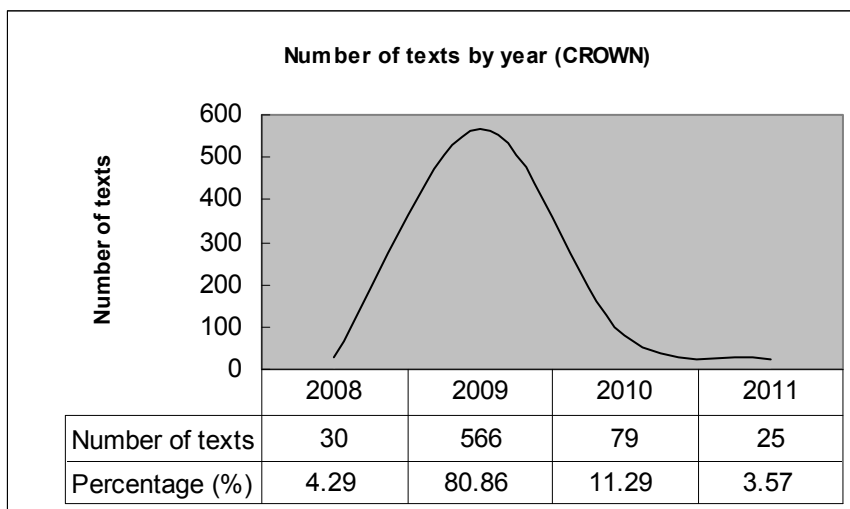


Figure 1: Number of texts by year (Crown)

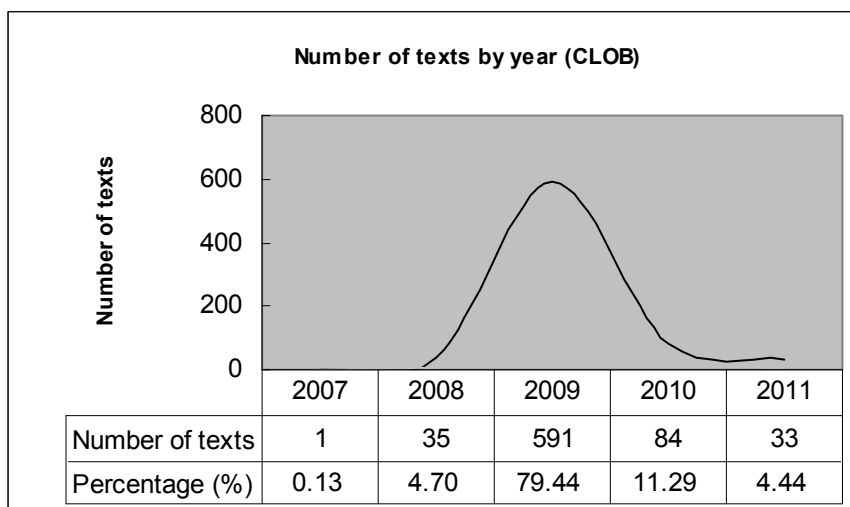


Figure 2: Number of texts by year (CLOB)

As seen in Figures 1 and 2, about 80 percent of Crown and CLOB texts were published in the year 2009, and 15 percent of the texts were published one year before or after 2009. Taken together, texts published between 2008 and 2010 account for over 95 percent of the total.

3 Data collection and processing

3.1 Sampling frame: The Brown Corpus model

The first standard release of Crown and CLOB contains one million words respectively, covering 15 categories (see Table 2) of texts published in 2009, or one year before or after 2009. In a few (less than 5%) cases, texts published in 2007 and 2011 were included.

Table 2: Text categories of Brown Corpus

	Text categories	No. of texts
A	Press: Reportage	44
B	Press: Editorial	27
C	Press: Reviews	17
D	Religion	17
E	Skill and hobbies	36
F	Popular lore	48
G	Belles-lettres	75
H	Miscellaneous: Government & house organs	30
J	Learned	80
K	Fiction: General	29
L	Fiction: Mystery	24
M	Fiction: Science	6
N	Fiction: Adventure	29
P	Fiction: Romance	29
R	Humour	9

3.2 The nationality of authors

Only the writings by the U.S. and the U.K. citizens or permanent residents were selected. The URLs, or the publishers, and the author profile pages were

recorded. In the case of multiple authors, the primary/first author should hold U.S. or U.K. citizenship or permanent residence.

3.3 *The length of texts*

The length of each text should be no shorter than 2,000 words. The word/token definition we used is [a-zA-Z0-9-]+. For some text types, one single article is not likely to be as long as 2,000 words. In that case, two or more articles of the same nature and on similar topics were gathered. If a text sample was going to be taken from a very long text, say, a novel, approximately 700 words from the beginning, the middle and the ending of the book respectively were selected.

3.4 *Text cleaning and processing*

The texts were cleaned and processed with regard to four aspects:

(1) To avoid, as much as we possibly could, mojibake or gibberish codes while processing the texts with corpus tools, some punctuation marks, mathematical symbols, non-English scripts (Greek, Nordic, German letters, etc.), and other special symbols were replaced with ASCII printable characters. For mathematical symbols, technical symbols, and Greek letters, the conversion table of HTML Symbol Entities Reference was consulted,³ and replaced with ASCII printable characters accordingly. For instance, α , β and Υ (in scientific discourse) were changed into ‘α’, ‘β’, and ‘γ’ respectively. All em dashes (—) and en dashes (–)⁴ were replaced by two hyphens (--) with a white space before and after the two hyphens. ö, æ, ë, ê, and ß (in German) in proper names were replaced by o, ae, e, e, and ss.

(2) The biographical information of all the authors was checked by the corpus linguistics team members at Beijing Foreign Studies University. Wikipedia and authors’ personal and/or institutional web pages were searched through to make sure that the authors were born and got their education in the United States or the United Kingdom.

(3) Advertisements, tables, figures, sideline links etc. were deleted.

(4) Partial or whole text duplicates were checked with Wcopyfind.exe.⁵ Duplicates found were removed and new texts were added in place. In such cases, we deliberately looked for some texts which were published in 2011; thus duplicate texts would be less likely.

3.5 *Format of texts and file names*

Plain text files (in ASCII encoding) are named A01A.txt, A01B.txt, A02A.txt, A02B.txt, B01A.txt, B01B.txt, etc. (the fourth character A for American, and B for British). There are 700 texts in Crown corpus and 744 texts in CLOB corpus

instead of 500 texts of 2000 words each in the Brown corpus model, because short texts are saved separately in Crown and CLOB. For example, A03 text in the Brown model consists of four short texts in Crown corpus and they are named A03AA.txt (449 words), A03AB.txt (449 words), A03AC.txt (737 words) and A03AD.txt (433 words). Short texts are especially frequent in news reports. The strength of saving the short texts separately is that each text represents itself. It is quite easy to merge the related texts as a 2,000-word one, if this is necessary. However, once the short texts have been put together in one single 2,000-word file, without explicit section markers, it is almost impossible to save them as individual texts as they originally were.

3.6 No reprinted works

The work had to be first published in 2009 (± 1 year). Works reprinted in 2009 (in other words, they were originally published earlier than 2009) were not included. For instance, reprinted detective novels of *Sherlock Holmes* were not considered.

3.7 Metadata mark-up

Metadata were marked up as follows. Metadata Encoder 2⁶ was used to add the following bibliographic information (see Table 3) to the texts:

Table 3: Metadata template for Crown and CLOB

Author	e.g. John Smith
Country	e.g. UK/US
Publication year	e.g. 2009
Publisher	e.g. New York Times
URL(s)	e.g. http://

4 Copyright and dissemination

The copyright of all the texts in Crown and CLOB belongs to the original copyright holders. We plan on a GNU⁷ distribution of the texts, given that they are not used for commercial purposes in whatever manner. Everyone who makes a positive contribution to the data collection will get a copy of the Crown and CLOB corpora for their personal study. Other users can access Crown and CLOB at our CQPweb site: <http://124.193.83.252/cqp/> (ID: test; password: test).

5 Case studies based on Crown and CLOB

The two new members of the Brown family, along with other Brown family members, can be a good dataset for studies on English language, for instance, for the contrastive study on lexis, phrases, grammar patterns, stylistic features favoured by the two varieties of English, synchronically and diachronically. The corpus research group at Beijing Foreign Studies University has investigated the lexical, grammatical, and stylistic aspects of the enlarged Brown family corpora.

Chen (2012) compares the 's-genitives' in the American and British press reportage. She concludes that: 1) American English and British English share similar 's-genitive' developmental patterns, i.e. the use of 's-genitive' increases from 1961 to 1991, and becomes stable after 1991; 2) Possessor animacy counts as one of the most powerful factors around the growing and declining popularity of 's-genitive' in contemporary modern English.

Ji (2012) conducted a quantitative and qualitative study on the variation of modals and semi-modals between American English and British English, and also along the timeline over nearly half a century from 1961–2009. Both tokens and modal meanings are examined. Extraction and comparison of modal occurrences are done with concordancers; modal meanings, such as obligation, permission, necessity, volition, possibility, however, are manually annotated. Data from the structural and semantic analyses show that possibility comes as the predominant meaning on top of all the other modal meanings, and it is used more in British English than American English, more in fiction than other genres, and sees a stable increase over half a century.

Zhang and Xu (2013) explore the diachronic changes of the distribution patterns and semantic shift of the English present progressive. The results show that, overall, frequencies of the English present progressive increase significantly since the 1960s, and the growth is more conspicuous at the beginning of the 21st century; as to genre and voice, the progressive instances exhibit some variability in addition to the general growing tendency; the frequencies of the non-present progressive uses (the futurate and attitudinal meanings) of the present progressive also increase dramatically. All these findings imply that the English present progressive in written English has a distinct tendency to be colloquialised, and in the meantime, the increasing uses of semantically shifted 'be V-ing' patterns seem to indicate that the usage of the present progressive is becoming more diversified and subjectified, suggesting that it is in a process of grammaticalisation, which is in concordance with the overall evolving tendency of English grammar.

Lu (2012) makes an in-depth corpus-based exploration of discourse representation (Leech and Short 1981, 2007; Semino and Short 2004), or speech, writing and thought presentation (SW&TP) in English romantic fiction based on six Brown corpora (Brown, Frown, Crown, LOB, FLOB, and CLOB). Statistical results show that both American and British English prefer to employ thought representation, and that there is more ‘telling’ than ‘showing’ in American romantic fiction than in British romantic fiction. A general trend of increasing direct thought, free direct thought and free indirect thought is observed over time, which brings about more dramatic effect and vividness.

6 Concluding remark

The two Brown family corpora have been made publicly available among Chinese researchers for some time, which has encouraged quite a number of comparative studies of English lexico-grammar. Apart from the studies of English per se, more often than not, Crown and CLOB, alongside other Brown corpora, serve as reference corpora of present-day native English in contrastive interlanguage analysis, and corpus resources for Chinese-English contrastive studies. We are now finalising our 2009 Chinese Brown family corpus (temporarily called CC2009) developed at Beijing Foreign Studies University. We hope that the corpus building work is but the beginning of more Brown family corpus-related corpus construction and research in the far east.

Acknowledgements

Our greatest debt is to all the contributors who helped collect texts for the corpus project. A special thank then goes to Jie Ji for coordinating the BFSU Corpus Research Group to check the text format, encoding, relevance of text content to their genre categories, and sources of the texts with extreme care, and all other kinds of proofreading work. We would also like to thank Gong Chen, Jie Ji, Guobing Liu, Lei Liu, Xia Liu, Gong Peng, and Baicheng Zhang for their valuable comments on the earlier drafts of the article. The compilation of Crown and CLOB is supported by the National Social Sciences Foundation Project “A Bilingual Corpus Based Study on the English Translation of Chinese Complex Verbal Constructions” (Ref.: 12CYY060) and “Program for New Century Excellent Talents in University” (Ref.: NCET-12-0790). We are also grateful to the participants at the *Corpus Linguistics in China 2011* conference, as well as the audience at the *Corpus Technologies and Applied Linguistics* (2012) conference, for their critical comments.

Notes

1. http://icame.uib.no/archives/No_5_ICAME_News_index.pdf.
2. The token definition applied was all alphanumeric character strings plus hyphens, and the regular expression for the definition is [a-zA-Z0-9-]+.
3. http://www.w3schools.com/TAGS/ref_symbols.asp.
4. <http://en.wikipedia.org/wiki/Dash>.
5. <http://plagiarism.bloomfieldmedia.com/z-wordpress/software/wcopyfind/>.
6. <http://ishare.iask.sina.com.cn/f/33571440.html>.
7. Refer to more information about the GNU General Public License at <http://www.gnu.org/licenses/gpl.html>.

References

- Chen, Gong. 2012. A diachronic study on English 's-genitives': Evidences from Brown family corpora (1961–2009). Paper presented at *Corpus Technologies and Applied Linguistics* conference. 29 June 2012, Xi'an Jiao Tong Liverpool University, Suzhou, China.
- Ji, Jie. 2012. Variational study on modal meaning – a corpus-based research of Brown family. Paper presented at *Corpus Technologies and Applied Linguistics* conference. 29 June 2012, Xi'an Jiao Tong Liverpool University, Suzhou, China.
- Leech, Geoffrey and Mick Short. 1981. *Style in fiction*. London: Longman.
- Leech, Geoffrey and Mick Short. 2007. *Style in fiction*. 2nd edition. London: Pearson Education.
- Lu, Lu. 2012. The speech, writing and thought presentation in English romantic fiction: A diachronic and cross-cultural study. Paper presented at *Corpus Technologies and Applied Linguistics* conference. 29 June 2012, Xi'an Jiao Tong Liverpool University, Suzhou, China.
- Semino, Elena and Mick Short. 2004. *Corpus stylistics: Speech, writing and thought presentation in a corpus of English writing*. London: Routledge.
- Xu, Jiajin. 2011. A tale of two C's: Crown and CLOB. Paper presented at the *Corpus Linguistics in China* Symposium. 19 November 2011, Beijing Foreign Studies University, Beijing, China.
- Xu, Jiajin. 2012. Comparing English varieties with Crown and CLOB. Paper presented at *Corpus Technologies and Applied Linguistics* conference. 29 June 2012, Xi'an Jiao Tong Liverpool University, Suzhou, China.
- Zhang, Baicheng and Jiajin Xu. 2013. A Brown family corpora based diachronic study on the English present progressive. *Foreign Language Education* 34: 42–47.

ToRCH 2009 平衡汉语语料库建成

近日，由我中心许家金教授主持，全国 60 多所高校的 115 位教师及研究生共同创建的现代汉语语料库 ToRCH 2009，正式对外发布供国内语言研究者免费使用。该语料库大小为 1,087,619 词, 1,703,635 字。该语料库所收文本绝大部分为 2009 年出版。语料库的名称 ToRCH 为 Texts Of Recent CHinese 的缩略词。该语料库可用于现代汉语研究，也可与此前中国外语教育研究中心组织创建的 Crown 和 CLOB 语料库¹构成英汉可比语料库（comparable corpora），开展汉英对比研究。

该语料库的一个突出特点是“共建共享”。它由上百位老师和同学共同完成，并与语料库的创建者、广大语料库研究者共享的一个新型语料库。

ToRCH 2009 及 Crown、CLOB 语料库皆可通过 BFSU CQPweb 语料库平台（<http://124.193.83.252/cqp/>，账号密码皆为 test）在线检索。ToRCH 2009 语料库的所有文本还可从 <http://ishare.iask.sina.com.cn/f/37898721.html> 下载。

¹ 有关 Crown 和 CLOB 语料库的详细介绍，可参看 http://icame.uib.no/ij37/Pages_175-184.pdf。

English Abstracts of the Full-text Articles in *Corpus Linguistics Research Highlights*

The one-*de*-preference of Chinese nominal phrases: Converging evidences from the corpus and cognitive linguistics

Appeared in *Foreign Language Education*

ZHANG Lifei and YAN Chensong (PLA University of Foreign Languages)

Abstract: A corpus-based study indicates that Chinese nominal phrases with multiple determiners prefer only one determiner-marker *de*. The supporting evidences include: 1) In actual language use, the majority of nominal phrases carry only one *de* in places where more than one *des* are possible; 2) The adjective *zhongyao* is more likely to omit *de* in nominal phrases with multiple connotative determiners than in those with only one connotative determiner. The cognitive process behind Chinese nominal phrase's one-*de*-preference is the activation of highly entrenched schema D *de* H to categorize nominal phrases with multiple connotative determiners. The omission or presence of *de* in nominal phrases is determined by two competing motivations, i.e. the Principle of Iconicity and the Principle of Economy. The former stresses the one-to-one relationship between form and meaning and tends to allocate one *de* for every connotative determiner, while the latter seeks to achieve the communication goal with the least cognitive effort and thus hopes to keep only *de* for every nominal phrase. The result of the competition is a compromise of the sort, i.e. using as few *des* as possible while conveying the intended meaning.

Typological generalizations and Chinese relative clauses produced by second language learner

Appeared in *Foreign Language Teaching and Research*

LI Jinman and WU Fuyun (Shanghai University of Finance and Economic and Shanghai International Studies University)

Abstract: Chinese has the VO word order, as English does, but its relative clause (RC) is prenominal, which is typologically rare. Thus Chinese RC might be a difficult structure to acquire for learners whose first language is English. Based on a corpus analysis of Chinese RCs produced by native English speakers, this study investigates the distribution of different types of Chinese RCs and the accountability of three typological generalizations for the patterns found. The corpus data show that the distributional patterns of Chinese RCs produced by second language learners conform to varying degrees to the three typological formulations derived from natural language observations. Moreover, typological generalizations vary in their strength and interact with each other, with animacy being ranked the strongest, accessibility the second, and embeddedness the weakest. Furthermore, animacy accounts for the partial disagreement between the patterns found in our corpus and those predicted by the latter two typological assumptions.

Demonstrative-classifier positioning in Chinese relative clauses and its underlying reasons: A

spoken corpus study

Appeared in *Modern Foreign Languages*

SHENG Yanan and WU Fuyun (Shanghai International Studies University)

Abstract: This paper discusses the distribution pattern of demonstrative classifiers (DCLs) in relative clauses (RCs) extracted from a Chinese live TV-show -- *Appointment with Luyu*. We found an asymmetric distribution pattern: pre-RC classifiers tend to occur in subject-extracted relative clauses, and post-RC classifiers in object-extracted relative clauses. This corroborates the findings of previous studies based on written corpora. The results are consistent with the audience-design model, suggesting that speakers are altruistic enough to provide early cues (i.e. the DCL) for the RC structure, and to avoid potential lexical disruptions or ambiguity induced by the mismatched or matched classifiers to ease comprehenders' parsing difficulty.

Patterns and functions of textual sentence stems in research articles

Appeared in *Journal of PLA University of Foreign Languages*

ZHANG Le and WEI Naixing (Shanghai Jiao Tong University and Beihang University)

Abstract: Textual sentence stems (TSSs) are semi-fixed sentential sequences that occur frequently in English academic texts and perform the function of organizing the text and expressing attitudes. Despite considerable interests in the way of facilitating the reader's understanding of propositional information and directing the reader to the writer's position, we still lack a comprehensive empirical study of how clausal expressions are used for interpersonal purposes. Based on an analysis of data in JDEST and with a corpus-driven approach to phraseology, this paper analyzes patterns and functions of TSSs. It is suggested that TSSs are important realizations of attitudinal meanings in research articles and perform a wide range of interpersonal acts. They are largely single choices in language use, involving the co-selection between lexis, pattern, rhetorical structure, register and so on. These findings may have practical implications for the teaching and learning of academic writing.

Undergraduates' acquisition of English present perfect tense: A corpus-based study – New evidences for the Saliency and Form-acquisition-priority Hypotheses

Appeared in *Foreign Languages and Their Teaching*

ZHANG Baicheng (Chongqing Jiaotong University)

Abstract: L2 present perfect tense acquisition is notoriously difficult, and the research on this issue has not got consistent findings. Based upon the four micro-corpora extracted from SWECCCL, CLEC and COLSEC, the present study explores the freshmen and sophomores' formal and semantic acquisition of English present perfect tense. Research findings support the Saliency Hypothesis but contradict the Form-acquisition-priority Hypothesis. Therefore, the study tentatively proposes that L2 tense and aspect acquisition is systematic and dynamic in foreign language contexts, and there is not a stable/unchangeable order. The acquisition order and proficiency of L2 tense and aspect is dynamically influenced, shaped and even decided by the

interaction of multidimensional factors including classroom teaching, the difficulty of certain L2 tense and aspect items, learner differences, learning stages, L1-L2 differences and some others.

On the development of psycholinguistic word information in Chinese EFL learners' written discourse: A study based on WECCL and MRC database

Appeared in *Modern Foreign Languages*

LI Hai and ZHANG Shujing (PLA University of Foreign Languages)

Abstract: Using the British Medical Research Center Psycholinguistic Database, this study aims to examine the developments of four types of psycholinguistic information, namely concreteness, 219 imagery, meaningfulness and familiarity, in L2 vocabulary produced by Chinese EFL learners in WECCL. The following patterns were found from freshmen to seniors: 1) concreteness and imagery of L2 vocabulary underwent a decline; 2) meaningfulness also proved to be on the decline, but not in a significant manner; and 3) familiarity of words declined first, but bounced back from sophomores onwards. The study indicates and reaffirms the feasibility of MRC database in L2 vocabulary research, and also introduces a different perspective from which research on L2 vocabulary can be carried out anew.

Translator's image schema and integrated conceptualization: A corpus-based study of "social face" in *Hong Lou Meng* and its translation

Appeared in *Foreign Languages and Their Teaching*

TAN Yesheng (Shanghai International Studies University)

Abstract: Via a corpus-based analysis of how "social face" is presented and conceptualized in two English versions of *Hong Lou Meng*, the paper intends to explore the connections between the creative cognitive processes of translation and the underlying image schema and embodied experiences. The analysis reveals that in the process of translation the source-domain image schema of "social face" can be instantiated or extended along various dimensions based on embodied experiences, involving different preferred modes of metaphorical conceptualization of social relations in the two lingua-cultures. And on the other hand, given the common underlying image-schema and the commensurable embodied experiences, language-specific modes of conceptualization can be integrated, and thus produce new ways of conceptualization and new linguistic expressions in the target language. The tentative conceptual framework of "schema-instance" hierarchy, the non-sentence-level parallel analysis and multi-corpus analysis adopted in the paper may shed light on the research of cognitive translology and contrastive cognitive linguistics in future.

The design and development of the R-gram based corpus analysis tool 'PowerConc'

Appeared in *Computer-Assisted Foreign Language Education*

XU Jiajin and JIA Yunlong (Beijing Foreign Studies University)

Abstract: This paper describes the innovative corpus tool PowerConc developed by the authors themselves. In its implementation, such key functionalities of corpus tools as concordancing, wordlist generation and keyword analysis, were redesigned, expanded and optimized. What underlies the whole design of PowerConc is the inventive synergy of regular expressions and N-gram, namely, R-gram proposed in this study. The R-gram feature allows for flexibility in concordancing, exhaustive listing of linguistic units, and key terms of varying length, and more likely than not, enables analyses of linguistic structures with uncertain words or categories. To minimize the inconvenience of operation, Smart Input has been introduced to facilitate easy search with enhanced returned hits. The user-friendly PowerConc is object-oriented software. The key functions have been packaged in different classes as a dynamic link library (*.dll) file, independent from the user interface, which warrants easy maintenance and expandability. It is hoped that PowerConc will be conducive to corpus-based research in its own way.

北京外国语大学语料库语言学团队介绍

语料库语言学是教育部人文社科重点研究基地——北京外国语大学中国外语教育研究中心的主要研究方向之一。经过数年的积累和发展，该中心的语料库语言学研究团队（以下简称“北外语料库团队”）已成为国内语料库语言学研究的主要力量之一，目前拥有专职研究员 5 人（王克非教授、梁茂成教授、李文中教授、许家金教授、熊文新副教授），兼职研究员 1 人，全部拥有博士学位，专业领域涉及基于语料库的理论语言学、对比语言学、翻译学、应用语言学、计算语言学等方面的研究，是一个典型的跨学科研究团队。该方向每年还招收一定数量的博士后、博士、硕士研究生和访问学者。

北外语料库团队的突出特点是：1）学术思想开放。不泥于一家之说，兼容并包不同语料库研究理念。鼓励语料库同其他相关领域的跨学科发展；2）建库和技术开发为先导，重视语料库语言学相关理论的探索；3）团队人员构成合理。顶层设计由王克非和文秋芳两位教授领衔，中间层以梁茂成、李文中、许家金、熊文新为骨干，加上一批极具活力的青年学者和学生积极参与。

北外语料库团队近年来积累一系列成果。目前，中国外语教育研究中心的语料库语言学研究主要集中在双语语料库和学习者语料库两个方向，近年来完成或在研的省部级以上研究项目主要有：国家社科基金重大项目 1 项（由王克非教授主持）、国家社科基金一般项目 5 项、教育部人文社科重点研究基地重大项目 3 项、教育部人文社科一般项目 3 项，发表 SSCI、CSSCI、EI 级别的相关学术论文百余篇、专著十余部。

北外语料库团队的相关成果在全国产生了重要影响。由王克非教授主持建成的北外英汉——汉英通用平行语料库是迄今为止国际上最大的英汉、汉英平行语料库。2010 年立项的国家社科基金重大研究项目“大规模英汉平行语料库的建立与加工”更将推出国际上前所未有的超大规模英汉、汉英平行语料库，研究团队中集结了国内外英汉、汉英双语语料库建设和翻译研究中的大批研究者。

学习者语料库建设与研究是北外语料库团队相关研究的又一亮点。由文秋芳教授主持建立的“中国学生英语口语语料库”和“中国学生口笔译语料库”已成为中国学习者语料库相关研究的重要资源。基于学习者语料库分析需求而设计的 PowerConc、Collocator、Colligator、ParaConc、PatCount、PatternBuilder、Parallel Text Aligner、Readability Analyzer、TreeTagger for Windows、BFSU Stanford Parser 等已成为国内语料库语言学研究耳熟能详的分析工具。近期开发的 Crown、CLOB、ToRCH 2009 语料库项目、PowerConc 软件包项目更将有力推动我国的语料库语言学研究；以中心学者为主编写的《语料库应用教程》已成为国内语料库语言学的一本重要教科书。

中心组织的语料库语言学沙龙、语料库语言学研讨会、语料库在外语教学与研究中的应用研修班、Corpus4U 语料库在线论坛（合办）等汇聚国内外大批语料库研究者和爱好者，为国内语料库语言学的普及和发展构建了多层次的学术交流平台。

衷心希望北外语料库语言学团队的研究得到广大语料库爱好者的支持。

Corpus Linguistics Research Highlights
of
The Linguistics and Applied Linguistics Research Highlights Series
Bi-annual

Managed by the National Research Centre for Foreign Language Education, Beijing
Foreign Studies University

Edited at the National Research Centre for Foreign Language Education, Beijing
Foreign Studies University

Editor: Maocheng Liang
Associate editors: Wenzhong Li, Jiajin Xu and Wenxin Xiong
Executive editor: Jiajin Xu
Editorial assistant: Jie Ji
Cover designer: Jie Ji

Advisory board

Wenzhong Li
Maocheng Liang
Kefei Wang
Qiufang Wen
Wenxin Xiong
Jiajin Xu