

# 语料库视角的英语口语“立标语块”探究<sup>\*</sup>

华南师范大学 何安平

**提 要:** 本文研究说话者表达对命题态度和立场的“立标语块”。笔者采用 Sinclair (2004) 的词汇语法分析手段对以往基于语料库的机切四词语块研究成果作二次加工, 构建起立标语块典型的形式、语义和语用共选范式; 然后带入两个同质同源的中、英大学生英语口语微型语料库作对比分析。研究发现这两类学生的立标语块使用不但在频数和种类上, 而且在语义和语用倾向上均有明显差异, 由此引出对国内英语教学的反思。

**关键词:** 语料库, 立标语块, 英语口语, 英语教学

## 1. 引言

说话者在表述话语命题信息的同时也表述自己对这些信息的立场态度, 这是构成话语交际能力的两个重要成分。本文探讨的是后一种能力, 即立场标记话语(stance, 以下简称立标语)的使用能力。从母语习得的角度看, 后一种能力的习得要比前一种晚(Kuczaj & Daly, 1979: 572; Perkins, 1983; Coates, 1988); 从外语教学的角度看, 这两种语言能力的基本表达形式通常在初级阶段都已经呈现, 但研究发现学习者对后一种能力的掌握却远不如前者(Cook 1978; DeCarrico 1986: 665; 桂诗春, 2009: 120)。针对这个问题, 本文尝试用语料库语言学的理念和研究手段来探讨英语口语交际中的立标语特点。切入点是英语口语语料库的机切四词语块。研究思路是先对以往基于英语本族语者语料库的同类研究成果进行二次加工, 以提取英语口语立标语块的典型表达范式及语义语用功能; 然后将此范式带入其他大型英语本族语者口语语料库加以验证和扩充; 最后利用扩充后的典型范式对比中、英大学生两个同质同源的微型口语语料库的立标语块, 由此凸现中国英语学习者的口语立标语使用特征, 并对英语教学做出反思。

## 2. 立标语块和机切四词语块

从语言序列的角度研究立标语是近年国内研究中国英语学习者立标语用能力的新视角。其中有梁茂成(2008)研究立标语的一种重要表现形式, 即情态动词的语法序列, 结果发现英语情态动词的语义分类(如认知情态类和义务情态类)与其

语法搭配范式高度相关。还有徐曼菲(2009)研究学习者“过度使用”的 I think 这一典型立标语的周边词汇、语法和语义共选, 结果发现它们不仅有渐强或渐弱的线性序列, 还有强弱转化的动态特征。本文则从语块角度来研究口语立标语, 这是基于语言的基本意义单位主要是词项或短语的这一语料库语言学理念, 即“语言意义的表达通常是一次性地选择两个以上的词……语言有很多意义都得靠一个以上的词组织起来才能实现”(Sinclair, 2004: 133)。本文使用机切四词语块作为研究立标语的基本单位则是受到国外学者(Biber, Conrad & Cortes, 2004 和 Biber & Barbieri, 2007)研究的启发。他们为了探究不同语域的语篇特征, 在逾 1200 万词次的课堂会话、日常会话、大学教材语料中用计算机软件自动切出了所有四词序列(lexical bundle, 以下称四词语块), 结果发现其中的立标语块(stance bundle)是口语语篇的重要构件。它用于表达说话者对所说命题的态度或者对所讲信息作肯定程度高低的评估, 常常位于命题(语句)的起始部分, 也时而出现在句中或句尾。形式上既有小句结构也有短语结构。语义上则可分为五小类: 认知类(如 I don't know what)、愿望类(如 I don't want to)、义务类(如 you have to do)、预测计划类(如 what we're going to)和能力类(如 to be able to)(Biber, Conrad & Cortes, 2004: 384-386; Biber & Barbieri, 2007: 8-9)。也有学者(Du Bois, 2007: 163)归纳出立标语的三大功能, 即“评价命题信息(evaluates an object)、表述自身立场(positions a subject, usually

<sup>\*</sup> 本文是“学术话语中的立场标记语研究”项目成果, 得到广东省哲学社会科学“十一五”规划基金(编号: 06GK-01)资助和匿名审稿专家的宝贵意见和指正, 谨此致谢。

the self) 和与对方互动 (aligns with other subjects)”。

研究语言序列从机切语块入手的最大优势是它能将大批量语言信息高度浓缩、简化从而突显其语篇意义 (Scott & Tribble, 2006: 5)。具体表现为三大特点:

1) 对语料处理的彻底性。它能够穷尽所有语料以便于更全面的观察与判断。例如对 I think it is a funny story 这句话语作四词语块的自动切分,就能一字不落地切出 I think it is, think it is a, it is a funny, is a funny story 四个语块。

2) 具有高频性和客观性。机切语块能够呈现很多不能靠说话者自身回忆和本能意识而得到的高度复现的词块。例如光凭感觉人们很难想象从一千万词次的剑桥国际语料库 (CIC) 机切出来的最高频三词语块是 I don't know (588 次/100 万字); 从 500 万词次的英国英语口语语料库 (CANCODE) 机切出来的最高频二词语块是 you know, I mean 和 I think (28~ 40 次/万字); 其中某些语块的频数甚至比某些常用单词还要高 (如 you know 就比 really 高频)。这些“反复出现的语言现象必然会引起注意,也就很有必要对其进行研究和解释” (Biber, Conrad & Cortes, 2004: 400)。

3) 具有预测性。表面上看多数机切语块的结

构和语义都不完整,然而仔细区别却可分出其中有一类是重复率很高但却无意义,如 are to my 和 this one for; 另一类是句子结构不全但有意义,如 a lot of 和 to be able to; 还有一类具有明显的语义和语用完整性,如 as a result 和 on the other hand (参考 O' Keeffe, McCarthy & Carter, 2007: 61-66)。

另外,口语交际的即时性特征使其话语结构单位较为短小; 频繁使用的短小“熟语序列”有助于减轻话语在线产出的压力,这就造成许多高频复现的语块能使人听到前面就能预知后面要说什么 (如 well I think it 往往预设后面会是以 it 引起的小句结构)。而本研究之所以采用四词语块而不是三词语块或其他,还因为会话中的四词语块绝大多数连接着两个结构单位:“语块开头通常呈现小句或短语的边界,并且‘勾起’了下一个结构单位”,起着“话语锚(anchor)”的提示作用或引领功能 (Biber, Conrad & Cortes, 2004: 377, 399), 这正是立标语在话语中出现的典型位置。可见,从机切四词语块入手更能揭示立标语典型的形式、意义与功能特征。

3. 研究语料及方法步骤

本研究的语料来自多个大型英语口语语料库的机切四词语表。详见表 1。

编号	名称	来源	频率
1	Biber 立标语表 <sup>①</sup> (共 36 种)	LSWEC <sup>②</sup> 中 640 万词次的英美会话语料	10~ 100 次/100 万词
2	Scott 最高频四词语表(40 种)	BNC-CONV 中 100 万词次的会话语料	(不详)
3	O' Keeffe 最高频四词语表 (共 20 种)	CANCODE 中 500 万口语语料	59~ 136 次/100 万词
4	NS 四词语表(共 3598 种)	LINDSEI-bri 中 11.7 万词次的英语本族语者大学生会话语料	2~ 37 次/10 万词
5	NNS 四词语表 (共 2009 种)	LINDSEI-chi 中 6.35 万词次的中国英语专业大学生会话语料	3~ 41 次/10 万词

表 1 本研究使用的语料

① 该表编号 1 的 Biber 立标语表(简称,下同)摘自 Biber, Corad & Cortes(2004: 384-386); 编号 2 的 Scott 高频四词语表摘自 Scott & Tribble (2006: 140); 编号 3 的 O' Keeffe 高频四词语表摘自 O' Keeffe, McCarthy & Carter(2007: 66); 编号为 4 和 5 的 NS 和 NNS 四词语表均从华南师大外文学院的英语教育教学语料库提取(详见何安平, 2008: 340)。

② 该列中 LSWEC 全称 Longman Spoken and Written English Corpus, <http://www.ebookee.net/search.php?cx=005418540955315608444%3Ahzcbi9hnswe&cof=FORID%3A11&q=longman+spoken+and+written+corpus&sa=Search#1533> (October-16-2010);

BNC-CONV 是 British National Corpus 中的会话部分, <http://www.natcorp.ox.ac.uk/what/index.html> (October-16-2010);

CANCODE 全称 Cambridge and Nottingham Corpus of Discourse in English, [http://www.cambridge.org/cn/elt/catalogue/subject/custom/item3646595/Cambridge-International-Corpus-Cambridge-and-Nottingham-Corpus-of-Discourse-in-English-\(CANCODE\)/?site-locale=zh-CN](http://www.cambridge.org/cn/elt/catalogue/subject/custom/item3646595/Cambridge-International-Corpus-Cambridge-and-Nottingham-Corpus-of-Discourse-in-English-(CANCODE)/?site-locale=zh-CN) (October-16-2010);

LINDSEI 全称 Louvain International Database of Spoken English Interlanguage, <http://www.uclouvain.be/en-cccl/lindsei.html> (October-16-2010).

本次研究采取的策略是以大型语料库的倾向性或典型性研究成果来引领对小型语料库更详尽的分析。即: 先从英语本族语研究者已认定的一批四词口语立标语块中归纳典型范式; 接着带到其他英语本族语者大型口语语料库的高频四词语表中进行该范式的验证和扩充; 然后再带到两个同质同源微型语料库作更大范围的四词语块表对比分析。

笔者首先收集了 Biber 等人在 6.4 百万词次的日常会话语料中认定的平均每百万词次出现 10~ 100 次以上的所有四词立标语块( stance expressions), 一共 36 种( 见表 1 和表 3), 约占该研究位于同样频数段的 52 种会话四词语块的 70%。基于机切语块是语篇功能建构单位的这一理念, 笔者采用 Sinclair( 2004) 对语言单位意义的四层构建法<sup>①</sup>来观察和归纳这 36 种语块的词汇、语法和语义搭配特征, 从而构建口语立标语块的典型范式。

类别	代词		动词		情态动词		其他	
词型: 频数	I( 含 I' m, me):	17	know: 9	going(to):	10	to:	23	
	you:	14	w ant: 9	have(to):	7	do/ don' t/ not:	19	
	it( 含 it' s):	3	think: 3	able(to):	1	be/ w as/ are:	10	
	w e( 含 we' re):	2				what/ how/ whether/ why/ if:	8	
						a/ and:	5	
						well/ oh/ go/ so:	4	
合计: 比例	36:	25%	21:	15%	18:	12%	69:	48%

表 2 Biber 立标语块表的词性分类

表 2 显示代词类占总数 1/ 4, 其中约 90% 为第一和第二人称代词( I, you 和 we); 其次是心理类动词和情态词, 共占了 1/ 4 强; 余下的动词不定式 to、助动词和疑问副词等则频繁用于与上述词类构成反复出现的语块, 如 I don' t know, do you want, going to 等等( 见表 3)。前 3 类词的词次总和( 36+ 21+ 18= 75) 超过了该语块表总词量的一半, 表明口语立标语块的用词非常集中, 且高频复现。其中最高频的 I 和 you 体现了交流双方频繁的互动; 与之高频搭配的心理类和情态类动词则表现出面对面交流时必要的表态; 凸显了口语交际的三大特点: “互动性、表态性和实时性”

第二步是将 Biber 立标语表带到另外两个英语本族语者大型口语语料库的最高频四词语表( 见表 1 编号 2 和 3) 中验证其代表性和典型性。一是看两者有多少在词汇形式上完全一致的语块; 二是看除此之外后者还有多少词汇上不完全匹配但结构和语义却可归属到该典型范式的其他语块, 以此构成扩充性的典型立标语表。

第三步是将扩充后的立标语表及其典型范式带入中、英大学生两个同质同源<sup>②</sup> 的英语口语微型语料库( 见表 1 编号 4 和 5) 的所有频数为两次及以上<sup>③</sup> 的四词语块表做对比分析, 以揭示中国英语学习者使用立标语块的状况和特征。

4. 结果与讨论

4. 1 口语立标语块的典型范式

对表 1 编号 1 的立标语表作词频分析, 发现这 36 个语块( 共 144 词次) 总共只由 32 种词型构成, 词性分类后的结果见表 2。

(Biber, Conrad & Cortes, 2004: 374)。进一步将这批立标语块进行句法结构分析( 方法参照的文献同上: 380-381) 则发现它们有近 86% 属于小句类语块( 即带有主谓结构), 而且主语基本是 I, you, we; 谓语基本是 know, want, think 或者 going to, have to 等; 余下的( 见表 3 划底线的语块) 则大都为表情态的短语。值得注意的就是其中有半数以上( 39% + 17% = 56%) 属于否定语序和疑问语序, 见表 3。

基于此, 笔者归纳出四词立标语块的关键词及其典型搭配范式<sup>④</sup>

① 即从词汇、语法、语义和语用四个搭配层面构建语言使用范式的词汇语法( lexical grammar), 详见 Sinclair( 2004: 164-176)

② LINDSEI-bri 和 LINDSEI-chin 均为 50 名大学生在同样时间内针对同样话题作口语表述的语料。

③ 此处参考 Sinclair (2004: 28) 关于语言范式观察的提取频数不少于两次的原则。

④ 该范式中的/表示“或者”, ( ) 表示“可有可无”。

肯定语序( 44% )	否定语序( 39% )	疑问语序( 17% )
and you have to <u>going to be a</u> <u>going to have a</u> <u>going to have to</u> I think it was I thought it was I was going to if you want to it' s going to be <u>to be able to</u> we' er going to have you have to be you have to do you know what I you want me to you want to go	and I don' t know I don' t know how I don' t know if I don' t know what I don' t know whether I don' t know why I don' t think so I don' t want to I' m not going to <u>not going to be</u> oh I don' t know well I don' t know you don' t have to you don' t want to	are we going to are you going to do you want a do you want me do you want to what do you want

表 3 Biber 立标语表的三种语序分类

( 疑问式) + 第一/ 第二人称代词 + ( 否定式) + 心理类动词/ 情态类词范式的核心由第一或第二人称代词加上心理类动词或情态类词两个部分组成。换句话说, 英语口语机切四词语块只要含有这两类关键词, 不论其搭配是肯定、否定还是疑问语序, 都具有潜在的立标语功能。其中半数以上的否定和疑问语序更表明说话者在使用立标语时往往采用否定或疑问的语气来表现自己对所表述的命题“ 不知道” 或“ 不肯定”; 或者用弱式语气表达某命题仅仅是个人意愿或预测; 又或者

是不得已向对方提出要求。总体上呈现一种非断然的、模糊的、或语气和缓的态度——这是立标语的语义和语用常态。

4. 2 立标语典型范式的验证与扩充

将 Biber 立标语表及以上归纳的典型范式带到 Scott 和 O' Keffe 的两个英语本族语者大型口语语料库的最高频四词语表( 见表 1 的编号 2 和 3) 里做验证和扩充: 一是做词汇形式的完全匹配; 二是将含有以上典型范式的关键词及其搭配模式的语块也纳入立标语范围。结果见表 4。

	与 Biber 立标语表完全吻合	可纳入典型立标范式的新增语块	扩充后的立标语块比例
Scott 最高频四词语表( 40 种)	17/ 40= 43%	I don' t know I know what I mean no I don' t think what do you mean what do you think	22/ 40= 55%
O' Keffe 最高频四词语表( 20 种)	7/ 20= 35%	know what I mean you know I mean do you know what what do you think	11/ 20= 55%

表 4 Biber 立标语表的验证和扩充

表 4 显示 Biber 立标语表在这两个英语本族语者大型口语语料库中有很高的认同度, 因为不论是在 Scott 的最高频 40 个, 还是 O' Keffe 的最高频 20 个机切四词语表中, 都有 1/ 3 以上在词汇形式上完全吻合, 更有一半以上( 55%) 的语块可被纳入立标

语的典型范式。这再一次证明口语高频复现的语块往往具有“ 语用整全性( pragmatic integrity)” (O' Keffe, McCarthy & Carter, 2007: 75)。如此突出的立标语功能似可从浮现语法的角度得以解释: 高频复现是语言动态发展的重要动因, 如同语法化

过程见证了某些曾经的实义词(如副词 so)在经过长期频繁使用之后演变成抽象语法词(如连词 so)一样,许多原先用于表达命题内容的语法结构经历高频使用之后也演变成语用程式语。例如 I think, you know, I mean 等现在已经语用化为表示说话者认知情态的立场标记语。

4.3 中、英大学生使用立标语块的对比

将 Biber 的 36 个立标语块以及以上新增的 7 个组成扩充后的立标语表(共 43 个)带到两个同质同源的微型语料库的 NS 四词语表和 NNS 四词语表(见表 1 编号 4 和 5)中调查形式完全匹配的语块,结果如下:

NS(英语本族语大学生): 29/ 43 = 67%, 194/ 117000x10000= 17(次/万字)

NNS(中国英语大学生): 10/ 43= 23%, 52/ 63500x10000 = 8(次/万字)

该结果显示,在表述同样话题的情况下<sup>①</sup>, NS 仅在 10 余万词次的会话中就使用了典型立标语块品种的 67%(29 种,共 194 次),平均每万字使用 17 次;而 NNS 使用的种类仅约为 NS 的三分之一(10: 29);使用频率不足 NS 的一半(8: 17)。倘若对比这 43 种立标语块中较常用的 32 种,即平均每百万词出现 40 至 100 次以上的语块,则结果如下:

NS(英语本族语大学生): 21/ 32 = 66%, 121/ 117000x10000= 10(次/万字)

NNS(中国英语大学生): 8/ 32= 25%, 44/ 63500x10000 = 7(次/万字)

以上数据显示,在同等时间内同样话题, NNS 对于这 32 种立标语块也只使用了其中的 25%,约为 NS 的 1/ 3。但与前面 43 种立标语块的调查结果相比,这批中国大学生对英语本族语者常用的立标语块(如 I don' t know how/ what/ whether, If you want to, what do you think 等)显然有较多的习得,因为他们使用的 10 种中就有 8 种为常用语块,而且占了使用总次数的 85%(44/ 52)。

为了进一步调查 NNS 立标词块的语义和语用特征,笔者使用了扩展式的范式调查,即将上述立标语典型范式的第一类关键词(第一和第二人称代词)作为检索项,再到 NS 和 NNS 各自的四

Rank	Freq	Stat	Collocate
1	58	7.61514	don't
2	52	7.20910	was
3	47	6.85527	mean
4	45	6.70521	think
5	40	6.32365	know
6	32	5.65683	i
7	22	4.68949	went
8	15	3.87260	had
9	13	3.60331	like
10	12	3.46223	can
11	12	3.46137	have
12	10	3.16196	didn't
13	8	2.82697	just
14	7	2.64570	could
15	7	2.64556	can't
16	7	2.64189	want
17	6	2.44929	got
18	6	2.44924	came
19	6	2.44918	really
20	5	2.23601	used
21	5	2.23585	thought
22	5	2.23434	erm

图 1 NS 的 I/ we/you 的 1R 搭配词

Rank	Freq	Stat	Collocate
1	129	11.35605	think
2	105	10.24694	i
3	41	6.40153	want
4	36	5.99842	have
5	34	5.82984	can
6	31	5.56631	like
7	30	5.47634	don't
8	27	5.19505	know
9	25	4.99711	was
10	21	4.58157	will
11	20	4.47116	went
12	18	4.24202	mean
13	17	4.12229	would
14	16	3.99897	just
15	12	3.45472	and
16	9	2.99585	so
17	8	2.82794	am
18	7	2.64316	we
19	7	2.63356	er
20	5	2.23534	should
21	5	2.23373	when
22	5	2.23128	you
23	5	2.22916	in

图 2 NNS 的 I/ we/you 的 1R 搭配词

词语表中提取其右一(1R)搭配词表,并观察其中显著性搭配(T-score 等于或高于 2.0000)的词汇(见图 1 和图 2),发现了两者在目标词与搭配词的共选范式上有语义和功能上的异同。

首先,图 1 和 2 再次证明机切四词语块中,但凡含有第一或第二人称代词时,其后面显著性搭配词大多为 mean, think, know, like, want 等心理动词或 have/ had( to), can/ could, will/ would/ should 等情态动词以及 just, really 等情态副词,其均呈现立标语功能倾向。但是,NS 比 NNS 有明显多的否定类搭配词,如 don' t, didn' t, can' 等,打开其四词语境也发现其后基本还是 know ,

<sup>①</sup> LINDSEI-bri 和 LINDSEI-chin 均为 50 名大学生在同样时间内针对同样话题作口语表述的语料。  
©1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

think, mind, remember, have to, need to 等心理动词和情态词<sup>①</sup>, 表达出说话者在陈述命题时的含糊或弱式否定的语气倾向。例如, 图 1 和图 2 显示 NS 的 I think 频数虽然仅为 NNS 的三分之一(45< 129), 但语境观察却发现前者使用否定式 I don't think 的比例是后者的 2 倍, 表现出说话者即便要否定命题也采用比较迂回的表达方式

I I'd like to  
but I'd like to  
erm I'd like to  
I'd like to see  
think I'd like to  
I'd like to do  
I'd I'd like to  
so I'd like to  
yeah I'd like to  
what I'd like to  
and I'd like to  
I'd like to go  
I'd like to teach  
I I do like  
and I was like  
I was like oh  
I I'd like to  
I I do like  
that I don't like  
I don't like that  
I don't like it  
I would like to  
I'd I'd like to  
I'd just like to

图 3 NS 的 I 和 like 搭配

I like to be  
I like to do  
I like I like  
I like to talk  
the film I like  
I I like the  
mm I like I  
and I I like  
I like the film  
yes I like the  
I like the I  
I I like to  
I like the the  
I I I like  
and also I like  
I I like I  
I I like gone  
erm I like the  
film I like most  
I like most is  
I like gone with  
I like this film  
I like it very  
I I like the  
- - -

图 4 NNS 的 I 和 like 搭配

Rank	Freq	Stat	Collocate
1	58	7.61514	don't
2	52	7.20910	was
3	47	6.85527	mean
4	45	6.70521	think
5	40	6.32365	know
6	32	5.65683	i
7	22	4.68949	went
8	15	3.87260	had
9	13	3.60331	like
10	12	3.46223	can
11	12	3.46137	have
12	10	3.16196	didn't
13	8	2.82697	just
14	7	2.64570	could
15	7	2.64556	can't
16	7	2.64189	want
17	6	2.44929	got
18	6	2.44924	came
19	6	2.44918	really
20	5	2.23601	used
21	5	2.23585	thought
22	5	2.23434	erm

图 5 NNS 的 I 和 will 搭配

图 3 和图 4 对比了 NS 和 NNS 的 I + like 搭配。前面图 1 和图 2 已显示 NS 的 I/you/we + like 搭配频数不及 NNS 的一半(13: 31), 但图 3 显示 NS 的 I 与 like 搭配基本用于表达意愿情态(如 I'd like to ...); 而图 4 显示 NNS 基本用它来表示“我喜好(I like sth.)”。图 5 显示 NNS 频繁使用 I 与较强势情态词 will 搭配, 但这种搭配在 NS 四词语表中显示为 0, 取而代之的是 I'll ..., I'd ..., I would ... 等弱势情态语。正如 Stubbs (1996: 202, 转引自桂诗春, 2009: 120) 指出, “当我们说话和写作时, 我们对所谈及的事物往往是模糊的、间接的、不清晰的; 表面看来, 这好像是人类语言的一种不适宜性, 其实只有那些对交际目的持粗浅看法的人才这样看。模糊和间接有诸多用途, 礼貌是偏离那些表面清楚或合理行为的一个明显原因, 只是在某种场合才需要精确。我们

(试比较 I don't think it is ... 和 it is not ...)。又例如, 图 2 显示 NNS 使用强势情态词 做搭配词的统计学频率显著高于 NS, 如 can (5.82984 : 3.46223), 而且图 2 中的 will 和 should 均未见于图 1。以下图 3 和图 4 是使用 Concgram<sup>②</sup> 提取的 I + like 和 I+ will 两个搭配式的实例演示。

经常发出信号, 以表明话语的模糊性”。如此看来, NNS 的立标语不仅品种稀, 数量少, 而且还显得语气断然决然而缺乏留有余地。桂诗春先生曾指出, “中国的英语学习者在论文写作中往往忽略表达作者对所表达内容态度的手段, 他们要么就是放弃使用, 要么就是以为这无非是一个可有可无的副词”(桂诗春, 2009: 120)。看来这批中国大学生的口语表达也不例外。

5. 反思和启示

本次研究凸显了英语口语立标语的广泛性和典型性, 以及中国高级英语学习者立标语使用能力的明显缺失, 引起我们反思以往英语教学中对立标语的忽视。国外已有语料库研究表明, 若按语义功能归类排行, 英语 2000 个最常用词汇中, 位居头三位的就有情态类和立标类(O' Keefe, McCarthy & Carter, 2007: 37-38)。本研究也发现

① 此处的强势是相对而言, 如 can 相对 could, will 相对 I'll 和 would, should 相对 will 和 would 等。  
② 该检索工具可以自动提取两个或多个搭配词之间跨距不同和排序不同的短语序列, 例如在四词语表中设置 I 和 know, 可呈现包括 far as I know, I know that I, but I only know, I really don't know, know what I mean 等语块。详见 Warren, 2008: 316-317。

多个大型英语本族语者口语语料库的机切四词语表中,均有 50% 以上的高频语块属于立标语类(尽管高频切割点有所不同)。这表明立标语其实是由最常用的基本词汇构成,但以往教学却很少关注这些基本词的常用范式和语用功能。如果说外语教学初始阶段的目标是学会表达简单的概念或命题,那么到了中、高级阶段就很有必要讲究说话者对命题内容的态度。因为“当说话人或作者谈及什么的时候,他们总是把自己对它的观点进行编码,这种表示方式渗透在语言使用的各种场合。所有的话语都要把这种观点进行编码,而对这些观点及其意义标记的描述就成为语言学的中心课题”(Stubbs, 1996: 202, 转引自桂诗春, 2009: 120)。因此,笔者认为重视立标语教学是提升外语学习者“语用域意识”(桂诗春, 2009: 117-118)的高层次教学标记。

那么应该教什么样的立标语? 是孤立地教某个情态动词或情态副词? 还是教立标语块? 具体教哪些语块? 本研究通过语料库的词汇语法手段,揭示了立标语块的典型范式,尤其是它高度集中的词汇、句法形式及其带倾向性的语义特征和语用功能,为识别和提取立标语提供了实证性参考。较常用的立标语块(如本文列举的 43 种)应该成为高级口语教学的显性目标和内容,因为它们已经是比较固定的程式语。但是教学又不必局限于这批语块,因为本研究构建的立标语典型范式更似一个半开放体,可以拓展到其他有着类似语义但形式却不尽相同的立标语,这就是归纳范式的优势。另外,通过机切语块凸显立标语的频数和种类的方法,还可能发展成对教材呈现立标语或学习者使用立标语的测评指标和手段,它涉及到对立标语的语料库研究成果进行“教学加工”的其他过程,对此我们将另文探讨。

### 参考文献

- Biber, D., S. Conrad & V. Cortes. 2004. "If you look at ...: Lexical bundles in university teaching and textbooks". *Applied Linguistics* 25/3.
- Biber, D. & F. Barbieri. 2007. "Lexical bundles in university spoken registers". *English for Specific Purposes*. doi: 10.1016/j.esp.2006.08.003.
- Coates, J. 1988. "The acquisition of the meaning of modality in children aged eight and twelve". *Journal of ChildLanguage* 15.
- Cook, W. 1978. "Semantic structure of English modals". *TESOL Quarterly* 12.
- DeCarrio, J. 1986. "Tense, aspect, and time in the English modality system". *TESOL Quarterly* 20.
- Du Bois, J. W. 2007. "The stance triangle". In R. Englebretson (ed.). *Stancetaking in Discourse*. Amsterdam and Philadelphia: John Benjamins.
- He Anping. 2008. "Application of Corpus Linguistics to EFL Teacher Education in China". In A. Frankenberg-Garcia, T. Rkibi, B. Cruz et al (eds.). *Proceedings of the 8th Teaching and Language Corpora Conference*. Lisbon: British Council & Benjamins.
- Kuczaj, S. A. & M. J. Daly. 1979. "The development of hypothetical reference in the speech of young children". *Journal of Child Language* 6.
- O'Keeffe, A., M. McCarthy, R. Carter. 2007. *From Corpus to Classroom: language use and language teaching*. New York: Cambridge University Press.
- Perkins, M. R. 1983. *Modal Expressions in English*. London: Frances Printer.
- Scott, M. & C. Tribble. 2006. *Textual Patterns: Key words and corpus analysis in language education*. Amsterdam: Benjamins.
- Sinclair, J. 2004. *Trust the Text: Language, corpus and discourse*. London: Routledge.
- Stubbs, M. 1996. *Text and Corpus Analysis: Computer-assisted studies of language and culture*. Oxford: Blackwell.
- Warren, M. 2008. "A data-driven approach to learning and teaching phraseology". In A. Frankenberg-Garcia, T. Rkibi, B. Cruz et al (eds.). *Proceedings of the 8th Teaching and Language Corpora Conference*. Lisbon: British Council & Benjamins.
- 桂诗春, 2009, 《基于语料库的英语语言学语体分析》, 北京: 外语教学与研究出版社。
- 梁茂成, 2008, 中国大学生英语笔语中的情态序列研究, 《外语教学与研究》第 1 期。
- 徐曼菲, 2009, On Epistemic Markers in Spoken English: A Case study of "I think", 广东外语外贸大学博士论文。
- 作者简介: 何安平, 华南师范大学外国语言文化学院英语教授, 博士生导师, 研究方向: 语料库语言学, 英语教育教学。
- (通讯地址: 510631 广州石牌华南师范大学外文学院)

(下转第 37 页)

London: Routledge.

Crystal, D. & D. Davy . 1969. *Investigating English Style*. Bloomington: Indiana University Press.

Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Johns, T. F. 1991. “Should you be persuaded: Two examples of data-driven learning”. In T. F. Johns & P. King (Eds.) . *Classroom Concordancing*. Birmingham: ELR Journal.

O’ Barr & M. William. 1982. *Linguistic Evidence — Language, Power, and Strategy in the Courtroom*. San Diego: Academic Press.

胡锦涛, 2005, 深化亚太合作, 共创和谐未来, 在韩国釜山第 13 届亚太经合组织领导人会议上的发言。

柯飞, 2002, 双语库: 翻译研究新途径,《外语与外语教学》第 9 期。

刘康龙、穆雷, 2006, 语料库语言学与翻译研究,《中国翻

译》第 1 期。

秦洪武、王克非, 2007, 对应语料库在翻译教学中的应用: 理论依据和实施原则,《中国翻译》第 5 期。

王克非, 2004, 双语平行语料库在翻译教学上的用途,《外语电化教学》第 6 期。

王克非、秦洪武、王海霞, 2007, 双语对应语料库翻译教学平台的应用初探,《外语电化教学》第 6 期。

吴春梅, 2007, 语料库在翻译教学中的应用,《山西广播电视大学学报》第 3 期。

张培基, 1980,《英汉翻译教程》, 上海: 上海外语教育出版社。

作者简介: 朱晓敏(1979—), 男, 江苏南京人, 南京大学外国语学院在读博士生, 常熟理工学院外国语学院讲师。

研究方向: 语言、翻译与信息技术。

(通讯地址: 210093 江苏省南京市南京大学鼓楼校区陶园 2 幢 908 室)

On the Employment of COCA and CCL in Translation Teaching

ZHU Xiaomin

**Abstract:** As a new approach of exploring language phenomena and nature, corpus linguistics has witnessed encouraging development since the 1980s. Corpus has virtually penetrated into every field of language research. At present, the employment of bilingual parallel corpora and translational corpora in translation teaching and research is well studied. However, study on the use of monolingual corpora in translation teaching and research is scant. By using freely-accessed Corpus of Contemporary American English( COCA) and Corpus of Chinese Language ( CCL) in translation teaching, this article endeavors to prove that monolingual corpora are useful tools in translation teaching.

**Key words:** corpus, COCA, CCL, translation teaching

(上接第 31 页)

Exploring Stance Bundles in Spoken English — A Corpus Perspective

HE Anping

**Abstract:** This is a study on stance bundles in spoken English corpora. Enlightened by Sinclair’s (2004) “lexical grammar” theory, the author did a “second processing” of some 4-word bundle lists automatically retrieved by previous corpus linguists from several large spoken English corpora, and worked out a typically formal, semantic and pragmatic pattern of stance bundles. The pattern is then used to compare two homogeneous mini-corpora based on British and Chinese university students respectively. The results reveal that stance bundles in these two types of students have marked difference not only in terms of frequency and types, but also in semantic orientation and pragmatic functions, thus bringing about reflections on EFL education in China.

**Key words:** corpus, stance bundle, spoken English, English teaching