

基于学习者英语语料的类联接研究 概念、方法及例析

许家金, 熊文新

(北京外国语大学 中国外语教育研究中心, 北京 100089)

摘 要: 本文主张类联接研究应兼顾词形、语法类别以及篇章和语类等不同语言层面来观察搭配行为。具体可包含: 单个语法范畴的类联接、语法串分析、词形与语法范畴的类联接、不连续结构类联接、语法范畴与话语、语体的类联接等。本文结合中国英语学习者语料进行了示例分析, 旨在说明可以结合类联接开展中间语对比分析、语体分析等。

关键词: 类联接; 学习者语料库; 语料库语言学; 研究方法; 类联接分析器

中图分类号: H319.3

文献标识码: A

文章编号: 1001-5795(2009)03-0018-0006

搭配(或称共现关系)是语料库语言学的核心概念之一。目前的搭配研究已不单限于词汇间的共现关系, 很多学者试图借助语料库语言学的方法观察语法及其他语言相关层面的共现现象, 其中以 Hoey 的“词汇触发”(lexical priming)理论涵盖层面最宽泛。Hoey 将词汇间的共现关系扩展到词汇与其在篇章中的出现位置以及在不同文体中的分布等十个层面(2005: 13)。事实上, 对于搭配范围的不同处理, 体现了不同的语境观, 即对共现语境范围的不同认识。本文将基于赋码语料探讨如何兼顾词形和语法类别, 以及篇章和语类等不同层面来观察搭配问题。本文示例所用语料主要为中国英语学习者英语。

1 类联接及其扩展

1.1 类联接

“类联接”(colligation, 又译作“类联结”或“类连接”)即语法层面的搭配关系。该术语由 Firth(1957[1968]: 181)提出, 他将 I watched him 背后体现的“第一人称代词 + 动词过去时 + 第三人称代词”语法类别共现关系称为类联接。这一术语中的“类”即词类或

语法类别之意。换言之, 类联接就是有关词类或语法类别的共现关系。而新近的一些文献, 如 Baker, Hardie & McEnery(2006: 36)的《语料库语言学术语汇编》里对 colligation 这一词条的解释除了原有的句法共现关系外, 还专门以 window 为例, 说明 window 常和介词构成类联接。

目前的类联接研究很大程度得益于一些优良的自动词性赋码工具(如 CLAWS, 准确率高达 96 - 7%)(Garside & Smith, 1997)。因此通过赋码语料研究词类或语法类别的共现关系成为可能。相关研究大致分三类。一、有关赋码串(tag sequences 或 POSgrams)和语法搭配的总体分析(如, Aarts & Granger, 1998; Tono, 1999; Hunston & Francis, 2000; 濮建忠, 2003b); 二、针对某一词类的总体研究, 通常以对动词的考察为主(如, Francis, Hunston & Manning, 1996, 1998; Mindt, 1995, 2002; 濮建忠, 2003a; 孙海燕, 2004; 孙海燕, 陈永捷, 2006); 三、还有一些专题研究。如 Biber & Reppen(1998)关于补语从句, Gabrielatos(2007)关于 if 条件句的类联接, 张军、李文中(2004)关于动名搭配, 以及梁茂成(2008)有关情态序列的研究等。

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。

熊文新: 男, 博士、副教授。研究方向: 计算语言学。

收稿日期: 2008-12-25

基金项目: 本研究得到教育部人文社会科学研究项目(08JC740002)资助; 本文的撰写还得到梁茂成教授的指导, 在此一并致谢。

研究类联接有助于从更抽象的语法层面探究语言结构间的组合关系,它超出了一般的词汇搭配研究。类联接虽可使语言研究向抽象层面推进一步,然而实际研究中有时我们会希望考察诸如 I think 这样的结构的词汇及语法层面的共现关系(见图1),以及像 so ... that, in order ..., as ... as 等结构同前后(及中间)的句法结构的搭配关系。此时,原本的类联接概念就需要作必要的扩展。

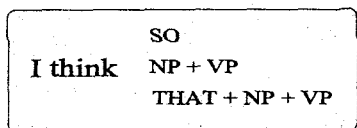


图1 I think 的类联接

1.2 类联接的扩展

以下,我们通过图1中 I think 这个例子来分析一下类联接概念扩展的理论和实践的必要性。不难发现,像 I think 这样的结构涉及到的语法成分共现关系应比 Firth 类联接的本义(即:词类+词类)要更灵活和多样。

从图1中可以看到,think 和 so 构成的是词汇层面的搭配;紧缩小句 I think 和副词 so 的搭配;I think 后面接 that 从句的用法;I think 后 that 省略的情形。可见:一、I think 的横向组合(syntagmatic)关系不光有属于词类(NP)、语法类别(that 从句)以及具体词形(so)。二、I think 右边的这种纵向聚合(paradigmatic)选择关系也是经常需要考察的。这里体现的横向和纵向关系也就是 Sinclair(1991)所谓的习语原则(Idiom Principle)和开放选择原则(Open Choice Principle)。只有两方面兼顾,对语言结构的描写才更完整。通过这个例子,大致可看出词语搭配关系同语法共现关系如何能加以整合。简言之,在对类联接概念的扩展中,只要有语法范畴参与就可以视作类联接,其中允许有具体词形的参与。

将类联接关系扩展到词形和语法类别两个层面,不仅仅有利于分析,事实上这也很可能是语言习得、心理表征的现实。语言学习初期,很可能学生先接触到 I think so 这样的表达。随着学习的深入,才逐步掌握了 I think 后接从句等其他用法。就心理现实性而言,并非像在做语言分析那样,按词汇层、语法层、音位层那样完全独立,而可能是有并行,有交叉。I think so 很有可能在表征时的联结强度与 THAT + NP + VP 等同,甚至更强,即有些词和结构始终绑定在一起。人们的词汇、语法知识很可能是在概念生成的基础上词、类、义

俱备的、交互的。语言知识是否按照语言学家分析的那样清晰划分为语音、词汇、句法、语义层呢?为了分析的目的,或许可以这么做,但人们的语言知识的事实很可能是上述层面交叉的。

按上面对类联接概念的扩展,本文主张类联接应既包括词汇和语法层面的横组合关系,也可以是词汇和语法的纵聚合关系。此外,如果研究需要,我们还可以结合语法类别同篇章和文体进行分析。以下是在我们对类联接内涵的扩展基础上提出的类联接的研究方法、分析工具和实例分析。

2 研究方法、工具及实例分析

2.1 基于赋码语料的类联接研究方法和特点

一旦语料施以了词性码,便可从中提取某些结构,并观察其典型分布和表现形式。语料库方法的优势就在于它可以在量化基础上发现语言现象的宏观使用状况。诚然,我们不应迷信量化数据,同时应充分利用量化分析方法带来的便利和效率。

本文涉及的类联接研究方法主要有以下特点:

- (1) 根据研究目的,词码兼顾,更完整地考察语言结构的共现关系;
- (2) 运用正则表达式,最大限度挖掘检索深度(可以检索词组、词性赋码串、不连续的结构,以及同时检索多个结构等);
- (3) 根据需要,以注重横组合关系为主,同时考察纵聚合关系。

2.2 类联接分析器、正则表达式

我们用于分析类联接的主要工具包括:类联接分析器(Colligator)和正则表达式(regular expressions)。其中类联接分析器是北京外国语大学中国外语研究中心研究人员自行开发的用于分析英语句型式(pattern)特征的工具。正则表达式可以用来提取较为复杂的句法结构。例如,英语被动语态可以用正则表达式一次提取出来(见图3)。

下面图2是利用类联接分析器检索情态动词的语境共现状况;如图所示,通过情态动词的词性赋码 VM^①,我们可以检索出学生语料中各种情态动词使用型式(此处左右语境分设为左1和右2),结果显示在左上方的 Results 窗口。第1列显示为其实际出现频率,第2列是情态动词的左边共现情况;第4列是情态

① 本文中涉及的词性赋码集采用的是 CLAWS7 Tagset, 详见 <http://ucrel.lancs.ac.uk/claws7tags.html>。

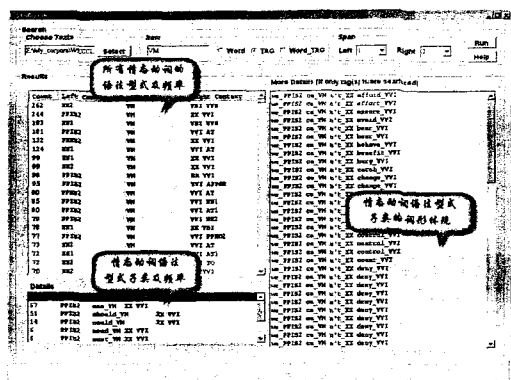


图2 类联接分析器提取的情态动词类联接的结果显示
动词的右边共现情况。Results 窗口中首行的抽象型式通过点击可展开为左下方 Details 窗口由词和码混合组成的搭配型式。若要进一步观察该动词型式,Details 窗口的结果还可被进一步具体化成右侧窗口的词码同现的形式。

类似地,图3是利用类联接分析器检索被动语态的语境共现状况(表达式为:“\S+_VB\S+_s(\S+_ [RX]\S+_s)*\S+_V\SN”,语境设置为左1)。

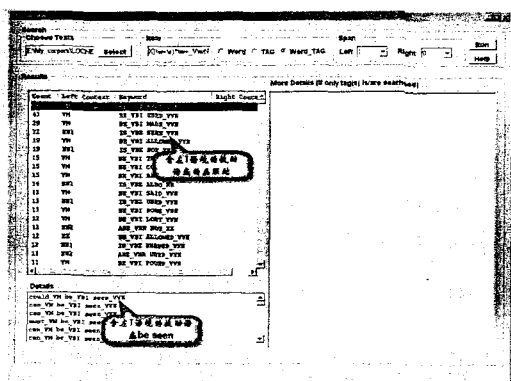


图3 类联接分析器提取的被动语态类联接的结果显示

该工具可较为方便地从大量语料中首先提炼出某语言形式的类联接状况,还可在限定某些词类或词形的情况下深入分析,最后还可以挖掘该类动词结构在语料中的实际使用状况。有关该工具的详细使用可参看软件附带的使用手册(梁茂成,熊文新,许家金,2008)。

同样的过程可用于口语、书面语语料库,学习者语料和英语母语语料,因而,可以进行对比分析。借助类联接分析器提供的频次信息,还可以进行统计分析等等。

由于 CLAWS 赋码集对动词的人称、单复数、过去式、过去分词等都有准确的赋码。我们可以借助这些

赋码信息和正则表达式对绝大多数动词语法型式进行分析。以下结合实例对扩展后的类联接分析方法做一简介。

2.3 单个语法范畴的类联接

表1是经由类联接分析器(参数设置为:检索项 AT(1)?,语境为 L0 和 R4)提取出来的学习者语料中所有冠词的用法。表1的上半部呈现的是冠词 the 的三种具体表现形式及在学习者语料库中的出现频数。表的下半部分是将 the 的最典型类联接“the + 单数名词 + 介词 + the + 单数名词”进一步展开,进而观察到学生们实际使用这类结构的情况。从纵聚合关系上来看,例句中的名词还可进一步归类分析。为了教学和研究的目,我们还可将英语母语者的同样数据提取出来加以比照,或者在学习者内部的低中高水平进行比较。

表1 冠词的类联接表现

冠词在语料库中的总频数	1265
the	859
a	362
an	44
定冠词 the 的典型使用 (语境设为右4)	16次 the + 单数名词 + 介词 + the + 单数名词 the development of the science the effect of the education
学生语料中的典型例句	the quality of the education the rest of the life the top of the world

2.4 语法串分析

2.4.1 任意赋码串

这里列出的是中国学习者语料库(NNS)和英美大学生语料库(NNS)排前10的3~4词的语法串(见表2)。总体而言,两组学生都较多使用介词短语(I开头的代码为介词)。然而,两组居第一的介词短语数量上存在显著差异(在0.05的显著性水平上,LL > 3.84)。通过类联接分析器可以进一步看出,在词形上,中国和英美大学生为什么存在差异,结果显示:在52个II APPGE NN1结构中,中国学生用了38次in_II my_APPGE opinion_NN1,其他还有5次in_II our_APPGE society_NN1,4次in_II our_APPGE country_NN1,3次in_II my_APPGE view_NN1,2次to_II our_APPGE nation_NN1,2次to_II our_APPGE society_NN1。在英美学生的275次AT NN1 IO AT结构中,有67次the_AT end_NN1 of_IO the_AT,排第一。根据Renouf(1992)

以及 PIE 网站上整个 BNC 的 POSgram 的结果, the end of the 都是排第一的四词短语。275 次中有 67 次 AT NN1 IO AT 只是一般的所属结构,往往是对其后的名词性结构的进一步描述。而中国学生显然在议论文中不恰当地加入了太多的个人判断和主观态度。这里仅举一例,对于表中其余的(乃至整个语料库里的)词性赋码串都还可进行深入研究。

表 2 中国大学生和英美大学生 3~4 词赋码串对比分析(前 10)

NNS 3 及 4 元结构(41,092)		NS 3 及 4 元结构(325,725)		LL
52	II APPGE NN1	275	AT NN1 IO AT	6.53
46	II AT NN1	273	II AT NN1 IO	3.08
41	II AT1 NN1	258	AT NN1 IO NN1	1.78
39	JJ II AT NN1	227	II AT NN1	2.95
35	IO APPGE NN1	200	II AT JJ NN1	2.95
33	VVI AT JJ NN1	178	AT JJ NN1 IO	3.77
29	NN1 IO APPGE NN1	172	NN1 IO AT NN1	1.95
28	NN1 II AT NN1	171	NN1 II AT NN1	1.54
28	AT JJ NN1 IO	156	AT JJ NN1	2.72
24	CSA PPIS2 DB VV0	151	AT1 JJ NN1	1.04

2.4.2 特定语法结构的类联接

如果希望了解中国学生在哪些情况下主谓一致容易出错,我们也可以用类联接分析器将学生语料中单复数不一致的情况检索出来。这里以“名词单数 + (实义)动词原型(检索设置:选择 TAG 检索方式,表达式为 $N \setminus w + 1 \ V \setminus w + 0$)”为例。检索可得表 3。

表 3 名词单数 + 动词原型检索结果(部分)

19	technology make	4	technology have
13	life become	3	coin have
6	society have	3	government have
6	world have	3	education have
4	government build	3	technology develop

表 3 显示的就是所用的语料里中国学生常见的名词单数同动词原型的主谓不一致的使用情况。通过进一步分析,我们可以进一步发现最常见的问题名词是: education, government, society, student, world, country, university, life, technology, patient, family, school 等。这其中 education 和 technology 同作文题目直接相关,因为使用频率极高,因而主谓一致出错率也很高。余下的词有一个共同的特点,就是多为集体名词,通过回溯原文,中国学生出现了过度概括的倾向,多将其视作为集体概念,其后的动词往往用原型。以至于 world, country, university, school 也视为集体概念。这当然也

或多或少与汉语没有单复数形态变化有关。当然,要深入挖掘还可以回到原文去分析。这些分析结果都可以作为教学实践的指导。

2.5 词形与语法范畴的类联接

2.5.1 含确定的语法范畴

比如若想知道学生英语中 the + 比较级形容词和副词的使用情况,在类联接分析器中,可以检索 the_\S + \s\S + _[JR]\S + R,从而得到类似于表 4 的数据。若要观察这一结构右侧语境的类联接,可以通过改变类联接分析器中的右侧跨距即可实现。

表 4 the + 比较级形容词和副词

71	The	JJR	形容词比较级
29	The	RGR	程度副词比较级
3	The	RRR	普通副词比较级

通过分析器中的纵向展开功能,还可以看到 29 个 The RGR 结构的一些具体用法,如 the less advantaged, the less motivated, the less privileged, the more active, the more common 等。

我们若想研究学习者英语中虚化动词 make 接单数和复数名词的使用情况(make a claim, makes a decision, making a comment, made policies 等),我们可以利用正表达式($ma[kd]e(s)?_V \setminus S + \setminus s(a|an_S + \setminus s)?_S + _N \setminus S +$)检索出相关用法,并在此基础上筛选我们感兴趣的用法。

2.5.2 语法类别不定的非连续结构

一些不连续的搭配和类联接也可以通过正则表达式进行检索。这同 Cheng et al. (2006)所提到的不连续的“跨词结构”(skipgram)和体现前后方向强共现关系的“框合结构”(concgram)有些类似,但这里我们更关心某些特定构式(如 It is + 形容词, as ... as, on ... basis 等)。例如,在类联接分析器中检索 it_\S + \s\S + _VB\S + \s\S + _J\S + (语境设置为 R3),可以得到“It + be 动词 + 形容词”后面通常都会接什么样的成分。示例结果显示,从我们试验语料中检索出的 264 次该结构中,有 82 次“It + is + 形容词 + for + 代词/名词短语”,48 次“It + is + 形容词 + 从属连词 that”和 39 次“It + is + 形容词 + to + 动词原型”。254 次英美学生语料中的同样结构显示的类联接分布呈现出的结果是:86 次“It + is + 形容词 + to + 动词原型”,58 次“It + is + 形容词 + 从属连词 that”,以及 20 次“It + is + 形容词 + for + 代词/名词短语”。中外学生在使用这一结构呈现出相反的使用倾向,尤其是“It is ADJ for sb. to

do sth”。的结构最为典型,这很可能就是在最初学习这一结构时老师教给的句型。

我们还可以检索近义形容词构成的这种结构:“It + is + obvious 类形容词”(表达式:“it_S + \s_S + _VB _S + \s((apparent)|(clear)|(discernible)|(evident)|(noticeable)|(observable)|(obvious))_J_S + ”)。以下我们简要介绍一下语法范畴在篇章层面和语体变异方面的关联和研究思路。

2.6 语法范畴与话语、语体和语用的类联接

从话语分析的各层面来看,我们可以观察某些句法结构的超句特征;作为使用中的语言的话语,我们观察某些句法结构体现出的语义韵;甚至还可以分析某些结构背后体现出的社会文化属性。因为有较大规模语料库的存在,我们除了可以研究某些语言结构在单个文本内的表现,还可以观察它们在文本间的共性和特点。这一方法特别适合中间语分析中的学习者和母语的对比,书面语体同口语体的对比。例如,一定的语法范畴往往倾向于出现在文本中的特定位置,或者出现在特定文本中。比如中国学生议论文中的顺序表达往往出现在段首,以标示文章的论述要点。英文中也有类似的结构,比如 Once upon a time 就总会出现在故事的开始。

一些语法结构更倾向于出现在口语中(比如,第一人称主格代词 + 动词原型 + 第二人称代词 + 动词原型,PPIS1 VVO PPY VVO,对应的结构为 I think you know 和 I mean you know 等),而书面语中典型的句法结构常为所属结构(如,定冠词 + 名词 + 介词,AT NN1 IO,对应的结构为 the end of 等)。这些都可以作为分析学生英语语体倾向的一些依据。

类联接分析器没有提供显示检索项在文本中的位置信息的功能。若要开展类似研究:一种方法是利用 WordSmith Tools 和 AntConc 的索引词图(concordance plot)功能辅助实现;另一种方法是将学生作文预先切分成引言、论述、结语分别存储,检索某句法范畴时对这三部分分别检索,这样可发现哪些结构在文章的特定部分容易出现。第二种方法常用于学术用途英语篇章分析中,类似研究可见 Biber, Connor & Upton (2007:第2、3、4、5章)。

若希望考察学生作文中表达复杂抽象概念的表述,我们可以尝试检索出每篇文章中的6个字母以上的名词(表达式:b_S{6,}_N_S +),以及6个字母以上的形容词(表达式:_b_S{6,}_J_S +)。同样的我们还可以通过动词的使用和分布分析文章中及物性的传

达。通过这种方法,我们比较高分学生和低分学生对复杂和抽象名词的选择情况以及它们在文中的分布和组织情况。这些名词的分布体现出学生的概念表达、组织和论述能力。从研究者角度,这些复杂抽象词组合可以揭示全文的“纲”和“领”。

而形容词体现出作者的态度、立场,对事物的评价。形容词的分布可以帮助了解学生的主观情态的表达。我们可以一次性地将同类型的语汇高亮显示出来,并放在全篇的上下文中来看,可以很好地揭示出内容和观点的衔接连贯的宏观结构。

3 结语

之所以我们主张用语料库的方法来进行类联接研究,是因为我们认同学习者的语法现状是一个群体发生的特征,或者说是这种语法体系是一种基于用法的语法(usage-based grammar)。而以往较大规模的语法范畴研究不太容易实现。通过前文的介绍,我们知道借助词性赋码的语料和合理的检索、分析方法,我们对很多句法结构的类联接特征都能够挖掘出来。

本文通过对类联接概念的扩展和研究方法的介绍,希望有助于发现和挖掘学生英语中的一些典型的句法特征偏误。尤其是在与英语母语者的相关句法范畴的语料对比中发现不足。本文介绍的工具和方法有效地将连续的横组合词汇搭配,扩展到语法层面以及词汇语法混合搭配,打破了词形、词类、句法的人为条块分割。在宏观研究方法上,可以结合中间语对比分析,文体分析,还可以进一步考察语义韵,以及同学生成绩参照对比。在应用层面上,类联接的研究可以应用于英语教学、词典编纂等方面。 □

参 考 文 献

- [1] Aarts, J. & S. Granger. Tag Sequences in Learner Corpora: A Key to Interlanguage Grammar and Discourse[A]. In S. Granger (ed.). Learner English on Computer[C]. London: Longman, 1998.
- [2] Baker, P., A. Hardie & A. McEnery. A Glossary of Corpus Linguistics [M]. Edinburgh: Edinburgh University Press, 2006.
- [3] Biber, D. & R. Reppen. Comparing Native and Learner Perspectives on English Grammar: A Study of Complement Clauses[A]. In Granger, S. (ed.). Learner English on Computer[C]. London: Longman, 1998: 145 - 158.
- [4] Biber, D., U. Connor & T. Upton. Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure[M].

- Amsterdam: John Benjamins, 2007.
- [5] Cheng, W., C. Greaves & M. Warren. From N-gram to Skipgram to Conegram[J]. *International Journal of Corpus Linguistics*, 2006, 11(4): 411-433.
- [6] Firth, J. A Synopsis of Linguistic Theory, 1930-55[A]. *Studies in Linguistic Analysis* (Special Volume of the Philological Society), 1957.
- [7] Reprinted in F. Palmer. *Selected Papers of J. R. Firth 1952-59*[C]. London and Harlow: Longmans, Green and Co., Ltd., 1968: 168-205.
- [8] Francis, G., S. Hunston & E. Manning. *Collins COBUILD Grammar Patterns 1: Verbs*[M]. London: HarperCollins, 1996.
- [9] Francis, G., S. Hunston & E. Manning. *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives*[M]. London: HarperCollins, 1998.
- [10] Gabrielatos, C. If-conditionals as Modal Colligations: A Corpus-based Investigation[A]. In M. Davies, P. Rayson, S. Hunston & P. Danielsson (eds.). *Proceedings of the Corpus Linguistics Conference: Corpus Linguistics 2007* [C]. Birmingham: University of Birmingham, 2007.
- [11] Garside, R. & N. Smith. A Hybrid Grammatical Tagger: CLAWS4[A]. In R. Garside, G. Leech & A. McEnery (eds.). *Corpus Annotation: Linguistic Information from Computer Text Corpora*[C]. London: Longman, 1997: 102-121.
- [12] Hoey, M. *Lexical Priming: A New Theory of Words and Language*[M]. London: Routledge, 2005.
- [13] Hunston, S. & G. Francis. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*[M]. Amsterdam: John Benjamins, 2000.
- [14] Mindt, D. A Corpus-based Grammar for ELT[A]. In B. Kettleman & G. Marko (eds.). *Teaching and Learning by Doing Corpus Analysis*[C]. Amsterdam: Rodopi, 2002.
- [15] Mindt, D. *An Empirical Grammar of the English Verb: Modal Verbs*[J]. Berlin: Cornelsen, 1995.
- [16] Sinclair, J. *Corpus, Concordance, Collocation*[M]. Oxford: Oxford University Press, 1991.
- [17] Tono, Y. A Corpus-based Analysis of Interlanguage Development: Analysing Part-of-speech Tag Sequences of EFL Learner Corpora[A]. In B. Lewandowska-Tomaszczyk & P. Melia (eds.). *PALC'99: Practical Applications in Language Corpora* [C]. Frankfurt: Peter Lang, 1999: 323-340.
- [18] 梁茂成, 熊文新, 许家金. *Colligator 使用手册*[Z]. 北京: 外语教学与研究出版社, 2008.
- [19] 梁茂成. 大学生英语笔语中情态序列的使用特点研究[J]. *外语教学与研究*, 2008, (1): 51-58.
- [20] 濮建忠. 学习者动词行为: 类联接、搭配及词块[M]. 开封: 河南大学出版社, 2003a.
- [21] 濮建忠. 英语词汇教学中的类联接、搭配及词块[J]. *外语教学与研究*, 2003b, (6): 438-445.
- [22] 孙海燕, 陈永捷. 中国英语学习者名词类联接的发展特征——基于赋码语料库的研究[J]. *外语教学与研究*, 2006, (4): 272-278.
- [23] 孙海燕. 基于语料库的学生英语形容词搭配语义特征探究[J]. *现代外语*, 2004, (4): 410-418.
- [24] 张军, 李文中. COL EC 中动名搭配模式及失误分析[J]. *外语教学*, 2004, (4): 30-32.

Learner Corpus Based Colligation Research: Concepts, Methods and Sample Analyses

XU Jia-jin, XIONG Wen-xin

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

Abstract: In this paper, we hold a more open view of co-selectiveness between grammar patterns and lexis, grammatical categories, textual properties and sometimes a combination of these categories. The object of inquiry for the present paper is Chinese learner English colligational behaviors on the basis of part-of-speech tagged texts. This paper offers a critical review of the concept, as well as tools, research methods and sample analyses.

Key words: Colligation; Learner Corpus; Corpus Linguistics; Research Method; Colligator