

国家哲学社会科学基金项目

YULIAOKU DE
DUANYU LINIAN JIQI JIAOXUE JIAGONG

语料库的 短语理念及其教学加工

主编 何安平

本书系国家哲学社科基金项目“‘短语理念’与语料
(编号09BYY067)结项专著

YULIAOKU DE

DUANYU LINIAN JIQI JIAOXUE JIAGONG

语料库的 短语理念及其教学加工

主编 何安平



广东高等教育出版社
Guangdong Higher Education Press

11314.5
198

图书在版编目 (CIP) 数据

语料库的短语理念及其教学加工/何安平主编. —广州: 广东高等教育出版社, 2013. 12

ISBN 978 - 7 - 5361 - 5040 - 9

I. ①语… II. ①何… III. ①英语 - 短语 - 研究 IV. ①H314.3

中国版本图书馆 CIP 数据核字 (2013) 第 300966 号

出版发行	广东高等教育出版社
	社址: 广州市天河区林和西横路
	邮编: 510500 电话: (020) 87553735 87551269
	http://www.gdgjs.com.cn
印 刷	佛山市浩文彩色印刷有限公司
开 本	787 毫米 × 1 092 毫米 1/16
印 张	13.25
字 数	320 千字
版 次	2013 年 12 月第 1 版
印 次	2013 年 12 月第 1 次
印 数	1 ~ 1 000 册
定 价	27.00 元



(版权所有 翻印必究)

前言

本书是国家哲学社会科学基金项目“‘短语理念’与语料库的‘教学加工’”(编号09BYY067)的最终研究成果。这是基于课题立项以来发表的一部专著、30余篇论文和完成的4部博士论文及20多篇硕士论文整合而成的学术论著,也是华南师范大学语料库团队全体成员将语料库语言学研究成果应用到我国英语教育教学实证研究的结晶。

全书分上、中、下三编,共18章。上编为语料库语言学的短语理论探究,分2章展开。该编首先论述了语料库语言学视角下短语研究的哲学和语言学理论基础,勾勒从Wittgenstein(1953)“意义即使用”的语言哲学观,到Firth(1957)的“语境意义论”,再到Sinclair(1996, 2004a)的“词汇语法理论和意义单位理念探究的历史沿革”。继而阐述语料库视角下六种不同于传统概念的短语形态,包括:意义转移单位、搭配框架、组织框架、模式语法、语义序列和“机切”语块。通过阐述每种短语形态的形式特征、表意功能和研究价值,凸显语料库语言学短语理念和研究方法的特色及其在构建语言意义单位,尤其是揭示语篇意义等方面的优势。

中编为多形态短语的应用研究,分9章展开。该编以国际通用语料库和自建的中国英语教育教学语料库为基础,探究我国英语学习者的短语应用能力。包括从语法模式视角探究英语大学生词汇产出能力的发展特征(第3章),从搭配框架探究大学生英语论文的词汇与文体特征(第4章),从高频“机切”语块探究大学生英语口语元话语能力(第5章),以及从语词的拓展意义单位视角探讨大学生英语口语立场标记语的表达特征(第6、7章),等等。同时该编还将多形态短语的研究视角延伸至教材话语(第8章)、课堂话语(第9章)和医学专业学术话语(第10章)以及外语学习者对高频语块的认知加工(第11章)等研究。

下编为短语的“教学加工”实证研究,分7章展开。该编介述语料库语言学的教学加工理念及国内外有关短语的教学加工研究现状(第12章),重点展示本课题应用短语理念和方法于国内英语教育教学方面的研究成果。其中包括大、中学英语教材中短语教学设计分析(第13章),基于语篇关键词与高级英语阅读教学探索(第14章),大型网络语料库辅助大学英语短语深度知识教学尝试(第15章),语料库辅助多模态词块教学课件的设计开发与



应用（第16、17章）以及语料库词块教学的效果评估等（第18章）。

本书的特色之一是在理论层面上厘清了语料库语言学有别于其他学派的短语研究历史沿革与理论发展，也提出了多形态短语研究的新理念。特色之二是在应用层面上从语料库短语理论，尤其是 Sinclair（2004）词汇语法理论的研究模式，分析我国大、中学英语教育的教材话语、课堂话语、专业学术话语以及学习者产出语等的短语相貌特征，获得了一系列有意义的新发现。特色之三是展示了该研究团队对短语理念和方法进行教学加工的多项实践成果，包括教材分析、教学设计、教学效果评估及多模态教学课件开发等，从一个侧面展示了我国语料库语言学应用于英语教育教学的最新成果。我们感谢国家哲学社会科学基金对本项目研究的资助，也感谢华南师范大学及其外国语言文化学院为研究提供的支持与便利。

何安平
2013年8月于广州

目 录

上编 语料库语言学的短语理论探究	(1)
第1章 从“意义即使用”哲学观到语言的意义单位探究	(1)
第2章 语料库视角下多种短语的形态、意义与功能	(8)
中编 多形态短语的应用研究	(21)
第3章 研究学习者英语的模式语法视角	(21)
第4章 搭配框架、意义单位与大学生英语论文写作	(30)
第5章 大学生英语口语元话语语块探究	(39)
第6章 大学生英语口语“立标语块”探究	(54)
第7章 认知类“立标语序列”探究——I think 的个案研究	(63)
第8章 英语教材话语的“立标语”探究	(77)
第9章 短语理论视角下英语教师课堂话语探究	(89)
第10章 医学英语关键词的语义关系探究	(97)
第11章 中国英语学习者对高频非习语英语程式语的心理表征研究	(106)
下编 短语的“教学加工”实证研究	(117)
第12章 语料库短语理念的“教学加工”研究综述	(117)
第13章 从“词汇语法”理论到 EFL 教材的短语教学设计	(123)
第14章 语篇关键词与高级阅读教学	(136)
第15章 大型本族语者语料库辅助英语词块教学初探	(147)
第16章 语料库辅助多模态积木式词块教学课件设计探究	(158)
第17章 “积木词块”英语教学模式的创设与应用	(167)
第18章 语料库课程与英语师范生的词块意识提升	(176)
结语	(187)
附录	(189)
参考文献	(191)

上编 语料库语言学的短语理论探究

第1章 从“意义即使用”哲学观到 语言的意义单位探究^①

1 引言

对语言意义的探索一直是语言哲学和语言学的核心议题。本章以文献述评的方式探讨从 Wittgenstein (1953) 的“意义即使用”语言哲学观,到 Firth (1957) 的“语境意义论”,再到 Sinclair (1996/2004a) 的语料库语言学对语言意义单位进行理论和实践探究的历史沿革,从而为语料库视角的短语研究奠定理论基础。首先分别述评上述三者的核心理念及相互关联,然后阐述短语作为意义单位的理据及研究方法。

2 Wittgenstein 的“意义即使用”语言哲学观

“意义即使用”是英籍奥地利哲学家 Wittgenstein 在 20 世纪其著作 *Philosophical Investigation* (1953/1959: 20) 中提出的著名命题——“the meaning of a word is its use in the language”。当时正值西方语言哲学的转折时期,语言哲学的焦点从抽象地分析语言转向了在社会语境中语言的日常操作,词语的意义问题成为西方语言哲学的基本问题之一 (Baghramian, 1998, 转引自钱冠连, 2001: 224 - 226)。当时一些哲学家 (如 Frege, 1892/1952) 将词语意义界定为它对某个具体目标的指认参照或者指向它所激活的心理联想,而 Wittgenstein (1953) 则提出了“不要问意义,要问使用”的主张 (转引自钱冠连, 2009: 9), 认为如果不了解词语的用法就无法理解语言的意义。在阐述他的语言游戏理论时, Wittgenstein 将语言的使用视为有规则的游戏并反复强调其使用语境对于理解其意义的重要作用。他认为参与语言交际的人之所以能够正确理解对方话语的意义,就在于他们的语言游戏有着能够相互理解的共同语境。正如没有必要去假定任何独立于 “good deed” 之外仍有被称为 “good” 的事情一样,人们也没有必要事前界定 “game” 一词,因为事前界定的结果往往使那些关于 “game” 即 “竞争” 的定义无法解释诸如 “捉迷藏” 和 “玩纸牌” 那样的 “game”。可见一旦不同交际者的语境发生了变化,他们对于同一词语的意义就会产生不同的理解和解释。以上论点 “表现出 Wittgenstein 在语言意义问题上的更彻底的语境决定论” (王健平, 2003: 125, 181), 同时也为 Firth (1957) 的

^① 本章作者为何安平 (华南师范大学), 基于其在《外语与外语教学》2013 年第 3 期和《外语研究》2009 年第 5 期发表的文章改写而成。



“语境意义论”和 Sinclair (2004) 基于语料库的“意义单位”探究提供了哲学层面的支撑。

3 Firth 的“语境意义论”

20 世纪 50 年代语言学伦敦学派的创始人 Firth (1957) 借鉴上述“意义即使用”的哲学观创建“语境意义论”(contextual theory of meaning) 学说。因为既然意义即使用, 那么使用就不能脱离语境, 所以只有基于语境的和功能的分析才能认识意义。关于语境的概念, 早在 20 世纪初就有英国人类学家 Malinowski (1993: 1-10) 指出语言的意义来自其使用的情景语境和文化语境: 前者指说话的时间、地点和对象; 后者指当地的文化习俗。但 Firth 的语境论更加关注语言符号和情景语境的相互关系并且将“宽泛的文化语境细分为语言学和社会学两个层面”, 其中特别强调语言学层面的“语音、语法、词汇和语义都是复杂语境的重要构成”(Firth, 1957: 19)。因此“研究语言应该基于来自语境中那些**典型的、反复出现的, 以及可重复观察到的完整话语事实**”^①; 而且“作为语境一个组成部分的语篇以及词语在该语篇中形成的语境模式(contextual patterning)都共同对该词语的意义起决定性作用”(Firth, 1957: 35)。这里的语境模式指的是词语在语境中形成的意义模式, 它“通过语言学技术或术语对反复出现的词语特征进行模式化的系统陈述”(Firth, 转引自 Palmer, 1968: 187)。“所有对意义的陈述其实都在陈述词语在语境中的相互关系(contextual relations)”, 即“通过词语的搭配了解词的意义(meaning by collocation)”(Firth, 1957: 11, 19)。这里的搭配包含了两个概念: 一是“词语搭配”(collocation), 指具体词语的重复共现(例如: dark 往往与 night 搭配); 二是“类联接”(colligation), 指语法上词性类别之间的习惯性搭配(例如: 形容词常与名词搭配、名词常与动词搭配)。前者强调词语搭配的意义跟词语本身所具有的固有特质一样重要; 后者则强调搭配在句法层面上的意义, 关注的是句法结构中语法类别的相互关系(转引自 Tognini-Bonelli, 2001: 5)。归结起来, Firth 的语境意义论有两个核心理念: 一是语言的意义来自其使用的语境, 这与“语言即使用”的哲学观保持一致; 二是语言的意义是一种模式(pattern), 一种基于典型的、反复出现的以及可重复观察到的语言事实而构建起来的抽象模式, 这是对上述哲学理念的拓展。虽然 Firth 本人没有对该理论实施大量的实证探究, 但是他提出的语境意义模式和搭配概念却为后人在语料库视角下研究意义单位奠定了理论基础。

4 Sinclair 的“词汇语法”理论和“意义单位”理念

以上“意义即使用”和“语境意义论”的核心都要透过语言使用中的语境来探究意义, 而在 Wittgenstein 和 Firth 身后到来的计算机时代催生的语料库语言学则更对此进行了理论上的拓展和实践上的方法创新。首先, 在对意义的理论探讨方面, Firth 当年的学生、语料库语言学的先行者 Sinclair (1991/2004) 及其研究团队提出了“词汇语法理论”(lexical grammar) 以挑战传统语法中“空位—填充”式(slot-and-filler)的意义构建模式, 从而将词语重新拉回意义研究的中心并提出了语言使用具有短语倾向性的论断。同

^① 引文中的黑体字系本章作者加附, 下同。

时,在意义探究的实践方面,他们又以语料库特有的计算机检索与呈现手段充实和深化了 Firth 的语境概念,而且提出了基于批量语境构建词项意义单位的工作模型。由于该节与本书的短语研究有直接联系,故以下用较大的篇幅介述。

4.1 挑战传统的“空位—填充”式语法

在传统语法中,语法和词汇一直被视为语言基本模式的两个范畴,其中语法的地位远远超过词汇。前者是基本框架,后者仅是填充物,语言的意义则来自预先由语法搭好的架构,而后填入符合语法词性的词。Sinclair (2004a: 165 – 169) 认为这种过于简单的“空位—填充”意义构建模式是“前计算机时代不够充分、不够完善的语言观察手段所导致”,如今已无法充分揭示语言的本质。例如,他在大量语料库研究中发现了以下一些无法用传统语法的词性概念来解释的词汇现象:

(1) 同一个词目 (lemma) 内不同词形 (type) 的意义其实并不一样,例如:动词 move (移动) 这个词目的其中一个词形 moving 就带有明显的感情意义 (如 sth. is moving and thought-provoking)。

(2) 形式相同而具有多种词性的词,其意义也不一样,例如:名词性的 combat 常指物理性质上的“格斗”,而动词性的 combat 则指社会意义上的“抗击”。

(3) 同一词性的词在语法结构中的出现位置或充当的成分并不一样,例如:人体部位名词 lap 前面通常只有带人所有格 (如: his 或 her) 并与介词搭配作状语 (如: on his lap), 而人体其他部位名称如 mouth 和 eye 却并非如此。

(4) 词性分类本身就有问题,例如:被称作介词的 of 其实极少用在动词后面作介词 (只有少数的情况如: remind of); 而多数情况下是用于连接两个名词或名词短语 (如: a thick carpet of snow)。

(5) 词或词组在现实使用中往往有语义限制,例如:动词词组 set in 的主语往往是不如意的情况 (如: the panic set in)。

(6) 有些修饰词其实并没有给被修饰词增添语义,例如: full range 中的 range 其实已经含有 full 的意思。

(7) 词汇搭配所形成的语义并非各个单词独立意义的叠加,例如: give them a speech 中的动词 give 已经语义虚化。

此外,语法概念的表现实体并非只有语法句式,例如:“否定式”这一个语法概念除了用 no 或 not 句式以外,还有诸如 hardly、scarcely、un_、in_ 等一批词汇或构词形式 (Sinclair, 2004a: 172)。

可见能够嵌入语法框架内的词并非只讲究词性相同;某种语法意义的实施也不单纯依靠句法形式。所以 Sinclair (2004a) 认为词汇和语法同等重要,而且要优先以词汇作为出发点来构建语言的意义单位。

4.2 充实和深化 Firth 的语境意义论

语料库语言学以计算机软件特有的研究手段 (如:语境共现行、词频表、关键词频表等) 在处理大量语料的过程中还前所未有地拓宽了对批量语境观察的视野,从而对 Firth 的语境概念 (见本章第3节) 有了更加细致的分类。例如 Sinclair (1991) 将 Firth 论述的语言学层面语境聚焦到以某个 (些) 词为核心并与其周边词共选的局部语境 (co-



text), 又称“词汇语境”。它通常为核心词前后各 4 至 5 个词的范围, 而且实践证明这一跨度 (span) 足以揭示大多数核心词与周边词语的共选特征 (Sinclair, 2003: 174)。可见 Sinclair 将局部语境和一般语境 (context) 区别开来是因为透过批量化的局部语境观察与分析可发现核心词在语境中的典型意义模式。此外还有学者, 如: Scott 和 Tribble (2006: 9 - 10), 将语境的定义分为范围和长度不等的多个层次, 其中包括: ①语篇中左右相邻的词。②语篇的完整句子。③语篇的段落。④陈述的事件。⑤书中的篇章。⑥整个语篇。⑦语篇所属的语域。⑧某语篇与其他相关的语篇。⑨语篇源出的文化环境。⑩读者接触语篇的时间及地点, 等等。从而廓清了语料库的局部语境概念在上述层级中的位置以及由此从小到大, 由近及远向外延伸的多个层面, 对研究词汇在其中各个层面的共选关系以构建词语的语境意义模式起着至关重要的作用。故有学者如 Stubbs (2001: 156 - 157) 称“语料库语言学本质上就是语境理论”。

4.3 创立“意义单位”的构建模型

在意义探究的实践方面, Sinclair (1996, 75 - 106/2004a: 24 - 28, 164 - 176) 还提出了构建词汇拓展意义单位的工作模型。该模型在 Firth 提出的搭配和类联接基础上增添了语义偏好 (semantic preference) 和语义韵 (semantic prosody) 的两个概念, 又被称为“3C2S”模式 (Römer, 2005: 12)。这 4 个符号分别代表了对任何词项进行意义分析的出发点及四个层面。其中 3C 指从核心词 (core, 通常形式不变) 出发, 通过其语境共现获取其与周边词语的典型词汇搭配和类联接 (collocation 和 colligation) 这两个层面; 2S 指基于前面两层分析而归纳核心词的语义偏好和语义韵 (semantic preference 和 semantic prosody) 这两个更高层面的共选模式。以下参考 Sinclair (2004a: 141 - 142, 200) 的论述分别对每个层面做进一步阐述。

第一层是词汇上的共选关系, 即词汇搭配。它指词与词的惯常相伴形式, 通常相伴跨度为核心词前后各 4 个词左右的范围。从横向看, 搭配词的认定比较简单和直观, 要确定某个词是否为典型搭配词的关键是看其频数, 可使用统计学上的 T 值或 MI 值来测定其显著高频性。然而从纵向看, 同享某个核心词的多个搭配词之间却不一定互为搭配, 例如 manual 和 restoration 都是 work 的惯常搭配, 但是彼此之间的搭配却不典型。

第二层是语法上的共选关系, 即类联接。它指实义词与语法词类之间, 或者是语法词类之间的惯常组合。典型的语法词类包括否定类 (negative)、情态类 (modal)、所属类 (possessive) 等。类联结以横向方式凸显核心词与搭配词在语境中形成的语法结构模式, 构建该模式尤其要关注搭配词的词性归属。此外类联结模式也往往融入词汇及其语义成分。例如: 名词短语 true feeling 的类联接是 possessive + true feeling, 这里的 possessive 类主要指人称代词如 my、your、his, 其中又因 your 与 true feeling 共现的频数最高而成为该类联接的典型表征; 而另外一种 possessive 类, 如: the true feeling of people (the ... of ...) 则是其次典型表现。

第三层是语义上的共选关系, 即语义偏好。它指词或词类之间在语义上的共选倾向, 即核心词与典型搭配词, 或者典型搭配词之间所共享的某些语义特征或语义范畴, 如 sports 类、suffering 类等。值得注意的是, 语义特征的归纳并不需计较词形或词性, 也不严格计较句法结构形式。例如: see 和 view 虽词形不同, see 和 visible 虽词性不同, to spot with 和 is invisible to 虽语法结构不同, 但都可归入 visibility 这一语义特征。

第四层是语用意义上的共选关系,即语义韵。它关注的是透过核心词及其周边语境所隐现的语言使用者的话语立场、态度情感或内含意义,通常弥漫在核心词周边多个词的范围内,是影响整个词项意义的决定因素,也是“语料库研究迄今最为重要的理论贡献”(Sinclair, 2003: 178)。例如:与 take place 相比, happen 的潜在语义韵往往暗指有坏事发生。

以上4个层面均“以语料库特有的语境共现行等多种频数界面揭示原本在一个(或几个)语例中难以觉察到的词语反复共选的特征,其中既包括横向的共选特征,也包括纵向的共选特征在多大程度上受到横向特征的制约”(Stubbs, 2001: 152)。Sinclair 把 Firth 的语境意义观和词语搭配观拓展到基于批量语境的语义模式研究而且落实到可操作的实施层面。可见这类语义模式涵盖了词汇、语法、语义和语用等多个层面,无论在深度和广度上都超越了 Wittgenstein 和 Firth 的研究。

5 语料库视角下的“意义单位”——短语

既然语料库视角下的意义单位是在批量语境下基于词语的多种共选关系而构建的语义模式,那么它与短语存在何种密切的关联?本节对这两个概念再做进一步探讨。首先看语料库视角下的意义单位概念与传统的有何不同。在传统语法中,单词一直被认为是语言最小的表意单位,但是上述 Sinclair 的词汇语法理论却认为构建语言意义的最基本单位不是单词而是对单词的拓展分析,这就是短语,因为只有短语才能将传统的语法和词汇统一起来。这一短语概念来源于 Sinclair (1991: 109 - 110) 早年所总结的英语词汇的两大选择规则,即开放性原则(open-choice principle)和习语性原则(idiom principle)。前者指在某个语法规则建成后可任意填充一个符合该规则所限定的词性的词(见本章第4节谈到的“空位一填充”模式);后者指词间组合之后生成完全不同于每个词字面意思合并起来的意义,又被称为习语(idiom)。传统语法将习语看作是语法规则的例外,并且因为其数量不多而处于边缘地位。但是语料库研究却发现所谓“习语”其实很具有开放性,而且这种由一定规则制约的开放性习语甚至占据了词汇使用的主导地位。人们经常注意到教科书上很少出现搭配非常固定的习语,就是因为习语在使用中往往变化无常(Sinclair, 2004a: 161)。于是 Sinclair (2004a: 29 - 31) 进一步提出语言词汇使用的两大倾向性:一是词汇的术语的趋势(terminological tendency),即词具有一个直接指向外部世界物体的固定意义,如果要指代这个物体就只能选择这个词,如 atom (原子),不过这是少数。二是短语倾向(phraseological tendency),即词必须组合起来才能产生意义,这是大多数。词语在语境中反复呈现的共选模式(pattern of co-selection among words)直接关联着意义,因为每一个单词都有可能处于以下几种语义状态之一:

(1) 单个词不能直接表达该单位的意义,例如: bear on 中任何独立词都无法表达出“与……有关”这种意思。

(2) 有的词可以表达该单位的意义,有的不能,如: to beat someone up (痛打某人)中的 beat 能而 up 却不能。

(3) 有的词仍然可以表达一般情况下的意义,如: the rain beats down (下倾盆大雨)中 beats 虽不能对整个短语的意义有深刻的影响,但至少会有微小的影响,因为它的字典意义就是“在一段时间内多次地或连续地用力打击”。



可见语言在使用中不仅受语法的制约，同时也受词汇的制约。词的选择受制于它的周边词，又或者说它与周边的词存在着多个层面和多种意义的关系。正是这些关系决定了词或词组在语境中的真正意义。例如：fortune 在 make a fortune 中表示“经济财富”；在 cost a fortune 中却表示“本不该支付的一大笔钱”；而在 had a fortune 中又表示“好运气”。语言使用的这种“短语倾向”引导人们将构建语言意义的最基本单位落实到词项 (lexical item) 而不是单词 (word)，后来被 Sinclair 称作“拓展式意义单位” (extended unit of meaning) (Sinclair, 2004a: 24)，再后来又被称作“意义转移单位” (meaning shift unit) (Sinclair, 2007a)。

此外，即便以往也有关注短语的研究，但是“传统的短语研究大多偏重形式固定的多词序列现象” (Granger 和 Paquot, 2008: 28)，而且“往往由某个理论驱动，采用自上而下的理性分析过程，对短语的内部结构、语义特征等语言学性质进行分类” (卫乃兴, 2009: 41)。但是，语料库语言学在运用现代计算机技术处理巨量语料的过程中却发现了远超越于上述短语概念的大量多词单位现象，并且使这种现象变得可观察、可计量甚至可预测，由此形成了以频数驱动和量化分布为特色的，即基于频数信息产生的大量语料实例进行自下而上的词语使用模式概括的语料库短语学研究。其最重要的发现是：“语言的使用具有短语倾向性”，“人们在多数情况下是一次性选择一串词来表达意义”，“意义从词语的特定组合中产生” (Sinclair, 2004a: 29, 148)。所以语言的基本表意单位应该是短语而非孤立的单个词，短语应该回归到语言描述和意义探究的核心地位 (Sinclair, 2008a: xvi; Teubert, 2011: 6)。尤其是 Sinclair (2008b: 407) 的论断“短语，完全是短语，除此之外还是短语 (the phrase, the whole phrase, nothing but the phrase)”更是将短语视同于意义单位。Sinclair 指出“意义单位的构建和成型是基于持续的反复出现的词语共选模式”，而且意义单位的“边界体现为一个共选模式的淡出以及下一个共选模式的涌现” (2008b: 410)。可见 Sinclair 对短语，也即意义单位的探究基础与 Firth 提出的基于语境中典型的、反复出现的以及可重复观察到的完整话语事实来研究词语在语篇中形成的语境模式 (见本章第 3 节) 十分吻合。下面以 Sinclair (2004a) 的一个经典范例来展示如何将 naked eye (肉眼) 放在短语的视角下构建其典型的意义模式。

it is not really visible to the naked eye .
cannot always be perceived by the naked eye a
ven though nothing is visible to the naked eye .
the opening is not visible to the naked eye .
hotoaging changes are visible to the naked eye .
ittle rooftop house. Viewed with the naked eye .
is visible with the naked eye .
utlets. These could be seen with the naked eye f
human ovum is barely visible to the naked eye .
small. It can easily be seen by the naked eye .
is large enough to be seen by the naked eye .
the surveyor's map. Invisible to the naked eye b
to iris borer. Invisible to the naked eye .
plants that you can see with the naked eye j
is almost invisible to the naked eye .
by many besides those visible to the naked eye .
i-rich Parsons could not see with the naked eye ;

图 1-1 the naked eye 部分语境共现行

Concordance	Concordance Plot	File View	Clusters
Total No. of Collocate Types: 203; Total No. of Collocate Tokens: 56			
Rank	Freq	Stat	Collocate
1	68	8.09980	to
2	46	6.11075	visible
3	45	5.85584	with
4	25	4.30223	be
5	14	3.27544	seen
6	13	3.25613	invisible
7	12	2.90458	see
8	10	2.58000	can
9	11	2.41080	is
10	8	2.38300	by
11	7	2.27946	not
12	9	1.99857	it
13	6	1.85603	are
14	4	1.70926	look
15	4	1.70926	cannot

图 1-2 the naked eye 左 1 至左 4 位置的最高频搭配词

图 1-1 是 Sinclair (2004a: 31-34, 40-47) 从 2 亿多词次的大型英语语料库 Bank of English 中提取所有的 (共 141 例) 以 the naked eye 为核心词的检索行局部截图^①。图 1-2 是核心词左边 4 个词跨度内 (L4 ~ L1)^② 的部分高频搭配词截图。其中有显著性统计学意义 (即 T 值 ≥ 2) 的搭配词有 to、visible、with、be、seen、invisible、see、can、is、by 和 not, 按其语法词性可分别归类为介词 (如: to、with、by), “是” 动词 (如: be、is)、形容词 (如: visible, invisible) 和动词 (如: seen、see) 等。统计显示仅在核心词左一 (即 L1) 位置的词就有 90% 以上是介词, 故形成 “形容词/动词 + 介词 + 核心词” 的类联接模式。聚焦其中的两类实义词 (动词和形容词) 则发现仅 visible、seen、invisible 和 see 四个词就占 141 例的 60% 以上, 余下的还有 look、view、perceive、spot、detect、recognize 等, 显示核心词具有与 “视觉类” 语义特征的词共选的语义偏好模式。它与先前的类联接整合后进一步形成 “视觉类词 + 介词 + 核心词” 词汇语法模式。再进一步关注核心词前后 4 至 5 个词范围内的语义内容, 更发现有 85% 以上含有 “不容易, 有困难” 的意思, 因为在上述视觉类词的周边总有诸如 can not always、not really、barely、small、little、faint (est)、weak、difficult 等词相伴, 由此总结出核心词的语义韵 (也可在广义上称为语用模式), 这就是: 当人们使用 the naked eye 时, 往往用于表示 “仅用肉眼不易看清某物”。虽然在每个具体语境中的措辞和语法形式各有不同, 但大批量语例事实已证明它们所传递的语用功能均具这一明显的倾向性。可见 the naked eye 的意义绝不仅仅是 “肉眼” 那么简单, 它的典型意义模式来自于反复相伴的词语所组成的短语。

6 本章小结

综合以上的理论阐述和实例演示, 我们似可理清支撑语料库短语研究的理论沿革。首先是 Wittgenstein 的 “意义即使用” 语言哲学观将语言意义的研究落实到语言的实际使用之中; 接着是 Firth 的 “意义语境论” 将语言实际使用中的各种语境作为研究语言意义的基础; 而 Sinclair 的词汇语法理论和意义单位概念则进一步将语言意义的研究聚焦于大批量具体语境中的短语意义单位模式上。他依靠大型语料库提取批量语言实例并从这些实例的前后语境中探讨周边词与核心词之间在词汇、语法、语义和语用等多个层面的共选与共存关系, 由此构建起语言 (具体落实到语篇) 的意义单位模式。语料库语言学的短语理论挑战了传统的短语、小句、句子等语法概念。它一方面沿用了传统语法中的一些基本概念, 如词性、语法类别等, 另一方面却更加注重语言在现实语境中使用的形式、意义和功能的三者共享。因而 “比以往任何语言意义描述方式都更加强有力” (Sinclair, 2004a: 133)。

① 因为在该语料库中提取的全部 (共 151 例) naked eye 的左 1 位置有超过 95% 是搭配词 the, 故将这三个词视为基本不变的核心词。

② 在本书中, L4 中的 L 表示核心词左边的第四个词, 若 R4 则表示右边的第四个词, 以此类推, 全书同。



第2章 语料库视角下多种短语的 形态、意义与功能^①

1 引言

本章侧重论述在语料库语言学视角下多种短语形态的研究。首先简述短语研究的历史沿革,接着阐述语料库语言学学者对短语的界定及研究方法,然后重点介述语料库的多种短语形态并聚焦其中的六种,即:意义迁移单位、搭配框架、组织框架、模式语法、语义序列和“机切”语块。通过阐述每种短语形态的形式特征、表意功能和研究价值,凸显语料库语言学的短语理念、研究方法以及在构建语言意义单位,尤其是揭示语篇意义方面的独特视角。

2 短语研究的历史沿革

短语又称多词单位,可泛指语言中各式各样的词语组合形式,对这一词语组合单位有各式各样的称号,包括源自心理语言学视角的“程式语”(formulaic language, 见 Wray, 2002; Wray and Perkins, 2000)、来自功能视角的“词汇短语”(lexical phrase, 见 Nattinger and DeCarrico, 1992)以及基于频数驱动的“意义拓展单位”(extended unit of meaning, 见 Sinclair, 1996a)、“词块”(lexical bundle, 见 Biber and Conrad, 1999)和“N 字语块”(n-gram, 见 Stubbs, 2004)等。

国外对短语的研究始于 20 世纪初。当时的英国学者 Palmer (1933) 系统收集了大批英语习语和短语动词并发表于《第二期英语词语搭配理论中期报告》(Interim Report on English Collocations)。其中的“通用词汇表”汇集了 3 879 个各种类型的词语搭配并在一年后增至 5 749 个。可见当时的短语研究关注的是“词语搭配”,即“由两个或两个以上的词组成的,必须作为一个整体来学习的,而不能从其组成部分任意拼凑起意义的词组”(Palmer, 1933: 扉页)。研究者从语法角度将词语搭配分为两大类:一类是由两种开放类词(指实义词)组成的搭配,如:“动词+名词”;另一类是由封闭类词(指语法功能词)与开放类词组成的搭配,如:“介词+名词”。到 20 世纪中期,苏联及东欧国家语言学者如 Amasova 和 Vinogradov 的短语研究更是将短语学奠定为语言学的一门分支学科。他们将语言中那些形式固定的多字语序列(例如:习语、惯用语、动词搭配、复合词等)按其形式上的固定程度和语义上的显性程度分列为一个连续体(参见 Cowie, 2002: 71)。英国学者 Cowie (1981) 继承了这一连续体模式,将短语单位从纯粹的习语(例如: figurative idioms)渐次过渡到词语的自由组合(free combination),力图通过制定语言标准来区别各类固定性程度不同的短语单位,结果是那些意义不能从各个组合词的表面意思合成的多词语被视为短语的核心形式;而那些仅受到某种语法或语义限制的词语搭配则只能算

^① 本章作者为何安平,基于其在《当代语言学》2013 年第 1 期和《外语研究》2009 年第 5 期发表的文章改写而成,林维燕博士增补了第 4.3 节。

作短语的外围形式。具体又制定出合成性的强弱（如：receive a lesson 比 by and large 强）、理据性的强弱（如：pretty girl 比 pretty village 强）、透明性的强弱（如：talk poetics 比 talk turkey 强）、比喻性的强弱（如：kick the bucket 比 from head to toe 强）等一系列区别标准，以此分出了自由搭配、有限搭配、修辞性成语、纯成语、程式语等多种短语单位的子类。以上这些短语研究虽然大多只关注形式固定的短语现象，然而它们对界定短语及划分短语种类的语法语义视角却为日后的语言多词语现象探究打下了基础并提供了参照。

3 语料库语言学视角的短语界定

语料库语言学界虽然迄今对短语尚未有统一定义，但是在描述短语时多谈到它们的以下特征：

（1）具有某种固定性，是“约定俗成而非依赖语法规则生成”（Altenberg, 1998: 101）；具有预制件或半预制件式的特点（Skandera, 2007: v）。

（2）是具有频数显著性的某个词与一个或多个语言成分的共现，并且可作为语句中的一个语义单位（Gries, 2008: 6）。

（3）具有多层共选性，即“将词、句法、语义和社会应用融于一身”（Ellis, 2008: 5）；“包容了一切人们偏好使用的语言序列以及所谓的固定习语”（Hunston, 2002: 138）；“包括了所有描写语言使用者在表达意义时的词语共选倾向”（Cheng, et al., 2009: 237）。

以上界定涵盖了语料库视角下短语的“词语共选”“持续复现”和“意义单位”三大基本属性（Sinclair, 2008b: 408, 410），凸显短语的核心理念是“共选”，是语篇中反复出现的词与词之间或词与语法之间构建的语境意义模式（contextual patterning, Tognini-Bonelli, 2001: 4）。Gries（2008: 3）曾为广义的语言共选现象及具体的短语识别归纳过六个考察指标：

（1）看短语内部成分的性质（单纯由词还是由词和语法构成）。

（2）看短语内部成分的数量（有两个还是多个成分）。

（3）看短语的使用频数。

（4）看短语内部成分的间隔程度（中间可否插入其他词或者句子）。

（5）看短语的可变性程度（可否有同一词根的词形变化或语法变化，如：主动语态变被动语态）。

（6）看短语的语义整体性、非整全性或可预测性。

可见，该视角下的短语界定特点是它较之以前的和其他学派的定义增加了词项共选的频数分布因素（即具有统计学意义上的显著性）、语义语用因素（即具有构建语义单位的基础性）和形态变化因素（即具有词语连贯的、间隔的或排序调换等多样性）。

4 语料库的多形态短语研究

以上阐述的语料库视角下短语概念的宽泛内涵决定了其形态的多样化。Warren（2008a: 3）曾从词语关联的角度将短语形态大致分为三类：意义迁移单位、搭配框架和语篇组织框架。Granger 和 Paquot（2008: 29）从词语合成形式的角度将诸如框架、搭配框架、类联接以及大致具有组构性的、反复出现的短语都纳入短语形态。Römer（2009:



143) 认为但凡词和语法结构的组合模式都是短语。Sinclair (2008b: 410) 还指出“除了最简单的、毗邻式的短语单位之外 (如: a lot of^①), 短语单位还可能呈现为多种复杂的形态。其中包括重合式 (如: what 在 you know what 与 what I mean 中的重合)、非连续式 (如: a ? of, ? 表示填充词, 下文同) 和包孕式 (如: do you know what I mean 包含 do you know 和 what I mean); 还可以是仅有统计学意义的独立单位 (如: lexical bundles)”。可见语料库视角下短语单位的包容性、多样性和延伸性。以下分别述评语料库不同于传统研究的 6 种短语形态。

4.1 意义迁移单位

意义迁移单位 (meaning shift unit, 简称 MSU) 是最简单的短语形态 (Sinclair, 2008b: 410)。其核心概念是词与词的频繁共选形成了一个语义并非等于各个独立词本意相加的短语单位。MSU 基本等同于 Sinclair (1996a, 2004) 早年提出的词项和意义拓展单位, 三者都是由五个词语共选范畴组成, 即①必须具备的、形式不变的核心范畴 (obligatory invariant core)。②必须具备的语义韵 (obligatory semantic prosody)。③可被选的语义偏好 (optional semantic preference)。④可被选的类联接 (optional patterns of colligation) 和⑤可被选的词汇搭配 (optional patterns of collocation)。具体见第一章第 5 节有关 the naked eye 的例子。但是 MSU 更强调构成这种单位的主要原则——即词语的共选——能够导致表面意义的转移 (Sinclair, 2007a)。例如, 近年 Cheng 等 (2009) 使用语料库检索软件 ConcGram^② 发现了 MSU 的多种变体。例如, 他们在从 5 百万词次的 BNC 语料库提取的 259 例 ROLE 和 PLAY 的语境共现行 (部分如图 2-1 所示) 中以频数高低为序归纳出 7 种变化模式^③:

Sample concordance lines for 'play/role'

1. that there was a need for a public authority to play a role in securing access to and observa
2. Industrial Development Board (IDB) hope will play a key role in financial regeneration of
3. Palestinian people and stated that it should play a full role in a UN conference to negoti
4. 1. What is market research, and why does it play an important role in the marketing funct
5. state enterprises have often been expected to play an exemplary role, because of concern fc
6. the equity provider or venture capitalist will play the most critical role in ensuring that
7. government actions and personalities that PPBs play a much more significant role in publiciz
8. biotechnology will have an increasing role to play in environmental quality during the next
9. Select Committees will have an important role to play in further developing the presentation c
10. financial support to the caring role that women play, both in terms of looking after the chil
11. about the role major companies should play in the community as a whole. He is chair
12. some clue to the role that dietary fibre can play in the prevention and control of adult-c

图 2-1 ConcGram 提取 ROLE 和 PLAY 搭配的部分语例 (Cheng 等, 2009: 242)

① 括号内的例子为本章作者加注, 下同。

② 此为 Greaves C. 设计的一款专门研究短语的软件, 详见 <http://www.benjamins.nl/jbpc/catalogs/concgrammanual>。

③ 模式中的词如 PLAY 是词目 (lemma) 的标记, 内含 play, plays, played, playing 等词形, “?” 表示一个填充词, “/” 表示 “或者”, “_” 表示 “从……到……”, 全书同。

- (1) PLAY ?? ROLE (如例 2 至例 5)
- (2) PLAY ? ROLE (如例 1)
- (3) PLAY ??? /????? ROLE (如例 6 和例 7)
- (4) PLAY ???????? ROLE (未在图 2-1 出现)
- (5) ROLE PLAY (未在图 2-1 出现)
- (6) ROLE ? PLAY (如例 8 和例 9)
- (7) ROLE ?? _ ?????? PLAY (如例 10 至例 12)

模式(1)至(4)作为 ROLE 和 PLAY 共选的主流倾向(占 259 例的 59%)体现为该短语内部两个成分之间多存在有距离不等的间隔,其中又以模式(2)(即可嵌入两个搭配词)为最常见的共选模式。模式(5)至(7)则是两个成分的排序置换,其中又以模式(7)最不稳定(即偏离主流模式最远)。模式(1)至(4)的 ROLE 和 PLAY 分隔使人们得以观察其嵌入词的语义偏好,图 2-1 中例 2 至例 7 的镶嵌词 key、full、important、critical、significant 均含有“关键”“重要”等的强势语义,并形成该短语的语义韵:即人们在使用 ROLE 和 PLAY 搭配时往往表达“以某种有分量、有意义的态度投入或参与某种活动或某个社团”(Cheng, et al., 2009: 247)。这一意义要比以往简单地把 PLAY/ROLE 理解为“扮演角色”或“发挥作用”更富有内涵。该研究表明 MSU 也具有非连续性或排序有变的词语共现模式的变体,这些变体“既有自身独特的‘局部语法’形式(Sinclair, 2007b: 5),又共享该模式的语义和语用功能(Sinclair, 2007b: 8, 转引自 Cheng, 2012)”。MSU 的优势在于它将 Firth 当年提出的搭配的词汇层面和语法层面进一步深入分析到语义层面和语用层面,即搭配词语的语义偏好和弥漫在整个短语使用语境中的语义韵,这后者正是“迄今为止语料库语言学最重要的贡献之一”(Sinclair, 2003: 178)。

4.2 搭配框架

搭配框架(collocational framework)较早由 Renouf 和 Sinclair (1991: 128)界定为由两个非毗邻的“语法词”共同搭建一个框架而让其他词汇有选择地填入其中。例如:a ? of 和 in ? with。之后又有研究者(如:Eeg-Olofsson and Altenberg, 1994; Marco, 2000)将搭配框架分为“语法词 ? 语法词”(如: as ? as)、“实义词 ? 语法词”(如: many ? of)、“实义词 ? 实义词”(如: wait ? minute)等多种搭配序列,而且搭配框架中的填充词也可从一个扩展至两个或多个(如: a ? ? of → a large number of)。此外,还有 Philip (2008: 97-99)把搭配框架归属于短语构架(phraseological skeleton)范畴,内含搭配框架、词汇语法框架(lexicogrammatical frame)和半预制包词组(semi-prepackaged phrase)三种。她关注短语架构的固定性和填充词的可变性,认为搭配框架是由语法词构成一个不变的模式,其中间的可变部分可以是任意语义的词语;而词汇语法框架和半预制包词组则除了大致固定框架外,其间的可变的的部分多是同义词。Stefanowitsch 和 Grices (2008: 933)则把搭配框架、语法形式(grammar pattern)、类联接和搭配结构(collostruction)统称为结构敏感搭配(structure-sensitive collocation),认为它们都是和语法结构有关的短语。Gries (2008: 20)还把搭配框架和类联接视为三大类短语之一,其介乎于连续的 n 元短语(n 的上限为 5,如: a very large number of)和连续的词性标记(如: Adj. + N.、N. + N.、N. + P. + N. 等)之间,并且提出了搭配结构分析(collostruction analysis)的概念。