



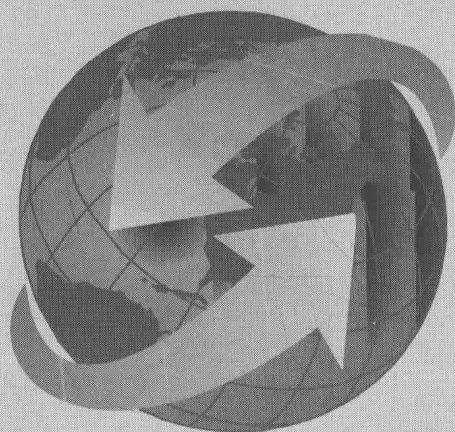
探索与反思

——外国语言文学热点问题聚焦

张生珍 ◎ 主编

Exploration and Reflection:
Focusing on the Hot Issues of
Foreign Language and Literature

中国社会科学出版社



探索与反思

——外国语言文学热点问题聚焦



中国社会科学出版社

图书在版编目(CIP)数据

探索与反思:外国语言文学热点问题聚焦 / 张生珍主编. —北京: 中国社会科学出版社, 2016. 3

ISBN 978 - 7 - 5161 - 7791 - 4

I. ①探… II. ①张… III. ①语言学 - 研究 - 外国②外国文学 - 文学研究
IV. ①H0②I106

中国版本图书馆 CIP 数据核字(2016)第 051396 号

出版人 赵剑英
责任编辑 任 明
特约编辑 李晓丽
责任校对 刘 娟
责任印制 何 艳

出 版 **中国社会科学出版社**
社 址 北京鼓楼西大街甲 158 号
邮 编 100720
网 址 <http://www.csspw.cn>
发 行 部 010 - 84083685
门 市 部 010 - 84029450
经 销 新华书店及其他书店

印刷装订 北京市兴怀印刷厂
版 次 2016 年 3 月第 1 版
印 次 2016 年 3 月第 1 次印刷

开 本 710 × 1000 1/16
印 张 18
插 页 2
字 数 305 千字
定 价 68.00 元

凡购买中国社会科学出版社图书, 如有质量问题请与本社营销中心联系调换
电话: 010 - 84083683
版权所有 侵权必究

江苏师范大学品牌专业培育项目资助

目 录

型式语法研究综述	于 涛 (1)
同形变异成语仿用的构式语法分析	毕 冉 (16)
族裔历史背景下恶作剧者的蜕变与偏离 ——评韦尔奇的《愚弄克劳人》	丁文莉 (24)
19 世纪俄国哲理诗人笔下的自然 ——以巴拉丁斯基和丘特切夫为例	董春春 (34)
二语心理词库形音建构问题与因应策略	韩玉强 (44)
《费城,我来了!》中的叙事策略分析	朱荣华 (57)
当代美国妇女剧作家概观	张生珍 司林榕 (65)
基于语料库的当代散文翻译规范研究	张继光 (79)
西方文学中的旅行书写	张 林 (96)
《第二十二条军规》中的悖论研究	邱 蓓 (107)
《谁害怕弗吉尼亚·沃尔夫?》的伦理意旨析评 ...	邹惠玲 李渊苑 (115)
伊丽莎白·毕晓普旅行叙事中的政治意涵	顾晓辉 (127)
寻找心灵的“香格里拉” ——试论《拯救溺水鱼》的魔幻现实主义元素	林晓雯 (141)
日本现实主义儿童文学的基石:论坪田让治的“童话小说”	刘 迎 (152)
中国学生俄语词汇教学的语言文化教学模式探究	杨 柳 刘华荣 (164)
英汉超音段特征对比分析与英语语调教学	杨 萌 (173)
两封信——两个童年 ——契诃夫与谢尔巴科娃的短篇小说《万卡》	郑广杰 (182)
北美印第安神话传说中的文化英雄	郑 佳 邹惠玲 (192)

协作话语后的会话流程之中日对比分析	熊红芝 (205)
突显和焦点实现的韵律研究新视角及其对英语语调教学的启示	高 薇 (222)
向心理论——话题与韵律接面研究的新方法	夏志华 (241)
从《苔丝》中译本的译者痕迹看经典文学作品的复译	张乐金 (251)
阶级意识与《金字塔》、《过界仪式》中人物的利益选择	肖 霞 (263)
文学翻译的隐喻之旅	张蓓蓓 (273)
后记	(284)

型式语法研究综述^{*}

于 涛^{**}

摘 要：本文对型式语法与型式识别相关文献进行了梳理和评析，主要涉及到型式研究的缘起与发展、型式语法的优缺点和相关应用研究，以及型式的识别方法。通过上述几个方面的论述，不仅可以帮助语言研究人员和二语学习者增进对于型式的理解，而且可以帮助二语学习者提高型式使用的准确率。

关键词：型式；型式语法；型式识别；研究综述

A Review of Pattern Grammar Studies

Yu Tao

Abstract: This paper aims to summarize and review related literatures on Pattern Grammar and pattern recognition, mainly covering the topics on the origin and development of pattern studies, strengths and weaknesses of Pattern Grammar, pattern-related empirical studies, and pattern recognition methods. Review on the fore-mentioned topics will not only enhance the linguistic scholars and second language learners' understanding of pattern, but also improve the precision of second language learners' use of patterns.

* 本文为2015年国家社科基金项目“基于型式语法及索引行聚类的动词型式自动识别与提取研究”(15BYY095)成果之一。

** 于涛：男，(1975—)，汉，江苏赣榆人，博士，副教授，硕士生导师。

一 引言

型式 (pattern) 这一术语在语言学界中的使用, 最早出现在霍恩比 (Hornby, 1954) 的《英语型式用法指南》一书之中, 距今已有 60 年历史。为了服务于语言教学, 霍恩比 (Hornby, 1954) 以句子结构为出发点, 对英语中的动词、名词和形容词型式进行了归纳总结。后来, 许多学者致力于语言型式的研究, 但由于研究出发点不尽相同, 他们对于“型式”一词的理解各不相同。例如, 语料库驱动研究范式强调以词汇为出发点观察语言型式, 其中辛克莱 (Sinclair, 1991) 以词汇为出发点研究扩展意义单位; 霍伊 (Hoey, 1991) 则基于词汇衔接理论对词汇模式 (Patterns of Lexis) 在语篇衔接中的功能进行了探讨; 汉斯顿和弗朗西斯 (Hunston & Francis, 2000) 则认为由动词、名词和形容词与其补足语成分组合而成的短语单位, 即型式, 如 *V from n into n*, 其中 *V* 表示核心词, *from* 和 *into* 指具体词项, *n* 表示名词短语。

近二十年来, 随着语料库语言学的发展, 型式研究已成为当下研究的重要议题之一。正如舒尔茨和罗默 (Schulze & Romer, 2008: 265) 所言, “很难想象语料库研究人员不对语言中的多词单位或型式感兴趣”。例如, 弗朗西斯 (Francis *et al.*, 1996, 1998) 在 COBUILD 项目基础上, 总结了英语动词、名词及形容词的各种型式, 并在此实践基础之上, 形成了《型式语法》(Hunston & Francis, 2000) 一书, 对型式的概念、识别标准、型式与意义的关系等问题进行了梳理及详尽的解释。因而“型式语法 (被视为是) 基于语料库驱动, 但却是在经过了语言学家的归纳提取之后才获得的一系列研究成果, 是经验主义和理性主义的结合” (陈功 2012: 64)。因而, 对于型式语法及型式识别相关研究进行述评, 有利于增强学界对于型式的理解, 提高学习者在语言使用中语言型式使用的准确率, 并有助于探讨型式在语言研究、辞书编纂、语言教学等领域的重要作用。

二 型式与型式语法

型式语法是第一部运用语料库数据编写的教学语法, 编写的过程中将

语法与词汇作为整体看待,其目的在于揭示传统语法中宾语、补足语等语法范畴的表现形式及其意义(Hunston & Francis, 1998: 62)。

(一) 型式语法

型式语法聚焦于词项及其相关联的型式,而非句法规则;“(型式语法)是将语法描写与词汇描写相结合,重点放在单个词项的使用行为上”(Hunston, 2002: 167);其前提是基于辛克莱(Sinclair)关于“词汇与语法不可分”的论断之上。正如弗朗西斯(Francis, 1995)所澄清的,“在交际过程中,交际双方并非优先选择句法结构,再选取词项来填补句法结构;相反,交际双方在大脑中预先存在所需要传递的概念信息,然后做出交际选择,而概念及交际选择首先要由核心词项来实现。最终,这些核心词项找到合适的符合句法规则的结构来表达意义”。

1. 型式语法与传统语法差异

型式语法与传统意义上的语法存有差异,其中传统语法通常建立在先前语言学理论基础之上,而型式语法完全是基于真实语言数据归纳总结而来,因此,型式语法对传统语法形成了强有力的冲击。这些冲击主要表现在如下方面:(1)词类概念受到挑战;(2)短语及小句等概念受到挑战;(3)句法功能类别受到挑战;(4)词汇语法二分的观念受到挑战;(5)语法语义二分的概念受到挑战(Hunston & Francis, 2000: 259)。因此,韩礼德(Halliday, 2005: 166)对型式语法给予了高度评价,认为“Hunston 和 Francis 所做的一切,非常成功。(型式语法)对语法的研究更加精细,使得词汇和语法的研究紧密结合在一起”。

2. 型式语法的优点

型式语法可能开辟了语法描述,尤其是词汇语法描述的一条新途径,对语法研究具有重要的引领作用(卫乃兴, 2011: 214)。相比英语结构分析,英语型式分析具有以下四点优势:(1)一致性问题得以解决(研究者只需进行少量培训即可);(2)型式分析通常指意义单位分析,因此可以在语言表层展开,从而避免了语言范畴的影响;(3)型式分析相对全面,不仅适用于所有的动词,而且适用于其他词类;(4)型式分析不受先前语言学理论的影响与制约(Hunston & Francis, 2000: 177)。正是因为上述优势,孙海燕(2013: 69)认为,在语料库驱动研究方法的指导下,型式语法挑战了以前的语言理论和范式,对传统的语法范畴提出质

疑,认为词汇和语法、型式和意义之间没有泾渭分明的界限,词汇和句法互相选择,特定的词项往往和特定的句法结构相关,型式与意义的一体性是语料库驱动的语言研究的一个重要维度。

正如上述论点中所言,型式语法是语言学研究的一大突破与创新,第一,摆脱了原有理论框架的藩篱,使用了全新的方式来解释语言现象;第二,型式语法打破了传统语法中词汇语法两分的观点,认为词汇语法不可分,是一个不可分割的整体,因为型式中各元素间既要满足语法结构的需求,又要满足词汇选择的限制,词汇语法之间并非界限分明;第三,型式语法是语言学家们基于大量的语言事实总结而来,改变了先前语言研究中仅依赖少数语言事实的现象,更加符合语言实际,更能反映语言的复杂性,是自下而上研究方法的典型代表;第四,型式语法为自然语言处理打下了坚实的基础,其总结的各类特征适用于语言模型的构建,可以更好地为自然语言处理服务。

3. 型式语法的不足

虽说型式语法是对语言研究的一大革新,但仍存在一些不足。在人工识别型式的过程和其后的理论归纳总结中,汉斯顿和弗朗西斯(Hunston & Francis, 2000: 77)就认为,“总体而言,我们对型式的归纳过于抽象。例如型式 V_n 中的 n ,既可以指名词,也可以指代词。当形容词型式中的组成元素中包括动词时,我们仅将系动词与其他动词区别开来,分类不够细化。从这个角度看,我们没有完全遵循辛克莱(Sinclair)的思想,因为他强调不同词汇形式出现在不同的型式列表当中,以及特定词汇拥有自身的短语学型式特点,部分辛克莱(Sinclair)想区别的东西在型式语法中被概括归纳。虽然我们的归纳总结可能不如传统语法那么宽泛,但仍存在归纳总结过当的迹象”。因此,型式语法并不能用来区别“相同句法结构中的一词多义现象。例如,型式语法中并没有区别 *firing a gun* 和 *firing an employee*,认为这两个短语都是 *fire* 一词的及物动词用法,同属于型式 V_n 。如若需要了解两者意义的差别,读者需要借助搭配信息,考虑哪些东西与枪的意义相关,哪些事物与雇员的意义相关”(Hanks, 2013: 6)。

除了上述学者对型式语法发出不同的声音之外,部分学者也表达了类似的观点。麦克内里等(McEnery et al., 2012: 213)就认为,虽然型式语法强调其研究方法围绕词汇展开,但事实上,在涉及到具体动词的词项

时对节点词实行了词形还原,忽略了其曲折变化形式。此外,金(King, 2005: 35)指出,“型式语法有其必然缺陷。为了尽可能囊括所有型式,许多型式被过度归纳。其实,诸多情况还需进行细致的区分”。正如汉斯顿和弗朗西斯(Hunston & Francis, 2000: 81)所言,“型式应该比当前的描述更加具体;具体化是指应该抛弃型式语法编写中所使用的语法标签”,因为“使用词类标签作为型式编码的简明方式不可能完全揭示短语单位或词块的意义与结构特征”(卫乃兴 2011: 213)。以上观点反映出汉斯顿(Hunston)始终坚持词汇为研究中心的观点,采用词类或语义类标签是一种权宜之计。然而,使用词类特征虽然致使透明度降低,但是型式传递的信息量增加了。如果完全使用词汇对型式进行重组,将会使得原本就复杂多变的语言现象越来越复杂,型式列表越来越长,更加不利于后续学习与研究中的应用。但在当前计算机技术广泛应用于自然语言处理领域的背景下,随着词汇语法界面研究的发展,可以在具体和抽象之间采取折中的做法。

综上所述,型式语法继承了语料库驱动研究的做法,坚持词汇语法不可分,并强调词汇在型式中的重要地位,反对传统语法的做法,摒弃复杂而抽象的句法概念,转而使用简单、通俗易懂的标签来标记型式,符合人类认知习惯,减轻了学习者的认知负担。但型式语法存在归纳过度的现象,在增加覆盖面的同时,却牺牲了语言描写的透明度,即没有反映出语言型式的具体呈现形式。

4. 基于型式语法的相关研究

型式语法在语言研究中的应用主要表现在以下几个方面,(1)词典编纂;(2)语言学研究;(3)自然语言处理。

型式语法在英语语言研究方面的应用主要表现在 COBUILD 项目的系列研究成果上(王勇 2008)。这些成果都是由 Collins 出版。其中词典系列包括:第一版的《柯林斯 COBUILD 英语语言词典》[辛克莱(Sinclair) 1987]、第二版的《柯林斯 COBUILD 英语词典》[辛克莱(Sinclair) 1995]等等;语法系列包括:《柯林斯 COBUILD 英语语法》[辛克莱(Sinclair) 1990]、《柯林斯 COBUILD 语法型式:动词》及《柯林斯 COBUILD 语法型式:名词和形容词》(Francis *et al.*, 1996, 1998)等等,共计三十多种。这些都是以型式语法的思想为指导编写而成的。

型式语法除了可以服务于词典编纂之外,还可以应用于语言学研究之

中,如语体差异研究、语言教学和中介语对比分析。例如,汉斯顿(Hunston, 2002: 181)在《运用语料库考察语言差异》论文集中详细论述了型式在语言教学、语体差异、语言地区差异等方面的应用价值。穆克吉(Mukherjee, 2001)统计了 PROVIDE 一词的四种型式在语料库中的分布情况,并在数据的基础上,讨论了四种不同型式分布的影响因素。威利斯(Willis, 2003)结合实例讨论了型式在语言教学中的功能及其应用价值。此外,国内学者濮建忠(2000)通过对3个较为常用英语动词使用情况的调查,认为在动词的使用过程中还包含一些重要的、关键性的信息,如动词的语法型式和词汇型式,没有为英语教师和学习者所普遍掌握。

型式语法在自然语言处理中的应用主要体现在自动语法检查领域。黄等(Huang *et al.*, 2011)尝试提取母语使用者的型式列表,并将其应用于学习者语言的自动语法检查。陈功(2012)则以型式语法为理论基础,通过链语法这一形式化语法体系,结合中国大学英语四六级和专业英语四八级考试大纲要求学生掌握的动词,对链语法词库进行了补充和调整,然后将动词型式语法形式化,实现了链语法动词词典的重构,使之更好地为学习者自动语法检查服务。

三 型式识别

当前型式的识别方法主要有两种:(1)传统的人工型式识别方法;(2)型式的自动识别。

(一) 型式的人工识别

传统的型式识别通常采用索引行解读的方法,这是因为索引行可以帮助使用者观察到语料库中出现了什么,以及文本中的意义是如何产生的,词汇是如何共现从而形成有意义的型式。将单个索引行视为单独文本,很大程度上能帮助考察节点词语境及语境中词汇型式的频数(Sinclair, 1991: 34)。

无论语料库语言学家,还是词典编纂学家都意识到,传统的索引行解读存在一定的局限性。例如,人工解读索引行,数量会受到很大限制(Hunston, 2010a: 162)。因为,“如果包含特定节点词的索引行数量有限,这种方法非常有效,但如果面对 THAT、STRONG 等常用词汇的索引行时,容易致使存在微小差异的型式淹没在成千上万的索引行当中”

(Church *et al.*, 1991: 116)。

由于存在上述局限性,许多学者在传统索引行解读的基础上,采取了随机抽样的方法。怀恩(Wynne, 2008: 722)就曾指出,“索引行自动抽样并提供典型实例,需要软件执行复杂的算法,从而解释共现语境中出现的型式,并选择典型的语言实例”。事实上,索引行的随机抽样,并非需要复杂的算法,我们可以采取分层抽样或是随机抽样的方法,选取部分索引行进行解读。但无论采取何种抽样原则,都有可能使得一些非常用型式被忽略,有时还会导致总结的型式用法或意义不够完整。辛克莱(Sinclair, 2003: ix)也意识到了同样的问题,认为“在计算机诞生以前,限于观察能力及记忆能力,语言学家们只能对小样本数据进行观察,即使花费了一生的时间也只能收集特定型式有限的实例”。

综上所述,传统索引行解读存在诸多局限性,(1)耗时费力;(2)有些型式被凸显出来,有些型式则被淹没在成千上万的数据之中;(3)随机抽样的方法,会导致部分包含非常用型式的索引行被筛选掉;(4)型式的人工识别结果有时会存在争议或不一致。

(二) 型式的自动识别

语料库语言学和计算语言学界迫切希望实现型式的自动识别,并且认为存在型式自动识别的可能性。汉斯顿和弗朗西斯(Hunston & Francis, 2000: 67)就曾指出,语料库语言学打开了文本自动处理的大门,其中之一便是自动识别特定词汇的型式。

梅森和汉斯顿(Mason & Hunston, 2004)基于现有的型式列表(Sinclair, 1995; Francis *et al.*, 1996),并结合字符串匹配方法对英语动词型式的自动识别进行了尝试。但是,其研究存在以下不足:(1)型式列表不够全面。(2)验证的数据量过小(仅100例包含节点词decide的句子)。(3)对型式内的插入语成分处理结果不令人满意,例如Mehari said yesterday he would decide next week whether the jury...中的next week作为插入语出现在decide和whether之间,并非型式的必要组件。(4)非常用型式(被动语态、倒装、从句修饰语)没有得到很好地解决。

梅森(Mason, 2004)采用基于句法剖析的语料进行英语动词型式的自动识别,但其在系统评测时使用的数据量过小,仅对包含BLEND的56个句子和包含LINK的116个句子进行了尝试,准确率分别为96.40%和

62.90%。虽然上述系统效果不错,但不同词汇间型式自动识别的准确率差异很大,因而很难得到广泛的应用。此外,此研究中使用句法分析器对句子进行浅层标注,中间错误率较高;在系统评测过程中,几个无法识别的实例均是出于句子切分错误导致。

梅森(Mason, 2006: 184—187)对以上两种型式自动识别方法进行了总结并指出了其中的不足,认为“上述两种型式自动识别方法局限于短语关系,预设所有短语是构成句子结构的重要元素。但是,在真实语言中,这种关系非常复杂。例如,真实语言中的插入语成分很多,每个小句都有其内部结构,因此 SPOCA^① 结构很难概括所有的句法结构”。

黄等(Huange *et al.*, 2011)设计了软件 EdIt, 试图通过对母语者文本的句法分析,实现语言型式的提取,然后应用于自动语法检查。其型式的提取过程如下:(1) 对语料库进行词元归并、词性赋码及短语层句法剖析等预处理,试图提取更多的符合型式规律的 N 元组和搭配信息;(2) 依据 MI 值提取语料库中的 N 元组和搭配;(3) 将每个词元、词性码,以及短语所在句子的信息放在倒排索引中,以便提高系统检索速度;(4) 在倒排索引中查询包含给定单词和短语的句子,并根据相应的词性码识别搭配中的词项信息,然后将该句转换为具体词形和词性码的混合串(hybrid Ngram),最后统计其频数信息。从上述方法看来,黄等(Huang *et al.*, 2011)将型式语法中的型式与搭配混为一谈,仍坚持使用搭配的统计手段来识别型式。此外,此研究中仅涉及少量语法现象的识别,并且没有报告该系统的查准率和查全率,真正的效果如何还不得而知。从形式上看来,此方法仍是一种有益的尝试,属于使用简单模式匹配的方法来进行语法检查。

综上,作为型式人工识别的补充手段,上述为数不多的研究采用基于规则和基于统计的方法相结合,在型式自动识别方面做出了诸多有益的尝试。鉴于语言的复杂性和句法剖析的准确率,上述研究的实验结果并不令人满意,这一切的原因归根结底体现在两个方面:第一,对于语言规则不应采取拿来主义、原封不动的加以利用,而应该对规则进行分析和归纳,如可以采取化整为零的方法对型式必要组成元素进行总结。第二,统计方法的选择上存在一定的问题;简单字符串匹配只能帮助发现前人总结的动

^① SPOCA 表示 subject predicate object complementary adjunct。

词型式，而对于非常规型式的识别却并不奏效。因而在未来的型式自动识别研究中，需要将规则的处理和统计方法的选择两个因素同时考虑在内，并找到两者之间的切合点，从而提高型式自动识别的准确率。

四 现有的语言型式网络平台数据库

为了增进语言使用者对语言型式的了解，提高语言型式使用的准确率，许多学者基于语料库数据，搭建了诸多网络平台，提供了大量的语言型式使用实例。现有的平台包括 StringNet, Justvthe-word, phrase in English, Sketch Engine, PDEV, FrameNet 等。

(一) 基于机器处理的语言型式网络数据库

为了辅助语言教学和语言习得，部分网络平台基于语料库数据，采用相关统计手段对原始数据进行了处理，如 StringNet 和 Just-the-word 等。

StringNet 网络数据库

StringNet 将词汇语法结合领域看作其核心目标区域，即坚持词汇语法不可分的观点 (Wible & Tsao, 2010)。在这一点上与型式语法有共通之处。但在操作层面上，仍采用基于频数统计的测量方法，与前人关于搭配的研究方法并无二致。

StringNet 数据来源于 BNC，其目的在于建立词汇语法知识库来揭示词汇型式，主要涉及到英语中构式现象，其中包括大量的混合 N 元组信息。但是基于 N 元组的做法存在一定的问题，汉克斯 (Hanks, 2013: 168) 就指出，基于词汇的 N 元组与直接采用跨距内频数统计的方法几乎没有差异。

No.	Sentences
1	In the new state, fostering was the only way to provide a home for orphaned and otherwise parentless children.
2	(He was later to commission Richard Rodney Bennett to provide a score for his Jazz Calendar based on the old nursery rhyme 'Monday's child')
3	Next wind some tape through the slots and around the core to provide a base for the pickup windings.
4	Conference provided a focus for Telford 1990, when delegates and visitors donated blue 'ears' to raise money for this charity.
5	The interest of the rest of the 'virtuosi' (as they were known) was a more general one in the 'new philosophy' and its aims, for the fact is that the Royal Society provided a focus for a whole movement of thought.
6	For example, a study of the debasement trends of sixth-to seventh-century AD European gold coinage provided a basis for dating the Sutton Hoo ship burial.
7	The same is true for many other periods; for instance, coins provide a date for the deposit of the great Viking hoard from Cuerdale in Lancashire of c AD905, and this in turn provides a date for the associated objects and hence a pivotal fixed chronological point for our understanding of Anglo-Saxon metalwork.
8	The same is true for many other periods; for instance, coins provide a date for the deposit of the great Viking hoard from Cuerdale in Lancashire of c AD905, and this in turn provides a date for the associated objects and hence a pivotal fixed chronological point for our understanding of Anglo-Saxon metalwork.
9	In a sense, therefore, his body provided a cover for the sharpness of his mind, and the cold baldness of his stare a cover for his crusades.
10	He wants to restore it to its original castellated glory to provide a house for wife Annie and two sons and realise a 15-year dream.

图1 StringNet 网络平台数据库中 PROVIDE 的部分索引行示例^①

① <http://www.lexchecker.org/>

JustTheWord 网络数据库

JustTheWord 是英语写作辅助平台,目的是给使用者提供合适的词汇选择。如其网站描述^①, JustTheWord 可以帮助英语学习者验证其词汇选择的正确性,或是提供改进意见。其数据同样基于 BNC 语料库,并将词语组合间搭配信息采用 T 值形式体现出来,并用绿色柱体表示搭配的强度。

JustTheWord 网络平台提供了大量的词语搭配信息,这无疑为语言学习者提供了大量的语言使用实例,在英语教学和英语写作中有较高的应用价值。但是 JustTheWord 的做法仍存在一定的问題,例如其语料选择标准仍沿用前人的搭配测量手段,局限于两词或三词连续或非连续的共现现象。事实上,有时在搭配统计上具有显著意义的两词之间在意义上不能构成一个整体。此外,JustTheWord 对常用词提供了大量的词语组合列表,分类过于详细,不利于学习者在短时间内查询。

jtw Help How PROVIDE	
	combinations alternatives for usage alternatives for new usage
A07 - 000972	ering is that introspective reports, while often providing helpful qualitative information, can never be reliable
A07 - 000280	ade Association for insurance companies which can provide information about its members.
A0V - 000492	rn, Help the Aged and Neighbourhood Energy Action provides information on heating and can be contacted on Freephon
A0V - 000715	This is a booklet providing information on all types of breaks, both at home and a
A0V - 000832	This free DSS telephone line can provide information about benefits to disabled people and their
A0V - 000863	(Tel: 081-675 6557/8/9/0) Supports families, provides information about all forms of dementia, runs local gro
A10 - 001291	It provides information on methods to improve the efficiency and ef
A10 - 001360	aves that general practitioners should be able to provide information about their services, and that there should
A15 - 001360	tells you how to decipher the drag listings which provide information such as: drag name, run hours, number of row
A19 - 002104	active, constant awareness of network conditions, providing accurate and timely information to permit the network
A26 - 00524	is large enough - covering 11,179 executives - to provide detailed information on particular groups.
A50 - 00497	d said they claimed a right under European law to provide information on services available legally in another EC
A55 - 00082	res member states to set rules on mandatory bids, providing information to shareholders and treating them equally.
A56 - 00220	er industry prospectus is published next month it provides comprehensive information on all 10 authorities.
A60 - 000895	A function of marketing research is to provide information that will help opportunities to be identifie
A60 - 000624	number of publications and organisations that can provide market information about different countries.
A60 - 000777	Secondary data also provide information relating to a wide variety of different aspe
A60 - 000781	Also, they provide information about political, financial, foreign exchange
A62 - 000158	formation on issues, but similar to the press for providing information on leaders, and close behind the press and
A62 - 001020	ugh they rated television as much more useful for providing information about issues, they did not rate televisio
A62 - 001058	(though much less useful than news broadcasts for providing information).

图2 JustTheWord 网络平台上 PROVIDE 的部分型式列表

(二) 基于人工处理的网络数据库

有些数据库的建设完全依靠手工完成,如 PDEV 和 FrameNet,分别由帕特里克·汉克斯(Patrick Hanks)和查尔斯·菲尔摩(Charles Fillmore)团队建设,并在进行不断的更新和完善。这类数据库通常从认知视角出发研究语言型式,因而数据库的主要目的是为了辅助语言学研究。

PDEV 网络数据库

语料库型式分析(CPA, Corpus Pattern Analysis)聚焦于在用法上具

① <http://www.just-the-word.com/>.

有横组合关系的词汇型式的原型。其平台建设步骤如下：第一，依据语义范畴，将包含相同横组合关系的索引行分成若干组；第二，将词汇的不同义项与型式建立关联（Hanks, 2004: 87）。关于数据的收集，首先从语料库中抽取包含目标词汇的索引行，先预览并对词汇的行为有大致地了解，然后随机抽样 200 到 1000 个索引行进行细致分析。在这个过程中，索引行按照意义相似或句法结构相似进行重新排序。最后对每一组目标词汇索引行，按照配价或论元信息进行语义标注（Hanks, 2004: 91—92）。

语料库型式分析所得出的数据，由于是人工归纳而来，具有准确率高，信息比较全面等优势，但耗时耗力，效率低下。截至目前^①，PDEV 数据库仅包含 1258 个动词的型式，并且其中绝大部分为非常用动词。但其部分做法与型式语法中的型式具有共通之处。

Sample size: all=1133 / 0

No. % Pattern / Implicature

1	<1% [[Wind Vapour {Stuff = Dust}]] blow [NO OBJ] [[Adv[Direction]]] [[Wind Vapour]] is an air current moving, at high or low speed (as the case may be), in a particular [[Direction]] and carrying [[Vapour {Stuff = Dust}]]	conc. explor.
2	<1% pr [[Wind = Storm]] blow [NO OBJ] (up) [[Wind = Storm]] suddenly occurs	conc. explor.
3	<1% pr [[Wind = Storm]] blow [[Self]] (out) Old-fashioned: [[Wind = Storm]] comes to an end after a period of time	conc. explor.
4	<1% [[Wind {Event = Explosion}]] blow [[Physical Object Stuff Vapour]] [[Adv[Direction]]] [[Wind {Event = Explosion}]] causes [[Physical Object Stuff Vapour]] to move [Adv[Direction]]	conc. explor.
5	<1% pr [[Wind]] blow [[Physical Object Stuff]] (away) [[Wind]] removes [[Physical Object Stuff]] from [[Location]] on a current of air	conc. explor.
6	<1% pr [[Event = Explosion]] blow [[Physical Object Physical Object Part]] (away) [[Event = Explosion]] destroys [[Physical Object Physical Object Part]] suddenly by force	conc. explor.
7	<1% pr [[Physical Object Stuff]] blow [NO OBJ] [[Adv[Direction]]] (away) [[Physical Object Stuff]] is removed from [[Location]] by the wind or by an explosion and moves [Adv[Direction]]	conc. explor.
8	<1% idiom pr [[Eventuality]] blow (colwrebs) (away) [[Eventuality]] has a cognitively refreshing effect on [[Human]] or [[Human Group]]	conc. explor.
9	<1% idiom pr [[Anything]] blow [[Human]] (away) Slang: [[Human]] is very favourably impressed by [[Anything]]	conc. explor.
10	<1% pr [[Human 1]] blow [[Human 2]] (away) Slang: [[Human 1]] kills [[Human 2]], typically with a [[Firearm]]	conc. explor.

图3 PDEV 网络平台中动词 BLOW 的部分型式实例^②

FrameNet 网络数据库

汉斯顿（Hunston, 2010b）指出，框架网络（FrameNet）与型式语法共同关注自然产出语言中形式和意义的一致性（形式和意义关系紧密）。框架网络关注具有相同意义的一组词，而型式语法聚焦于高频共现的搭配、类联接，进而识别核心词项意义之间的相似之处。两者相同点体现在以下四个方面：（1）关注语义一致的一组词；（2）关注实际产出语言中的词汇行为；（3）关注语义角色的相似性；（4）语义角色到型式的

^① 2015 年 1 月 7 日采集的结果。