

型式语法的产生、特点及其应用价值^{* 1}

陈功 梁茂成

(对外经济贸易大学 北京 100029; 北京外国语大学 北京 100089)

提 要:型式语法是 Hunston 和 Francis(2000) 在 COBUILD 项目语法研究成果的基础上,对全新的语料库描写语法的深入理论探讨。本文将着重梳理型式语法的产生背景、理论渊源以及突出特点。最后,本文探讨型式语法在英语教学和自然语言处理方面的重要应用价值。

关键词:型式语法; 语料库描写语法; 英语教学; 自然语言处理

中图分类号: H314

文献标识码: A

文章编号: 1000-0100(2017)01-0017-8

DOI 编码: 10.16263/j.cnki.23-1071/h.2017.01.004

The Origin, Features and Applications of Pattern Grammar

Chen Gong Liang Mao-cheng

(University of International Business and Economics, Beijing 100029, China;
Beijing Foreign Studies University, Beijing 100089, China)

Pattern Grammar is an in-depth theoretical investigation into a brand-new corpus descriptive grammar by Hunston and Francis (2000) on the basis of the COBUILD project research. This paper mainly focuses on the background, the theoretical origin and the features of Pattern Grammar. It also explores the important applications of Pattern Grammar in English teaching and natural language processing.

Key words: Pattern Grammar; corpus descriptive grammar; English teaching; natural language processing

1 引言

型式语法(Pattern Grammar)是在 COBUILD 语法研究成果的基础上产生的,是对 COBUILD 成果的全面总结。作为第一部语料库驱动的英语描写语法,它颠覆以往语言学家看待语言、获取语言知识的传统,给语言研究提供一个全新的视角。为了让更多的研究者了解型式语法在理论和实践方面的贡献及价值,本文从型式语法产生的背景、理论渊源、内容、特点和应用价值等方面对其进行详细的梳理和探讨。

2 型式语法的产生

2.1 产生背景

作为一种全新视角的语言描写理论,型式语法的产生和发展绝不是偶然的。下文将从3个方面详细探讨型式语法产生的时代背景。

2.1.1 词汇—语法的二元划分问题及其转向

在现代语言学研究中,各个时期占据主导地位的语法研究都体现出重结构、轻词汇的倾向。无论是索绪尔的“组合关系、联想关系”,布龙菲尔德的“直接成分分析法”,还是乔姆斯基的“句法研究”,“都将语法视为独立于词汇的一个系统”(Hunston, Francis 2000: 7),“结构第一位,词汇第二位”(Owen 1993: 168)。由于结构主义和乔姆斯基的强大影响,词汇—语法二元划分根深蒂固,影响到理论语言学界和应用语言学界的大

* 本文系教育部人文社科基金项目“面向中国学生的英语动词语法错误自动检查研究”(14YJC740006)、教育部人文社科重点研究基地重大项目“英语学习者语误自动检测系统的研制”(11JJD740011)和对外经济贸易大学科研课题“中国学生英语书面语动词形式错误自动检查研究”(12QD16)的阶段性成果,并得到对外经济贸易大学商务英语与跨文化研究中心的资助。

批学者。

20世纪70年代以后,语言学研究开始越来越多地关注词汇,并将词汇看成语言的核心(Hunston, Francis 2000: 7)。到90年代,尽管词汇—语法二元划分的影响依然深刻,但学者对词汇研究的观点已经逐渐达成共识。Owen指出,“尽管传统主流的力量仍就强大,但在过去的十几年中,研究方向已经发生了明显的转变”(Owen 1993: 169),即“从对结构事实的解释向对词汇事实的解释转变的趋势”(Hudson 1990: 3)。这是语法研究的第一次转向。

如果语法研究第一次转向的目的是为了纠正当时“重语法、轻词汇”的倾向,寻求语法研究和词汇研究之间的平衡,那么,第二次转向则开始真正深入看待语法—词汇之间的关系。对此,不少研究者意识到,语法和词汇并不是非此即彼的关系,两者是不可分割的(Sinclair 1991, Halliday 1976, Francis 1993)。“每个单词都有自己的语法”(Lewis 1993: 142)组成语言的单位是“语法化了的词汇,而不是词汇化了的语法”(同上: vi)。

20世纪80年代之后,关于多词单位或短语的研究开始出现,如Pawley和Syder(1983)、Melcuk(1995)等。尽管短语研究的时间长,范围广,研究者们却“始终没有将短语看作是一种正常的语言组织原则”(Hunston, Francis 2000: 21),相反,“只有当语法不能解释一些语言现象时,我们才会转而寻求短语的帮助”(同上: 254)。短语所受到的空前重视与其在语言研究中的实际地位极其不相符。正是在这样的需求下,3代语料库语言学家Firth, Sinclair, Francis, Hunston等人在搭配、短语学等领域不断探索,推陈出新,逐渐形成对语言规律的独特认识。

2.12 经验主义的回归

“在20世纪50年代以前,现代语言学的传统……都注重语言事实,提倡经验主义”(Hunston 2006: 11),即“从实践出发,以行为数据为研究对象,对言语进行科学、系统的描写,并从中归纳出语言理论”(梁茂成 2010: 90)。然而,随着乔姆斯基在语言研究中主导地位的确立,经验主义的传统逐渐被主流遗弃,语言研究进入主要通过内省和直觉来解释语言的理性主义时期。

理性主义强调研究者的分析解释能力,本身并无过错。然而,乔姆斯基对经验主义的排斥以及对真实语言的忽视却逐渐引来质疑声音。一方面,“理性主义的研究所得出的语言规则似乎只能适用于一种子语言,而不能推广到该子语言之

外的其他语言现象,具有很大的局限性”(冯志伟 2006: 12)。另一方面,语言学家的直觉是否如他们所相信的那样可靠,是否可以说基于直觉的语法“更多的是对直觉的研究,而非对语言的研究”(Sinclair 1991: 4)。

在对理性主义不断质疑的同时,语料库建设却如火如荼地展开。由于语料库研究尊重语言事实,注重对真实语言的描述,经验主义的传统开始回归。受这一趋势的影响,“大多数学者认为:语言学的研究必须以语言事实作为根据,必须详尽地、大量地占有材料,才有可能在理论上得出比较可靠的结论”(冯志伟 2006: 12-13)。

经验主义的回归是另外一个语言学研究时代——“后描写主义”时代的开始,型式语法是这个时代的典型代表。

2.13 研究方法的创新

新方法的出现不仅可以对以往观点或理论进行证实或证伪,还能激发思想,催生新的理论。Hunston和Francis指出,“不管在任何研究领域,新的方法不仅带来新的数据,更能带来不同类型的数据,从而进一步启发出新的理论”(Hunston, Francis 2000: 250)。型式语法的出现就是得益于语料库研究方法在研究思路和研究工具上的新进展。

前计算机时代的语言学研究者在搜集、整理和加工语言材料时,基本上完全靠手工进行,而且获得的语言材料极其有限。“大规模机器可读语料库的出现和使用,把语言学家从艰苦繁重的手工劳动中解放出来,使语言学家可以集中精力来研究那些更加重要的问题,这对于促进语言学研究的现代化具有不可估量的作用。”(冯志伟 2010: 34)

另外,语料库检索软件的出现也使我们能够更加方便地观察语言,获取语言知识。Hunston指出“如果没有语料库检索软件的帮助,语料库充其量只是一大批实际使用语言的集合。而检索软件能够以不同的方式对语料进行重新排列组合,可以使进行多方面的观察”(Hunston 2002/2006: 3)。它的出现使语料库研究方法成为语言研究中的“显微镜和望远镜”(Stubbs 1996: 231),扩大语言学研究的视野,“使他们看得更远,看得更细,从而使他们能够发现更多的语言现象,挖掘出更多的语言事实”(冯志伟 2006: 14-15)。和20世纪美国的结构主义语言学相比,语料库语言学同样重视描写和分析,但是在方法上却有质的飞跃。

2.2 理论渊源

型式语法的理论渊源可以追溯到伦敦学派的奠基人 J. R. Firth(1957) 在“Modes of Meaning”一文中对搭配(collocation) 及其意义的研究。不过, 对型式语法产生直接影响的则主要是 Sinclair 的语言思想以及 Francis 对前人思想和方法的继承与发展。

2.21 Sinclair 的语言思想

Sinclair 主张语法和词汇不可分割, 并从根本上与传统语法不同的思想, 对型式语法的产生和发展起到直接的推动作用, 主要表现如下:

(1) 语料库驱动语言描写对语料的要求: Sinclair(1991) 对语料库语料的关注构成语料库驱动语言描写的基础。Hunston 和 Francis(2000: 15) 总结 Sinclair 对语料 5 个方面的要求: 第一, 语料要真实, 依靠语言学家的内省或直觉获得的数据是不可靠的; 第二, 不能根据研究需要选择语料, 传统的语言学研究通常会使用例句对语言点进行阐述, 这一做法等于“按图索骥”, 是不可取的; 第三, 语料数量要大, Sinclair 的一个观点就是 bigger is always better(同上: 16), 语料库的容量越大, 我们就越有可能对低频词项进行更加准确的描述; 第四, 语料的组织要系统, 语料的组织决定文本以何种方式呈现, 从而最终决定研究者对语料的观察; 第五, 不要根据现有理论对语料进行标注, 除了对标注的准确性有所顾虑之外, Sinclair 最担心标注依据的理论对语料库研究创新的束缚将会使发现新的语言现象或规律的可能性大大降低。

(2) 意义与结构(sense and structure) 的联系: Sinclair(1991) 通过对多义动词及其所在结构进行考察发现, 意义与结构有千丝万缕的联系。具体来说, 如果一个词有多个义项, 并且该词可以用于多个型式, 那么每一个型式总会相对应地与其中一个义项频繁共现, 这样, 我们反过来可以通过型式来预测该词的意义。因此, Sinclair 明确表示, “意义和句法之间的联系表现出了很强的共现倾向”(Sinclair 1991: 65)。“在正常的语篇当中, 结构确实能够有效地区分意义”(Hunston, Francis 2000: 21), “形式和意义根本上是不可分的”(Sinclair 1991: 7)。

(3) 成语原则: Sinclair(1991) 的主要贡献之一是提出观察或解释语言的两个不同模型: 开放原则和成语原则。开放原则, 也称为“空位—填充”模型, 主要为词汇的选择提供基本的语法约束, 保证语言的合语法性, 是传统研究看待或描写语言的方式。然而, Sinclair 同时也发现, 更多的词汇选择是开放原则无法解释的, “语言使用者

掌握了大量的半预制的短语结构, 虽然每个短语看上去都可以分成不同的部分, 但却构成使用者的唯一选择”(Sinclair 1991: 110), 这就是成语原则。它表明语言具有短语倾向, 即词语不是孤立存在, 而是通过它们的共选关系而获得意义。大部分文本都可用成语原则来解释, 具体来说, “80% 以上的词项共现具有共选关系, 只有 20% 的词项是通过独立的语法聚合选择出现的”(Sinclair 2000)。

(4) 意义单位(units of meaning): 开放原则作为语言的语法模型, 将单词的词汇语义层面和句法组织割裂开来(Barnbrook 2007), 句子结构空位对词汇的选择基本上是基于单词的词性, 完全不考虑语义。乔姆斯基转换生成语法的短语结构成分就是很好的例子。

Sentence \rightarrow NP + VP
 NP \rightarrow T + N + Number
 Number \rightarrow { sing , pl}
 VP \rightarrow Verb + NP
 Verb \rightarrow Aux + V
 Aux \rightarrow Tense
 T \rightarrow the
 N \rightarrow man , ball , etc.
 V \rightarrow hit , took , etc.

(Malmkjær 1991: 483)

尽管这一系列规则可以保证生成的句子符合语法, 但所有词汇和语义方面的问题都未加以考虑。

与开放原则不同, 成语原则的基本单位不是词, 而是词所重复出现的短语。在对搭配、短语的研究中, Sinclair 发现短语的意义“分布于整个短语的所有词汇中”(Sinclair 1987: 110), 如 have a bath(Sinclair 1991: 113), 其中, have 义项众多, 通过与短语中其他词的共选才获得该短语作为一个整体唯一的意义。这很好地解释语言中的意义是如何产生的, 而作为意义单位的短语也被认为是语料库描述语法的核心单位。

(5) 搭配框架: Sinclair 和 Renouf 提出“搭配框架”(collocational framework) 的概念, 如 a + ? + of 等。他们的研究证明, 搭配框架和其中的搭配词高度共选, 而且这些搭配词属于特定的意义范畴。如 an + ? + of 框架中的名词可分为 7 个不同的意义组(Sinclair , Renouf 1991: 136 - 137)。

这一研究对型式语法具有两大积极的影响。第一, 对于同一种语言现象的描述可以有很多种不同的方法。该发现令 Hunston 和 Francis 喜忧

参半,喜的是“对英语的描述原来可以有多种视角”;忧的是多视角下的“语言描述将永无止境”(Hunston, Francis 2000: 26)。因此,尽管“搭配框架”的真正描写还不无问题,但型式语法的方向却清晰很多。“一部语料库驱动语法需要多方面、多视角地对英语进行描述”,而且“让这部语法做到非常完备、全面几乎是不可能的”(同上: 27)。第二,对搭配框架提取方法的思考。Sinclair和Renouf(1991)的研究完全依靠计算机自动提取,该方法省时省力,但却容易遗漏其它一些重要的语法信息。基于此,Hunston和Francis提出两种完全不同却又互补的方法:一是完全依靠计算机软件根据共现频率提取搭配;二是采取更有解释力的人工查找的办法(Hunston, Francis 2000: 27)。尽管型式语法的最终提取方法并不仅仅是两者的折中,但方法的思考却对型式语法产生至关重要的影响。

2.22 Francis 对前人思想的继承和发展

Francis是Sinclair语法思想的继承者和发扬者,主要在以下4个方面发展了Sinclair的思想,也逐渐成就了型式语法。

(1) 意义和型式: Francis的“意义和型式”(同上: 29)虽然在措辞上只是略不同于Sinclair的“意义和结构”(Sinclair 1991: 53),却是对Sinclair思想的深化,主要表现为3个方面:第一, Sinclair的“结构”概念较大,一个结构可以是短语,可以是小句,也可以是整个句子。而根据Sinclair(1991)考察的例子,即多义动词的义项与其所在结构的关系,这个“结构”显然应该是一个局部概念,称为“短语结构”或“型式”更为合理。第二,“型式”这一概念可以更好地将Sinclair的“意义和结构”与成语原则下的“意义单位”结合起来。第三, Sinclair(1991)仅探讨多义词的不同义项可以在各自典型的结构中得以区分,而Francis则着重考察另一个方面,即“特定的型式能够选择具有特定意义的词”(Hunston, Francis 2000: 29)。

(2) 词汇语法和短语学: 在承认语法和词汇不可分割的基础上, Francis(1993, 1995)首先肯定词汇的重要地位,并指出“句法是由词汇驱动的,词汇在交际中具有优先性”,或许这就是为何Hunston和Francis(2000)将其专著*Pattern Grammar*的副标题定为“语料库驱动的英语词汇语法”的原因。

Francis指出,词汇语法“认识到横组合型式(syntagmatic patterning)以及受其影响的词汇选择在语言架构中的重要性”(Francis 1993: 142 -

143)。换句话说,“句法结构和词项之间具有共选关系,不能将其分开考察”(Hunston, Francis 2000: 30)。具体表现为:一方面,特定的句法结构通常与特定的词项共现;另一方面,词项通常只出现在有限的结构中。词汇语法的目标就是关注处在“短语渐变区”中具有不同程度共选倾向的短语,并对其进行忠实的描述(Francis 1993: 143, 146)。

(3) 对语料库驱动描写语法研究方法的探索: 对Sinclair和Renouf(1991)研究方法的思考给Francis非常大的启示,Francis提出“从环境^①到词项,再到环境”的观察路径(Francis 1993: 146)。在COBUILD语法型式系列(Francis et al. 1996, 1998)的编撰过程中,上述方法也得到进一步的细化和完善。在大多数情况下,观察是从词项到环境;需要补充语法信息时,再从环境回到词项,目的是为了尽量完整地列出各个型式分别所选择的词项。此外,还需要对由此产生的词项列表进行义项的分组归类,这个任务主要由语法学家人工完成。

值得一提的是,在整个语法系列的编撰过程中,除语料库驱动的方法外,语言学家的人工处理也占据很大比重。Francis和Sinclair明确表示,COBUILD的工作离不开语言学家的直觉(Francis, Sinclair 1994: 191)。COBUILD提倡使用真实语料,忠实语言事实,但这并不等于对直觉的压制或抛弃。相反,语料库在辅助人们的语言直觉和内省的同时,离不开研究者本人的语言分析和判断能力。

(4) 对语料库驱动语法的进一步探讨: 对语料库驱动语法和传统语法的不同,Francis总结4点内容:第一,语例真实,COBUILD项目语法研究的观察语料和语例均取自“Bank of English”语料库;第二,旨在构建一部非对比语法,即只展示词项本身的用法,不与其他词项或其他范式的用法做对比;第三,定位为数据驱动语法,通过观察大规模语料来获得对语言的认识,是语料库时代“基于观察的语法”,而不是“基于直觉的语法”(Aarts 1991: 46);第四,旨在构建一部真正意义上的词汇语法(Francis 1993: 137 - 139)。

3 型式语法的特点

3.1 型式语法的编码

和传统语法标注相比,型式语法的编码设计有较为大胆的改革,基本上不采用语法功能标签,如vt.(及物动词)、V+O(动词+宾语)等作为语法编码的做法。取而代之的是一串字母加具体单词或相关说明文字的组合,例如, *V n as n* 型式的

基本要素包括 3 个部分。为了便于解释,笔者将 3 个部分分别称做核心词项、补足成分和其他成分。其中,核心词项(V)是构成型式最为关键的要素,是任何型式都不可缺的成分;补足成分和其他成分则根据语言的不同使用规律附加于核心词项,或为具体单词(如 V so/not, I hope so.) ,或为成分说明(如 V n to-inf, I need you to be there.) ,或两者同时用于型式的构成(如 V n at n, He shot a glance at her.)。根据需要将这些编码进行组合便可构成核心词项的不同型式。

型式语法的编码主要遵循灵活性、透明性和一致性 3 个原则(Hunston, Francis 2000: 33)。首先,编码的灵活性是指语法标注不受短语长度的限制。由于编码只使用表层的分类范畴和具体单词,因此不论型式有多长,都可以使用有限的编码将其表示出来。在例句 It seems to me that I had done everything I could 中,动词 seem 的型式可以通过简单的编码表达为 it V to n that。第二,透明性在一定程度上反映语法编码的易懂性,即语法编码不考虑句法方面的抽象概念,只反映词项的表层使用以方便学习者的理解和记忆。第三,一致性是指编码要保持统一,进行表层描述。在传统的词典中,make her happy 往往会被标注为 verb + object + adjective,其中不仅包括功能标签 object,还包括词类标签 verb,这是语法编码的混合使用。

尽管型式语法的编码也存在多种类型,如词类标签(n, v, adj 等)、具体词项(on, into, of)、语言使用说明(amount, with quote 等),但笔者通过文献学习发现,这样的编码设计有其特定的考虑。首先,型式语法的编撰者关注的只是对语言表层的描述,不论采用何种编码,只要不违背表层描写原则并能方便学习者的理解和记忆,都可为我所用。其次,型式语法编码体现对传统语言学研究成果的尊重和认同,如词类划分标签,但并不是简单的继承,而是结合型式语法的编码需求,对其内涵或外延进行适当扩展。例如,在型式语法中,传统语言学名词词类标签 n 并不单纯指一个词,还可以指整个名词词组。正如 Hunston 和 Francis 所言,“型式语法的编码体系是语言使用概括与具体的折中,尽量能够做到在描述一棵树的同时,脑中还装着整片森林”(同上: 249)。

3.2 型式语法的理论主张

型式语法的两个理论主张是词汇和语法不可分与型式和意义相联系(同上: 250)。

3.2.1 词汇和语法不可分

在型式语法中,词汇和语法互不可分。也就

是说,核心词项与其他词项或其所在结构存在共选关系,相互之间具有形式上的依赖性,而且词项和结构之间并不是简单的一一对应关系,本文以动词为例进行阐述。

首先,词项对型式有选择,而且一个动词词项可以对应多个不同的型式。动词的行为是复杂的,在大多数情况下,一个型式根本无法满足描写的需要。例如,通过观察动词 explain 与周围其他词项的共现情况,我们可以归纳出该动词所依赖的 9 个不同型式(如 V n, V wh, V about n 等)。

其次,型式对词项也有选择,而且一个型式可以选择若干个不同的动词。也就是说,这些动词的用法在特定的语境中相同,可以出现在同一个型式中。例如,对于型式 V over n,有许多动词符合要求,其中包括 argue, compromise, disagree, quarrel 等。

3.2.2 型式和意义相联系

型式和意义相联系是指核心词项与其他词项或其所在结构共同构建一个稳定的意义,该意义与型式不可分割。例如:

表₁ 型式 V as n 的动词词项列表

act	end	originate	serve
begin	end up	parade	shape up
come	enlist	pass	sign up
come across	figure	pose	stand
come out	finish	practice	stand down
come over	finish up	put up	start
continue	freelance	qualify	start off
count	function	rank	start out
double	go down	rate	start down
double up	masquerade	remain	train
dress up	moonlight	resign	volunteer
emerge	operate	run	work

通过对表₁中动词型式 V as n 的动词列表进行简单观察便可发现,共享同一型式 V as n 的这组动词并不是任意出现的,其中很多词项在意义上非常接近,例如, enlist 和 volunteer, resign, stand down 和 step down, begin 和 start, end, end up, finish 和 finish up 等。根据意义的相关性我们可以把动词划分成不同的意义组,根据 Francis 等(1996: 159-162),这 48 个动词可以被划分成 6 个意义组(见表₂给出的其中 3 个意义组)。

观察表₂ 可以发现,这些动词之所以能按照意义进行分组是因为对于共享同一型式的众多词项来说,虽然它们常用的词典意义不尽相同,但由于同一型式所造成的语言环境,或多或少产生意义上的交集。简单来说,共享型式的词项同时也共享其词意的某些方面(Hunston, Francis 2000: 256)。例如,动词 qualify, train 是在它们共同的使用环境 V as n 中体现出与“入职培训”相关的意义。我们很难说是意义的构建产生相应的型式,还是型式的构建产生相应的意义,总之,型式和意义在很大程度上是不可分的。

表₂ 型式 V as n 中的动词意义分类(部分)

意义组 1	The work group: 与工作有关的动词
	1. 入职培训相关: qualify, train
	2. 求职相关: enlist, put up, run, sign up, stand, volunteer
	3. 从事某一职业相关: freelance, moonlight, practise, serve, shape up, work
4. 离职相关: resign, stand down, step down	
意义组 2	The function group: 与具有某一角色或职能有关的动词
	act, double, double up, figure, function, operate, serve
意义组 3	The begin and end group: 与开始或结束相关的动词
	1. 开始: begin, originate, start, start off, start out
	2. 持续: continue, remain
	3. 结束: end, end up, finish, finish up

4 型式语法的应用价值

不可否认,型式语法是语料库理论建设的优秀成果,是英语语法理论的一个突破。除了在 COBUILD 词典中做语法标注之外,型式语法还在英语教学领域和自然语言处理领域有不可估量的应用价值,这和型式语法一开始的定位——教学语法和描写语法——不无关系。

4.1 英语教学领域

型式语法在教学中的应用价值主要有 4 个方面:

第一,型式语法有助于提高学习者的语言意识。大规模语料观察表明,型式语法重在对词项表层行为的关注,词项的行为具有高度的重复性和规律性,让学习者认识到这一点非常重要。另外,型式语法的编码在设计之初就已经考虑到要尽可能方便学习者的接受和理解,因此,其简洁透明的特性在某种程度上可以帮助学习者更加清晰地把握语言规律,并掌握词项的典型用法。

第二,型式语法有助于提高学习者语言产出的准确性和流利性。型式语法是一部词的语法,提供具体词项的详细信息,学习者在了解词项及其型式的同时,就可获得具体词项的用法,有助于提高学习者语言产出的准确性。例如,通过学习动词 discourage 及其型式 V n from n/-ing,便可知 discourage 的典型用法是 discourage + 名词词组 + 介词 from,其中介词 from 不可随意替换。而一旦学习者掌握该词及其词项,其语言产出也会相应地是整个词项的短语。

第三,型式语法有助于学习者词汇的积累。型式语法不仅归纳出所有词项的所有型式,还对同一型式下的词项进行意义分组。换句话说,有一部分词项不仅用法相同,意义也相似。例如,在学习动词 discourage 及其型式 V n from n/-ing 时,教师可以根据学习者的词汇水平列出与 discourage 同一意义组的词项,如 stop, keep, prevent, prohibit, restrain, restrict, 等等。这样,学习者既学到 discourage 的用法,也了解到其他类似的词汇及用法。此外,由于这些词项之间存在意义上的联系,更方便学习者记忆。

表₃ 用不同的型式和词项表达同一意义

型式	例句
V n	She <u>liked</u> the idea.
V to n	She <u>warmed</u> to the idea.
V n n	She <u>thought</u> the idea a cracker.
V n adj	She <u>considered</u> the idea brilliant.
V n as adj	She <u>regarded</u> the idea as brilliant.

第四,型式语法有助于学习者语言产出的灵活性和多样性。共享同一型式的词项通常存在意义上的共享或意义交集。但是,某一方面意义的表达却并非只能通过一种型式来实现(Hunston et al. 1997: 215)。表₃ 显示的就是用不同的型式和词汇来表达“某人喜欢一个想法或观点”。通过对词项和型式的学习,学习者逐渐会用灵活多样的词汇来表达思想。

4.2 自然语言处理领域

《型式语法》(Hunston, Francis 2000) 推出之后,很快受到自然语言处理界的关注。次年,《计算语言学》(Computational Linguistics) 第二期就刊登出 FrameNet 项目组前计算词典学家 Christopher Johnson 的书评。Johnson 指出,型式语法的语言思想对于计算语言学和理论语言学的重要性不可低

估(Johnson 2001: 138)。同时,他还建议,如果计算语言学家想获得型式描述的相关资源,并将其应用到自然语言处理当中,还须要参考 Francis 等(1996, 1998)的型式语法系列。不过,从计算语言学的视角 Johnson 指出,尽管型式语法有很多值得学习之处,但其本身既不属于计算语言学,也不是形式语法体系,因此,希望今后能看到一部更加形式化、更加系统的型式语法(Johnson 2001)。

近年来,随着自然语言处理界对语料库的不断重视,型式语法也开始受到愈来愈多的关注。型式语法逐渐体现出在自然语言处理领域的价值。正如 Johnson 所期待的,已经有研究者开始着手构建基于型式语法的语法形式体系,或者将型式语法的理念用于自然语言处理程序的设计当中。Otero 和 López 认为,型式语法提出的“语法—词汇不可分”和“意义—结构相联系”的观点是当前其他形式语法体系所或缺的。词汇—语法分离、意义—结构分离的后果是无法处理搭配、习语等占据语言很大部分的固定或半固定短语结构,而形式化体系在发展过程中一直都忽略这个问题。鉴于此,他们结合型式语法和依存语法的思想,建构一部语法形式体系——Dep Pattern。Huang 等(2011)则将型式语法的方法和理念运用到学习者语法检查系统“EdIt”的设计中,最终测试结果显示,EdIt 的查错结果要好于微软的 ESL Assistant。

陈功(2012a)以型式语法为理论基础,通过链语法形式化语法体系对动词型式(如 V as n 等)进行形式化,并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,构建一个面向中国学生的英语书面语动词形式错误检查系统。该研究测试结果显示,加入动词型式后,链语法分析器的查错性能和句法分析能力得到很大的提高。对包含动词形式错误检查的召回率比原系统提高 4.5 个百分点,准确率提高 15.7 个百分点;对本族者正确句子分析的准确率提高 12.2 个百分点。该研究表明,型式语法不仅能够较好地运用于自然语言处理领域的形式模型,还可以更好地服务于中国学生书面语形式错误检查系统的构建。在此基础上,陈功(2012b)进一步指出,基于型式语法理论构建的错误检查系统不仅能帮助学习者检查动词形式错误,还能够借助型式语法的理念,设计动词型式参考句库和动词型式自动反馈工具,为学习者提供更好的语言学习平台。对于学习者来说,不论是错误检查,还是诊断、纠错,其最主要的目的是提高语言技能,避免在同一个语言点上重复出错。这也是

我们今后努力的方向。

这些最新的研究成果表明,自然语言处理领域的研究者已经逐渐认识到型式语法在语言描写方面的优势,并正在将这一语料库语言学的理论成果积极运用于自然语言处理的实践中。今后,型式语法的应用研究非常值得期待。

5 结束语

本文对型式语法的产生背景、理论渊源、主要内容、特点、问题及其在英语教学和自然语言处理领域的应用价值进行了详细的梳理。不难看出,型式语法已经被越来越多的研究者所接受、认可。而近年来,型式语法在自然语言处理领域的应用也表明,语料库语言学的思想和理论还有更大的发展前景和应用价值。

注释

①“环境”指“型式”(pattern),即词项出现的环境。Francis 指出,环境/型式和词项本来是不可分的,之所以作区分,为了从不同的视角观察语言型式。

参考文献

- 陈功. 中国学习者英语书面语动词形式错误自动检查: 一项基于链语法的研究[D]. 北京外国语大学博士学位论文, 2012a.
- 陈功. 自然语言处理技术在语言教学和研究中的应用初探——以重构后的链语法分析器为例[J]. 中国外语教育, 2012b(4).
- 冯志伟. 自然语言处理的形式模型[M]. 合肥: 中国科学技术大学出版社, 2010.
- 梁茂成. 理性主义、经验主义与语料库语言学[J]. 中国外语, 2010(4).
- Aarts, J. Intuition-based and Observation-based Grammars [A]. In: Aijmer, K., Altenberg, B. (Eds.), *English Corpus Linguistics* [C]. Harlow: Longman, 1991.
- Barnbrook, G. Sinclair on Collocation [J]. *International Journal of Corpus Linguistics*, 2007(2).
- Firth, J. R. Modes of Meaning [A]. In: Firth, J. R. (Ed.), *Papers in Linguistics* [C]. London: Oxford University Press, 1957.
- Francis, G. A Corpus-driven Approach to Grammar: Principles, Methods and Examples [A]. In: Baker, M., Francis, G., Tognini-Bonelli, E. (Eds.), *Text and Technology: In Honour of John Sinclair* [C]. Amsterdam: John Benjamins, 1993.
- Francis, G. Corpus-driven Grammar and Its Relevance to the

- Learning of English in a Cross-cultural Situation [A]. In: Pakir, A. (Ed.), *English in Education: Multicultural Perspectives* [C]. Singapore: Unipress, 1995.
- Francis, G., Hunston, S., Manning, E. *Collins COBUILD Grammar Patterns 1: Verbs* [M]. London: Harper Collins, 1996.
- Francis, G., Hunston, S., Manning, E. *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives* [M]. London: Harper Collins, 1998.
- Francis, G., Sinclair, J. M. I Bet He Drinks Carling Black Label: A Riposte to Owen on Corpus Grammar [J]. *Applied Linguistics*, 1994(2).
- Halliday, M. A. K. *System and Function in Language: Selected Papers* [M]. London: Oxford University Press, 1976.
- Huang, C., Chen, M., Huang, S., Chang, J. S. EdIt: A Broad-coverage Grammar Checker Using Pattern Grammar [OL]. <http://www.aclweb.org/anthology/P11-4005>, 2011.
- Hudson, R. *English Word Grammar* [M]. Oxford: Blackwell, 1990.
- Hunston, S. *Corpora in Applied Linguistics* [M]. Beijing: WPC, 2002/2006.
- Hunston, S., Francis, G. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English* [M]. Amsterdam: John Benjamins, 2000.
- Hunston, S., Francis, G., Manning, E. Grammar and Vocabulary: Showing the Connections [J]. *ELT Journal*, 1997(3).
- Johnson, C. Review of "Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English" [J]. *Computational Linguistics*, 2001(2).
- Lewis, M. *The Lexical Approach: The State of ELT and a Way Forward* [M]. Hove: LTP, 1993.
- Malmkjær, K. *The Linguistics Encyclopedia* [M]. London: Routledge, 1991.
- Melcuk, I. Phrasemes in Language and Phraseology in Linguistics [A]. In: Everaert, M., van der Linden, E., Schenk, A., Schreuder, R. (Eds.), *Idioms: Structural and Psychological Perspectives* [C]. Hillsdale: Lawrence Erlbaum Associates, 1995.
- Otero, P. G., López, I. G. A Grammatical Formalism Based on Patterns of Part of Speech Tags [J]. *International Journal of Corpus Linguistics*, 2011(1).
- Owen, C. Corpus-based Grammar and the Heineken Effect: Lexico-grammatical Description for Language Learners [J]. *Applied Linguistics*, 1993(2).
- Pawley, A., Syder, F. H. Two Puzzles for Linguistic Theory: Native-like Selection and Native-like Fluency [A]. In: Richards, J. C., Schmidt, R. W. (Eds.), *Language and Communication* [C]. London: Longman, 1983.
- Sinclair, J. M. Grammar in the Dictionary [A]. In: Sinclair, J. M. (Ed.), *Looking Up: An Account of the COBUILD Project in Lexical Computing* [C]. London: Collins ELT, 1987.
- Sinclair, J. M. *Corpus, Concordance, Collocation* [M]. Oxford: Oxford University Press, 1991.
- Sinclair, J. M. Lexical Grammar [J]. *Naujoji Metodologija*, 2000(24).
- Sinclair, J. M., Renouf, A. Collocational Frameworks in English [A]. In: Aijmer, K., Altenberg, B. (Eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik* [C]. London: Longman, 1991.
- Stubbs, M. *Text and Corpus Analysis* [M]. London: Blackwell, 1996.

定稿日期: 2016-04-25

【责任编辑 王松鹤】