

数据驱动学习与专业英语写作创新

蒋 婷, 杨 梅

(重庆大学 外国语学院, 重庆 400030)

摘要:基于语料库的数据驱动学习已然为外语教学的变革提供了创新工具,而目前的专业英语写作教学仍然固守着以教师和教科书为中心的、自上而下的传统外语教学模式。本文针对专业英语写作教学中存在的问题,从词汇和语篇等方面,探讨如何将数据驱动学习方式创新地应用于专业英语写作教学,确立以学生为中心、充分开发学习者创作潜能,并由此提高其语言应用能力的英语写作教学模式。

关键词:数据驱动学习;专业英语写作;词汇;语篇

中图分类号:H087 **文献标识码:**A **文章编号:**1008-5831(2006)06-0115-06

一、引言

大学专业英语写作课在专业英语教学中是不可或缺的教学科目,它以培养学生写作能力为教学目标,以传授英语写作技巧为教学依托,以提高学生英语语言的表达水平为最终目的。但就目前的实际情况看,专业英语写作教学尚不尽人意,存在不少问题,主要是:(1)教学形式比较单一,教学方法缺乏创新。表现为写作培训的安排很少,有限的时间大多放在应试讲评上,缺乏对语域等全局层次的特征传授(黄大网,2003:37)。(2)教材编排过于陈旧。受教材篇幅的限制,传统的英语写作课本所包含的练习和例文非常有限,且呈现方式受文本线性特征约束,不可能将教学内容定量、定时地呈现给学生,很难高效率地培养学生遣词造句和布局谋篇的能力,故常常难以达到预期的训练效果(陈敏哲,2005:98)。(3)缺乏对学生学习主动性的调动和激发。目前的写作课堂上,大部分教师鉴于课时限制,倾向于过多依靠教师的语言叙述,即便是涉及学生练习,也常为简单的摹仿写作。

总体来看,现时的专业英语写作教学仍然固守以教师和教科书为中心、自上而下的传统外语教学模式,授课方式和内容集中在抽象的讲述和简单的摹仿写作练习,这与它作为写作教学高级阶段的目标定位之间存在不少差距,导致长期以来不能收到理想的实践绩效。因此,如何给专业英语写作教学注入新鲜血液,如何赋予它崭新的内容和机能,是我们亟待解决的问题。

数据驱动学习(data-driven learning,简称DDL),是一种新的基于语料库的外语学习方法,它的基本原理是指引导学生利用海量的语料库数据,观察、概括和归纳语言使用现象,自我发现语法规则、意义表达及语用特征,这是一种以学生为中心的发现式学习。所以,把数据驱动学习应用于专业英语写作教学中,不

收稿日期:2006-10-10

基金项目:重庆大学外国语学院2006年度科研教改资助项目

作者简介:蒋婷(1974-),重庆大学外国语学院讲师,硕士,主要从事专门用途英语、语料库语言学研究。

仅将为教学活动提供丰富的语言材料,而且必将引发教学活动中语言学习者的学习方式、学习内容以及学习材料等方面的革命性变化。这是对以教师和教科书为中心的传统外语教学模式的挑战,同时也为有效地解决目前专业英语写作外语教学中的各种问题提供了一个崭新的思路。

二、数据驱动学习的理论基础及其特征

数据驱动学习是20世纪90年代初由Tim Johns提出的,其基本理念在于鼓励学生积极主动地从真实语料中去发现语言事实(act as 'language detectives')(Hunston, 2002: 170)。事实上,数据驱动学习并不是直接教授某一语言特征,而是向学习者提供语言事实,引导他们据此提出假设并最终得出结论。可以说,整个过程其实是围绕学习者而设计的一系列活动,旨在提高学习者对有关知识点的敏感度(consciousness-raising)。Jane Willis进一步将其归结为以任务为中心的学习模式(如下图)。

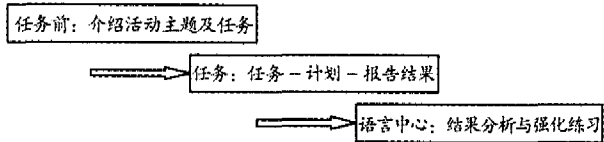


图 数据驱动学习的模式

Tim Johns将数据驱动语言学习过程分为三个阶段:(1)提出需要解决的问题(Identify)。这些问题可以是教师根据教学大纲而提出,也可以是在学习过程中学习者自身产生的疑问。(2)材料分类(Classify)。在传统的教学中,分类通常是通过查询词典,参考书籍,或者由教师通过内省的方法编造例句等途径实现。这样所得的材料极为有限,而且有些不是真实的语言材料。在数据驱动语言学习中,语言材料是通过语料库检索工具对语料库中海量的真实语料进行检索而获得的。通常检索和分类需要在课堂教学之前完成。(3)归纳总结(Generalize)。这是数据驱动学习中最重要的一個阶段,即在教师的引导下,学习者阅读、理解和归纳分类后的语言学习材料。在学习者独立地按照自身的认知方式对各例句进行理解的同时,他们也不断地进行自我策略调整。

由此可见,数据驱动语言学习与传统的教学模式存在很大区别。其特征可以归纳为四个方面(甄超, 2005: 20)。

第一,以学生的自主学习为中心。传统的教学模式是与学习者的自主性(learner autonomy)不相适

应的。数据驱动学习强调学生的自主学习,完全以学生为中心,发挥其个性特点。学习者不是被动的语言知识的接受者,而是一个主动的语言知识的发现者和研究者,并根据需要自我调整学习策略。此外,不同于普通意义上的“自学”,自主学习要求学生加强相互间的合作,共同探索和发现语言规则和使用特征。

第二,以真实、丰富的海量语料为主要语言输入。目前的外语教学仍然以教师的语言输出和教材为学生的主要语言信息来源。由于多数教师的外语“语言能力”尚未达到外语本族语者的水平,加之教材编写中的各种人为因素,学生接触的真实语言有限,非真实的语言输入则有可能导致各种语言错误的产生。而基于语料库的数据驱动学习提供给学生的质量、海量的语言数据都来源于真实的交际活动。简言之,数据驱动学习能够为学生营造真实的语言环境,提高他们的语言直觉,锻炼他们处理语言变体(variation)的能力,帮助他们掌握地道的语言。

第三,强调探索和发现的学习过程。建构主义学习理论认为,语言知识的习得不是一个从教师到学生的简单过程,而是一个由学生自己发现和探索的过程。数据驱动学习尤其重视这一学习过程。它基于语料库提供给学生大量真实的语言数据,引导学生自己观察学习过程,根据自身的需求去经历、探索、发现语言知识。这样一来,学生获得的语言知识会更地道、印象更深、体系性更强、储存更持久。

第四,主张自下而上、归纳式的学习。目前的外语教学多体现自上而下、演绎式的特征。比如,由于缺乏足够的语言材料和应有的语言分析知识,学生对语法规则的理解不够深入,靠死记硬背强化记忆的较多,使用时则生搬硬套。此外,教师讲解的语法规则同真实语言往往不尽吻合。在数据驱动学习方式中,学生首先接触到的不是规定性的语法规则,而是大量的真实语言数据,经过他们的独立观察,从中归纳出语法规则。当然,在这种“自下而上、归纳式”的学习方式中,学生也需要借助语法描写的手段来分析观察到的语言现象,并总结出规则。

可以说,数据驱动学习代表着教学模式的转变,揭示了新的学习方法和理念。它要求学习者作为一个积极主动的语言特征的研究者和语言规律的发现者进行语言学习,其学习过程是自主的。同传统外语教学模式相比,数据驱动学习更适合进入专业英

语学习的具有“发现式”能力的高级水平学生。

三、数据驱动学习在专业英语写作教学中的创新应用

(一) 数据驱动学习在词汇教学中的创新应用

词汇是语言的“建筑材料”,是理解语言和表达语言的基础。教学大纲明确规定,词汇教学是大学专业英语教学的重要组成部分。然而,教师应如何在有限的课时内让学生扎实地掌握和应用词汇,从而提高教学质量,这是对专业英语教学的一个挑战。目前,数据驱动在英语教学中运用最多的正是词汇教学,其中包括搭配、类联接(colligation)与语义韵(semantics prosody)。

1. 搭配

词语搭配无疑是语言学与应用语言学领域最为重要的概念之一。就一般的语言教学与研究而言,搭配是“在文本中实现一定的非成语意义并以一定的语法形式因循组合使用的一个词语序列,构成该序列的词语相互预期,以大于偶然的几率共现”(卫乃兴,2001: 100)。Firth将词语搭配分为一般搭配、有限的专业搭配和个人搭配。在专业英语中,对有限的专业搭配的研究词语行为更具有重要的意义。它可以确定这些词汇所用于的“语域”和发现某些特有的专业风格。词语搭配明显地受到语域的影响。在一种语域里使用频率极高的词语搭配,在另一种语域里可能很少使用,或者几乎不出现,这就是专业性搭配的概念和理论基础。由于学术研究和科技领域特有的域、式和旨,一些特有的词语组合(characteristic lexical combination)常被频繁地使用,并形成典型的意义和形式特征。

在专业英语写作教学中,教师对学生写作中词汇运用方面的问题,尤其是搭配感到很棘手,很难找到一套明确的规则来指导学生或对此做出解释。如果说基于规则的过程能有效地应付语法问题,搭配问题则可以尝试着靠数据驱动式教学来解决。

我们以专业英语中情态模糊限制语的用法为例,探讨数据驱动在写作词汇教学中有关搭配方面的具体使用情况。

一般看来,作为正式书面语体的学术英语应该力求做到语言准确、明晰,观点客观、严谨。然而,客观世界中,事物之间的联系都具有模糊和非定化的特点,因此,作为反映客观世界、表达思维结果的语言就必然存在着模糊形式。学术文体也不可能例

外。在专业英语文献中存在着大量的情态模糊限制语,专业英语写作课老师如何让学习者在写作中更准确地掌握它们的用法呢?在课堂上用有限的例句和语法讲解是远远不够的,相反,却可以利用现有的专业英语语料库,如JDEST,并选择2个库容相当的通用语料库(BROWN和LOB)作为参考语料库进行对比研究,用海量的真实的语料,让学生发现它们在专业文本中的使用规律。首先,在课前教师必须借助检索软件Wordsmith,分别提取三个语料库中九个情态模糊限制语的观察频数,然后加以标准化,将三个语料库中的九个模糊限制语的词频由大到小进行重新排序,得出下表的数据(蒋婷,2006:47)。

表 三种语料库中九大情态模糊限制语的使用频率(按降序排列)

	JDEST		BROWN		LOB
can	26.19%	would	19.27%	would	20.2%
will	18.67%	will	17.32%	will	16.69%
may	18.66%	can	16.92%	can	13.45%
would	11.22%	May	11.95%	could	12.24%
should	8.88%	could	11.37%	may	10.91%
could	6.35%	must	8.33%	should	9.5%
must	5.53%	should	7.75%	must	8.35%
might	3.37%	might	5.01%	might	5.95%
shall	1.13%	shall	2.08%	shall	2.71%

教师随后在课堂上,用多媒体的形式向学生呈现从语料库提取的数据,并给出相关的索引行,如:

1. ion should never be used , others that it can be justifiable under certain circ
2. t a multiplication by a large matrix it can be worthwhile to look for a facto
3. n - fusion hard facing deposits, but it can be important on fusion deposits t
4. t iron that is 99. 94 percent pure. It can be purchased in shot form for mak
5. lded in a protective cardboard cover , it can be mounted easily for display
6. reaches 113 deg as a result of this . It can be supposed that such zones are d

接下来,学生在课堂上分组讨论这些真实的语言现象后面隐藏的规律。在条件允许的情况下,还可以让学生独立使用索引软件,从语料库中提取更

多的真实语料,进而加以分析。最后,在老师的指导下,总结归纳出专业英语中情态模糊限制语的特点。在这种发现式学习过程中,学生可以掌握在写作中如何正确地使用情态模糊限制语的基本规律,形成对此类模糊命题的精确概念。比如,如何在专业英语写作中做出适度推测,从而达到直接缓和语气、保护自己、实现合作原则的语用功效,或者如何使用“倾向”作者的模糊限制语,以“作者缺席”为特征,目的是为了给作者留有更大的余地,减轻自己的责任,让话语更加客观和可信。(蒋婷,2006:50)

2. 类联接

类联接是词语搭配研究中的一个重要概念,它指的是文本中抽象的语法范畴间的结合,即词语搭配的语法结构和框架。如 V + N 表示一个类联接。掌握词汇的语法形式特征对于学生在写作中准确使用词汇极为重要(甄凤超,2005:23)。通过建立类联接,我们可以发现词汇的语法模式。单词因意义不同而具有不同的语法型式,而相同语法型式下的不同单词具有某种意义上的联系。下面我们仍以情态模糊限制语为例,讨论如何利用数据驱动给学生讲解情态模糊限制语类联接构成和它们的使用意义,以便让学生在写作中更准确、地道地使用这些常用结构。

很多情况下,在连续的学术文本中情态模糊限制语并不是作为单独的情态动词出现,而是与其他词搭配,构成固定结构。常见的类联接有:“It + 情态动词 + be”、“One + 情态动词”。在 JDEST 中,检索到固定结构“*It + 情态动词 + be*”的出现频数为 1698 次,结构“*One + 情态动词*”的频数为 707 次。以下分别是提取该二结构的索引行:

7. ion should never be used , others that it can be justifiable under certain circ

8. t a multiplication by a large matrix it can be worthwhile to look for a facto

9. n - fusion hard facing deposits, but it can be important on fusion deposits t

10. t iron that is 99. 94 percent pure. It can be purchased in shot form for mak

11. lded in a protective cardboard cover , it can be mounted easily for display

12. reaches 113 deg as a result of this . It can be supposed that such zones are d

13. dary layer and for a turbulent one . As one would expect from the preceding de

14. camera and so a compromise is sought . One would like to use all the incoming l

15. nsidered a component of commitment, one would expect that the features of lo

16. freezing/changing/refreezing model, one would predict that OD activities whi

17. have to serve as the primary source. One would think that the topic of color

18. irrored by the opposite trends in Uos. One would anticipate that the difference

这两种语法式结构特征在学术英语中如何使用,并且具有什么样的意义呢?

基于对数据的归纳分析,我们可以发现,“*It + 情态动词 + be*”的类联接常出现两种形式:(1)*It + 情态动词 + be + 形容词*;(2)*It + 情态动词 + be + 过去分词*。在结构(1)中,be 动词后面往往接 significant、possible、desirable、advisable、worthwhile、difficult、fair、feasible 等判断形容词,间接表达了作者对该命题的态度。结构(2)中的被动语态是科技文体中常用的结构,是“倾向作者的模糊限制语”的一种。使用被动语体,表明该事实或命题不是作者提出,而是通过一些有根据、有来源的说法,由此间接地表达作者对某问题的态度。如根据某人所说,依照某种信念,或者凭借某本著作等等,以示这种断言并非作者本人的估计、推测,避免了作者把个人意见强加于人,加强了文章的客观性和可信度。这种模糊限制语使话语更具客观性,同时也减轻作者应负的责任。

而在 *One + 情态动词 + 动词* 的结构中,情态动词后面通常接 argue、assume、conclude、notice、remark、observe 等推测性动词。当推测性情态动词和推测性动词连用,就达到了双重推测的效果,强化了对命题的不确定性和作者的推测态度。

3. 语义韵(semantic prosody)

语义韵即关键词项的典型搭配词在其语境中营造的语义氛围(卫乃兴,2002:302)。语义韵大体可分为积极(positive)、中性(neutral)和消极(negative)等三类。语义韵是语料库语言学开辟的一个新研究方向,它提供了观察和描述词项行为的新视角(卫乃兴,2002:175)。根据卫乃兴(2002)的讨论,语义韵研究可以通过:(1)建立类联接,基于数据来概括和描

述关键词的语义韵;(2)计算搭配词,用数据驱动方法研究语义韵;(3)基于数据与数据驱动相结合,建立语义韵结构。不管以何种方式,研究必须是数据驱动的,唯有依据语料库的强大依据,通过对数据进行统计、分析与概括,才能科学地描述词项可能有的语义韵。通过对语义韵的学习,学生在写作中可以避免按照语法规则生造搭配,从而掌握地道的英语,在文章中准确恰当地使用语言。(甄凤超,2005:25)

因此,基于语料库的数据驱动学习能够为写作教学提供大量的文本作为研究素材,不仅考虑到大量的语言特性,而且可以在不同语境间进行定量的对比,这为专业英语写作教学中的词汇教学提供了更加便捷、客观的途径。

(二)数据驱动学习在篇章教学中的创新应用

学术论文有其固定的语篇模式。受其制约,作者都要在文章中依据一定的形式实施某些具体的行为,这就是语篇行为。从宏观上看,任何专业领域的学术论文都含有一系列的语篇行为(discourse acts),这些语篇行为主要由一些词语搭配序列实施;从微观上看,学术论文的各个部分都由一系列的语步(discourse moves)构成(卫乃兴,2001:194),这些语步中的信息安排和功能也都主要由词语搭配序列来实现。因此,在专业英语写作教学中的语篇部分也可以充分利用数据驱动学习的模式,来培养学生在写作中的语篇构建能力。

1. 语篇行为

虽然学术论文的语篇行为因专业习惯而异,但有一些语篇行为是各专业共有的。学习和掌握这些语篇行为的共性,对学生阅读学术英语论文和写作都十分有益。卫乃兴(2001)提出一般的学术论文基本有如下几种语篇行为:界定;举例;陈述结果;比较;对比;分类;证实;推理与讨论;总结与结论。我们以“界定”为例来说明“数据驱动学习”的具体应用。

在专业英语中,所有的作者都会对新的概念、研究框架和研究对象等进行正确的界定。一般作者会用类似这样的句子:A failure is defined as a physical defect of one or more circuit component that...但在写作实践中,学生需要更多的词语搭配序列来界定对象,让文章的语篇结构更生动、丰富。因此,教师首先在课堂上提出学生所熟悉的一些词或结构,如 define, definition, know, mean 等,然后要求学生课后用

索引软件在专业英语语料库中进行检索,并对检索行进行分类、总结。在课堂上讨论后,大致可以得出以下有代表性的表示界定的搭配序列:(1) X be defined as;(2) define X as;(3) By X be meant...;(4) By X we mean;(5) X is known as;(6) X is what is know as;(7) A (another, second) definition of X is...;(8) By definition X is...

通过以上的归纳总结,学生能够掌握更多更丰富的结构来表达“界定”,并在写作中灵活运用。

2. 语步分析

科技论文写作教学是写作教学的高级阶段,但在写作实践中,无论在内容的组织、篇章的形式抑或是语言表达等方面都存在着大量、严重的不规范现象。如果我们将论文写作分解为摘要、前言、正文、结论等部分,并运用“数据驱动学习”模式,结合相关的语步理论,对这些部分的内容成分与语言形式做出细致的研究,就能给学生的科技论文写作提供切实可行的帮助,从而提高这门课程的教学质量与效率。我们以论文摘要的写作为例加以分析。

论文摘要是位于论文题目与正文之间,对整个论文内容进行概述的一个独立语篇。好的摘要能让读者迅速对论文的目的、方法、主要研究结果、意义及推论有一个概括的了解,帮助读者对该论文的学术价值做出比较准确的判断,从而决定对该论文的取舍。因此,摘要写作成为科技论文写作中不容忽视的重要环节,对摘要做写作范式方面的研究,使论文摘要无论在格式和内容上都更趋规范和全面,无疑有助于论文研究成果的推广。(余莉,梁永刚,2006:36)

摘要由四个语步组成:提供背景信息、提出当前研究、描述研究方法和展示研究结果。我们在教授这一部分的写作时,可以鼓励学生从网上或其他渠道,自建小型语料库。目前,许多大型语料库及其专业分析软件的知识产权保护,给数据驱动学习在专业英语教学中的推广和普及造成了障碍,不少专业英语教师只能“望库兴叹”。其实,我们可以依靠自己的力量打破这种垄断,在自己的教学实践中,按照科学方法自建语料库,并通过网络上提供的免费专业分析软件获取自己教学所需的数据驱动信息。如以电子期刊全文数据库文献为主,采集某一时期的学术英语文章作为语料。这些自建语料库的语料可以是同一学科,也可是跨学科,但必须遵循建库的四

个原则:(1) 操作性;(2) 代表性;(3) 随机性;(4) 库容合理性。建成小型摘要语料库以后,教师可要求学生根据摘要写作的四个语步构成,对每篇论文的摘要进行附码、标注。然后,教师和学生可以在课堂上针对摘要中出现的各种语言现象,如情态动词、动词时态和语态、第一人称代词、连接副词等,进行数据提取和处理,综合分析讨论,得出使用这些语言现象在各个学科领域的使用规律和跨学科的同异。

四、结论

综上所述,数据驱动学习在专业英语写作教学中的应用,不仅在于为教学活动提供丰富的语言材料,更重要的是促使语料库研究工具得以直接应用于学习者的学习活动,进而引发了语言学习方式、学习内容以及学习材料形式等方面的重大创新。在这种发现式学习过程中,学生由被动接受转变为主动探索,学习兴趣能够充分激发,内在的学习潜力得以挖掘,自主学习能力得到了锻炼。毋庸置疑,数据驱动学习给专业英语写作教学带来了新的革命,它有助于从根本上改变以教师和教材为中心的单一讲授的教学模式,逐步确立以学生为中心的、充分重视和开发学习者积极、主动的创作潜能,并由此提高其语言应用能力的英语写作教学模式。

参考文献:

- [1] 黄大网,等. 数据驱动学习与英语写作教学中的语域知识传授——基于 CLEC 语料库的个案研究[J]. 外语电化教学, 2003, (94): 37-41.
- [2] 陈敏哲. 高等院校专业英语写作教学改革探索[J]. 高等教育研究, 2005, (5): 97-99.
- [3] HUNSTON S. Corpora in Applied Linguistics [M]. Cambridge: Cambridge University Press, 2002.
- [4] 甄凤超. 语料库数据驱动的外语学习: 思想、方法和技术[J]. 外语界, 2005, (5): 19-27.
- [5] 蒋婷. 论学术英语中的情态模糊限制语——一项基于语料库的研究[J]. 外语电化教学, 2006, (4): 47-52.
- [6] 卫乃兴. 词语搭配的界定与研究体系[M]. 上海: 上海交通大学出版社, 2001.
- [7] 卫乃兴. 语义韵研究的一般方法[J]. 外语教学与研究, 2002, (4): 300-307.
- [8] 余莉, 梁永刚. 英语科技论文摘要的写作范式研究[J]. 外语教学, 2006, (1): 34-37.
- [9] 刘艳晖. 基于多媒体语言实验室的专业英语写作教学模式[J]. 宁夏大学学报(人文社会科学版), 2006, (1): 126-128.
- [10] 温剑波. 专业英语写作教学探究[J]. 辽宁教育学院学报, 2005, (1): 87-88.
- [11] 王金铨. 数据驱动学习——英语教学模式的新动向[J]. 扬州大学学报(高教研究版), 2004, (6): 88-90.
- [12] 周冬梅. 数据驱动语言学习——教学模式的转变[J]. 零陵学院学报(教育科学), 2003, (2): 125-128.

New Application of DDL to EAP Writing

JIANG Ting, YANG Mei

(College of Foreign Languages, Chongqing University, Chongqing 400030, China)

Abstract: DDL based on corpora has provided a novel tool for the foreign language teaching. However, the temporary EAP writing course still adheres to the traditional up-bottom pattern of teacher-centered and textbook-centered teaching. Regarding the problems in EAP writing, this paper probes into the new application of DDL to EAP writing in two aspects: vocabulary and discourse teaching. Consequently, a new student-centered teaching model will be established to inspire learners' potential and improve their ability to use standard language.

Key words: data-driven learning; EAP writing; vocabulary; discourse