

对语料库中高年级学生写作词汇搭配错误的分析

秦悦

(上海外国语大学国际金融贸易学院, 上海 200083)

摘要:分析和探讨了《中国学习者英语语料库》中标注的动词+名词和形容词+名词搭配错误及其原因。研究发现:对于动词+名词搭配错误,由于搭配的意义主要是由名词决定的,忽视或违背搭配的固定性并不影响对整个搭配意思的理解,但缺乏地道的搭配会影响写作的准确性。对于形容词+名词搭配错误,L1干扰是主要的原因。词语搭配运用是否得当,会影响到写作的准确性和地道性。帮助学习者掌握搭配类型,利用语料库学习和探索搭配,能有效地解决学习者写作中词不达意的问题,提高写作表达的准确性。

关键词:《中国学习者英语语料库》;词语搭配;搭配错误

中图分类号:F592 **文献标志码:**A **文章编号:**1673-1883(2018)02-0097-05

An Analysis on Advanced Learner's Lexical Collocation Errors in CLEC Writing

QIN Yue

(School of Economics and Finance, Shanghai International Studies University, Shanghai 200083, China)

Abstract: The use of collocations will affect the accuracy and idiomaticity of a piece of writing. The present paper analyzes and traces the causes of two types of lexical collocation errors by advanced learners in Chinese Learners English Corpus (CLEC): verb + noun and adjective + noun collocation errors. It is found that for the verb + noun collocation errors though it is the noun decides the meaning of the collocation, the inadequate combinations of verb + noun undermines the accuracy of writing to a large extent. For the adjective + noun collocation errors, L1 interference is a major cause. It is believed that exploring learner corpus can be an effective way to improve learners' writing. Corpus allows students to observe an extensive data, which help them understand collocations and finally master them. Utilizing corpus to explore collocations will solve the problem of inaccurate expression in learner's writing.

Keywords: *Chinese Learners English Corpus*; collocation; collocational errors

一、引言

许多语言学家和研究者,都十分重视搭配在写作中的运用。词语搭配的好坏,会影响到写作的准确性和地道性。二语习得在对英语本族语者和非本族语者的写作进行的研究中发现,搭配与写作能力有很大的关系,不同水平的学习者使用不同的搭配组合。L2水平高的学习者,较之L2水平低的学习者,使用的搭配显著地多,而且更为正确,类别也更多。总之,运用搭配的多寡,运用搭配的多样性和使用的准确性,是写作顺畅的保证(Lewis 2000, Granger 1998, Howarth 1998)。掌握搭配,是自如运用一门语言的前提。而遵循一门语言的搭配传统,会使学习者的语言使用显得很地道。

根据《高等学校英语专业英语教学大纲》的

规定,在大学学习两年后,学生的词汇量应达5 500~6 500个单词,并能有效地使用其中3 000~4 000单词及其基本搭配。在三年级的时候,提高课堂教学和其他途径认知词汇应该达7 000~9 000,且能正确而熟练地使用其中的4 000~5 000及其最常用搭配。对非英语专业的学生也有类似的要求:在5 500基本词汇中,学生应该知道3 000个单词及其搭配。为了解大学生英语写作中搭配使用的情况,本文采用《中国学习者英语语料库》(Chinese Learners English Corpus, CLEC,以下简称语料库)中标注的词汇搭配错误,分析高年级学生写作中出现的搭配错误及其成因,讨论搭配与写作质量的相关性,并指出语料库的运用有助于对加深搭配理解和提高写作精准性。

二、以往研究概况

自弗思在1957年提出“搭配”(collocation)这个概念以来,作为一个受到普遍关注的语言学话题,学者们的看法不一,搭配被定义为一种词汇现象、语法现象,或更笼统一点地说,语言现象(Stubbs 2001; 卫乃兴 2002)。“搭配”虽然在语言学习和语言教学中有着模糊的意义,但归纳起来有两种主要的观点(Nesselhauf 2005):以数据统计为导向或以频率为依据的观点认为,搭配是词语的同现,区别仅在于频率的多寡。以意义为导向的观点认为,搭配是词语的组合,有一定程度的固定,但不完全固定。因此,搭配是“词语在实际运用中的习惯性结伴”(actual words in habitual company)(Firth),搭配是词语的可预知的组合,但却是受限制的组合(卫乃兴 2002)。搭配具有任意性(arbitrariness)、可能性(likelihood)、高频性(frequency)的特点(Stubbs 2001, Howarth 1998)。因此,搭配并不是词语的相互替换,而是哪些词经常相伴出现,哪些词语组合同现的更有可能,出现的频率更高,更可预期。一般说来,词汇方面的错误并非出在单个的词上,而是错在词的组合上。Carter(1998)认为,由于学习者接触到一个词或词语组合的频率不够高,不足以把它缩小到固定的伙伴关系,因此,错误的组合不可避免地成为二语输出中的普遍现象。

许多研究都证明:搭配错误在L2学习者所犯的的错误中占了很高的比例(Howarth, 1998; Nesselhauf, 2003)。他们的研究发现,即使是程度高的学习者在词语搭配方面也有不小的困难,常出现不恰当的、不可接受的搭配。虽然这些学习者已经成功地驾驭了有相当难度的语法结构,也有着很好的观点,但由于他们不能准确运用关键词的搭配,导致无法清楚地表达所写的话题。所出现的搭配错误不仅“影响到文章的好坏”,而且“这些搭配错误加起来就会使文章有失精准”(Howarth 1998)。对搭配了解不充分,已经成为高级学习者在语言输出时的一大障碍。

学习者语料库是运用计算机根据一定的标准,对能代表某一学习群体的语言学习者产生的文本,进行系统的计算机选择收集整理。Nesselhauf(2005)认为,与本族语语料库不同的是,组成学习者语料库的文本往往不是“自然发生的文本”。因为在外语学习中,学习者语料库的语料通常是经过一定程度的设计而引发的语言产出,收集的大多数命题作文或口语访谈。很多学者利用语料库对搭

配进行了研究。濮建忠(2003)研究了中国英语学习者的英语动词的使用情况。他利用《中国学习者英语语料库》(CLEC)的一个子语料库描述并解释动词行为:类联接、搭配及词块,并将学习者使用动词时所呈现出的特点与控制语料库Brown和LOB进行了参照,以找出学习者与本族语者之间的异同。Shei & Pain(2000)利用研究搭配错误的学习者语料库,提取标准英语搭配的参考语料库两个不同的语料库,设计了在线纠正(on-line correcting program)侦查学习者英语写作中出现的搭配错误,并提供语料库中标准搭配的例子,以供参考。Yu-Chih & Li-Yuch(2003)用检索系统研究了归纳和演绎两种方法学习搭配的相对有效性。认知方法与搭配的难易程度也得到了研究。结果表明,用归纳法学习搭配的小组掌握的情况明显比用演绎法的要好。

三、本研究对《中国学习者英语语料库》中标注的搭配错误的选取

学习者语料库收录的语言产出,让我们有可能探究了解学习者在特定情形中的实际产出,更好地解释学习者真正想表达什么,让我们能发现和提出新的假设。对学习者的语料库进行语言分析,已经成为研究二语学习者写作的一个重要手段。本研究的目的是描述并解释高年级英语学习者写作中的词汇搭配错误,因此,我们采用《中国学习者英语语料库》来进行描述。

《中国学习者英语语料库》(CLEC)共采集了不同英语水平的超过100万词语料:ST2(高中生);ST3(CET4。非英语专业的大一和大二的学生,他们中大多数都会参加大学英语四级考);ST4(CET6。非英语专业的大三和大四的学生,他们中大多数都会参加大学英语四级考);ST5(英语专业的大一和大二的学生);以及ST6(英语专业的大三和大四的学生)(桂诗春,杨惠中,2003)。

CLEC对所收集文本的错误进行了标注,从句子到词组,再到词汇层面,一共标识了61种错误。

表1 CLEC 搭配错误代码及其描述

代码	类别	描述
cc1	n/n collocation	improper noun (phrase) and noun (phrase) combination/semantic
cc2	n/v collocation	improper noun (phrase) and verb (phrase) combination/semantic
cc3	v/n collocation	improper verb and noun (phrase) combination/semantic
cc4	a/n collocation	improper adjective and noun (phrase) combination/semantic
cc5	v/ad collocation	improper verb and adverb (or adv) combination/semantic
cc6	ad/a collocation	improper adverb and adjective combination/semantic

来源:桂诗春&杨惠中2003:8。

与本研究相关的是搭配错误(collocational errors), [cc]是搭配错误的代码。表1是对各个代码的描述。

由于我们的研究对象是高级学习者,我们进一步把所标注的语料缩小到ST4(非英语专业的大三和大四的学生)以及ST6(英语专业的大三和大四的学生)的写作搭配错误(表2)。

表2 ST4及ST6写作搭配错误数

错误类型代码	cc1	cc2	cc3	cc4	cc5	cc6
ST4	95	60	505	163	36	8
ST6	69	26	160	61	8	2

来源:桂诗春&杨惠中2003:48。

搭配错误是这样标注的:[错误类型代码,错误范围]。[cc3, -4]代表的是动词名词搭配错误,发生的范围是在表示错误标识的方括号前的这个词和包括被标识这个词在内的第四个词。在下面的这个例子中,动词+名词搭配错误发生在see和story之间:discuss the issue, let's see [cc3 -4] a Chinese story first.

从[cc1]到[cc6]这六个错误类型中,[cc3]动词+名词搭配和[cc4]形容词+名词搭配最具有数理统计意义。[cc3]是61种错误中最常见的,[cc4]紧跟其后。因此,我们就深入探讨并详细分析《中国学习者英语语料库》(CLEC)所标注的这两种错误类型。

(一)对动词+名词搭配错误的筛选

在ST4子语料库中,共标注有505个动词+名词搭配错误;在ST6子语料库中有162个。这些错误都表明,动词+名词直接的搭配偏离本族语者的表达习惯,或不是典型的本族语者的表达。

然而,我们进一步审视却发现:有些组合并不具备搭配的某些特质,它们更接近于词语的自由组合(free combination),或者它们不太可能形成组合,如reverse its foreign debt, torment their senses, wear husband's surname。换句话说,它们只是临时性的一次组合,并不具有讨论的意义。因此,并没有被视为搭配,但give a conclusion和make success则被算作是搭配错误,因为它们很显然是想表达draw a conclusion和achieve success意思。筛选出来的搭配再去BBI Dictionary of English Word Combination和LTP Dictionary of Selected Collocations两本词典里去比对,以验证提取的组合是否可接受。这项工作拯救了show one's ability, occupy your time, meet this situation, fulfill one's fantasies, cut down forest等,使它们免于错误之列。在一番批判筛选后,最后只有ST4里的274个动词+名词搭配错误,ST6里的89个动词+名词搭配错误得到进一步的研究分析。

(二)对形容词+名词搭配错误的筛选

在ST4子语料库中,共标注有163形容词+名词搭配错误;在ST6子语料库中61个。

形容词+名词搭配也同样经过[cc3]一样的甄别筛选程序。一些与good、bad、true、real的搭配,如bad men, good commodities, good reputation并未入选。诸如a fired job、metal result这样的组合也没有考虑进去。如果是名词误用的情况,如price和prize在it is necessary to give a great [cc4 -] price to the man who reports the source的误用,也不视作为搭配错误。一些应该视为是语法层面的搭配错误也不在本文讨论之列,如:we let him still have so many [cc4 -1] freedom to misbehave and why

经过甄别和筛选之后,来自两个子语料库的110个形容词+名词搭配错误进入了下一步分析研究。由于来自两个子语料库的搭配错误样本不够大(ST6的只有32个),因此我们就不再作具体的区分,而是统一进行分析和讨论。

四、发现与讨论

(一)[cc3]动词+名词错误

对于[cc3]即动词+名词搭配错误,我们有如下的发现:

1. ST4学习者在使用含有失去词汇意义的动词/轻动词(a de-lexicalized verb)的搭配时不够正确。类似do harm to/make harm to, try my best/make my best的例子很多,部分原因是这些动词本身不具有特别明显的词汇意义,搭配的意义主要是由名词决定的。最重要的原因是搭配正确与否并不影响对整个搭配意思的理解。

2. ST4学习者忽视搭配的固定性。搭配有其固定性,每个词只有那么相容的几套搭配伙伴,有的甚至是只有固定的某几个词或某一个词,任何替代都不行。例如, pay attention to永远不能是*put attention to。这表明学习者对搭配所要求的固定性还不是很了解:词的选择是受限制的,如“遭受损失”应该是suffer loss而非lose benefit。词是不可以不受限制地简单地组合成词组,或受制于它前面出现的词,抑或后面出现的词。

3. ST4学习者在运用一些意为“成功、实现预期、达到要求”的动词如achieve, gain, get, fulfill, make和reach等不加区别:gain achievement/aim/goal/success, get much achievement, get to the goal, get to the purpose, make the success, reach the effects。这表明,即使是这些常用的组合,学习者对其固定搭配

也未牢固掌握。

4. ST4学习者把do当成了万用动词,替代搭配要求的其他动词。*do success (achieve), *do activity (take part in), *do any measures (take), *do examination (take), *do penalty (impose), *do preparation (make), *do some achievement (make), *do effort (make)。过度使用(overuse)是词语搭配中的一个重大问题。在需要更为具体和更贴切的动词,学生的选择仍然是集中在那么几个动词上,如get,即使是这时用accumulate、acquire和gain更为可取。“缺乏习得那些能让语篇流利和地道的语块(language chunks)”(Aghbar 1990)是导致写作的准确性大打折扣的原因。

5. ST6中此类错误的成因更多地是违背了搭配的固定性,出现错误的动词也不局限于那么几个失去词汇意义的动词:cause imagination, extend one's resignation, strike violent crime, and theory concluded等。尽管do仍然无所不在,但它所想表达的意义远比ST4丰富得多:*do contribution (make), *do great threat (pose), *do punishment (carry out), *do some activity (take part in), *do the crime (commit)。另外一个特点是:ST6学习者在此类错误中犯得更多的是语义方面的错误。这表明在动词+名词搭配方面,ST6学习者已经掌握了基本的固定搭配,但更高层次搭配知识还有待完善。

总的来说,动词+名词搭配对高级学习者来说依然是困难。

(二) [cc4] 形容词+名词搭配错误

Gitsaki(1999)认为,形容词+名词的搭配是很早习得的。因而可以假设形容词+名词的搭配问题没有形容词+名词那样严重,原因是第一语言(L1)的影响已经淡出。然而,Carter(1998)却提醒我们,形容词+名词的搭配是最难掌握的。哪一种说法更接近事实呢?根据CLEC标注的错误分析来看,L1干扰是主要的原因。搭配错误既是语内问题,又是语际问题(Gitsaki 1999)。

1. 字面翻译在形容词+名词的搭配中无所不在。从deep question到constant place,翻译会‘促进’L1对L2输出的影响。这些中式英语所犯的错误让L1的读者很容易认同,也很难觉察到,因为他们都不需要解码过程。L1的语义范围越大,所犯的推断(inference)错误就越多,搭配中出现的同义词或近义词就越多,学习者也就很少能给出限制紧密的搭配。学生写作中出现的偏离限制的回答(deviant answers),表明他们英语搭配知识十分有限。

2. 意义相近的形容词是引起形容词+名词的搭配错误明显原因之一。与学习者在动词+名词搭配出现错误的成因不同,形容词+名词的搭配错误产生的主要原因是学习者对意义相近的形容词的意思一知半解,所犯的错误完全是忽视名词所要求搭配的形容词。“经验丰富”应该为rich experience而非下面例子中的ample experience和wealthy experience

job then we can have ample experience [cc4 1-] on the basis of the good

sent [vp6,s-] them home. the wealthy [cc4 1-] experience on war is a country's

3. 形容词+名词的搭配的规定性,没有动词+名词的那么强。但是,学习者创造性的形容词+名词的搭配应谨慎对待。如果出现搭配词不属于某个词的词汇集(lexical set)或者超出了它的搭配范围(lexical ranges),那么,这个搭配组合就是错误的。如injure一词只能和injure somebody或somebody's pride/feelings,但不能说*injure the right。但是,形容词+名词的搭配却并非如此。我们不会在success的搭配集里发现sweet,但在某一特定上下文中,我们却能理解最终取得成功那种喜悦的心情。换句话说,形容词+名词搭配正确或可接受与否,有的时候更多的是取决于上下文(context-dependent/context-specific)。虽然搭配反应的是一种预期,违反预期组成不同寻常的组合却能取得某种效果,但像sweet success这样的组合在学习者还没有搭配意识的情况下,不宜多鼓励:

feminism does obtain sweet [cc4 -1] success in the fields as was [vp3,s-] stated

虽然形容词+名词的搭配是最早习得的搭配类型,在构建时学习者依然过多依赖于字面翻译。如果读者不是和作者有一样的L1的话,像a flat life(平凡的生活)这样的搭配根本无法理解。另一方面,有real和true这样的搭配虽然也是可以接受的表达,但会令整个写作平淡无奇,不甚精准。

5. 搭配与写作质量的提高

无论是动词+名词还是形容词+名词的搭配,高级的英语学习者的表现都差强人意。要提高写作质量,应该从平时对搭配的点滴关注开始。Fernado(2000)指出,搭配都是特定话题有关的(topic-related)。搭配围绕一定的话题组织展开,在类似的语境中反复出现。如讨论“分班教学的利弊时”,与这一话题相关的搭配,无外乎关于班级的regular class, slow class, accelerated class, mixed class/heterogeneous class;关于分班好处的lessen anxiety,

broaden interest 等等。提前做有关搭配的功课,让学生找出与该话题关键词有关的一系列搭配,以避免写作时的生译硬套。如在写商品打假前,先找出 counterfeiting activity, proliferation of counterfeiting, crackdown on counterfeits, shoddy and counterfeit goods 这样的搭配,学生的写作就会更流畅。

搭配更容易激发与特定情境的联想 (situationally evocative), 相应地也更容易激发对写作的想象。例如 widely available 这个搭配无法不让人想到市场里某件很常见的商品。写作之前,不妨先头脑风暴一下与话题有关的搭配,不失为解决才思枯竭的一个办法。关注搭配不仅能减少搭配错误,也能提高表达的准确性,既规避了 L1 的影响,也

避免了直接翻译带来的错误,是提高写作质量的一个有效手段。

“以语料库为基础的研究不是简单地汇报一些量化的发现,更重要的是探讨这些发现对学习语言类型的重要性”(Biber et al)。《中国学习者英语语料库》(CLEC)提供了观察语言大数据的机会,让我们更好地了解某些语言层面的过度使用以及使用不足 (overuse and underuse)。利用语料库学习和探索搭配是个有效的手段。如果语料库涵盖本族语者和非本族语者的语料,学习者还能将自己的写作与本族语者的比对,学习到更多的表达方法,也能加深他们对搭配的理解,掌握搭配类型,最大程度地削减 L1 的负面影响,进而提高写作表达的准确性。

参考文献:

- [1] AGHBAR, A. A. Fixed Expressions in Written Texts: Implication for Assessing Writing Sophistication[M]. Paper Presented at a Meeting of the English Association of Pennsylvania State System Universities, 1990.
- [2] BARNBROOK, G., O. Mason & R. Krishnamurthy. Collocation: Applications and Implications [M]. New York: Palgrave Macmillan, 2013.
- [3] BENSON, M et. al. The BBI Combinatory Dictionary of English: Your Guide to Collocations and Grammar. [M].3rd edition. John Benjamin's Publishing Company, 2010.
- [4] BIBER, D. et al. Corpus Linguistics [M]. Cambridge: CUP, 1998.
- [5] CARTER, R. Vocabulary: Applied Linguistic Perspectives [M]. New York: Routledge, 1998.
- [6] FERNANDO, C. Idioms and Idiomaticity. [M]. Shanghai: Shanghai Foreign Language Education Press, 2000.
- [7] FIRTH, J. R. "A synopsis of linguistic theory, 1930-1955", in Palmer, F. R. (ed.) Selected papers of J. R. Firth 1952-1959[C]. London: Longman, 1957:168-205.
- [8] GITSAKI, C. The Development of ESL Collocational Knowledge[M]. San Francisco: International Scholars Publication, 1999.
- [9] HOWARTH, P. "The Phraseology of Learners' Academic Writing and Second Language Proficiency", in Cowie, A. P. (ed.) Phraseology: Theory, analysis, and applications[C]. Oxford: OUP, 1998:161-186.
- [10] HILL, J. & M. Lewis. LTP Dictionary of Selected Collocations. [M]. Beijing: Qinghua University Press 2006.
- [11] LEWIS, M. Teaching Collocation. [M]. Hove: Language Teaching Publications, 2000.
- [12] NESELHAUF, N. The Use of Collocations by Advanced Learners of English and Some Implications for Teaching" [J]. Applied Linguistics, 2003(24)/2: 223-242.
- [13] NESSELHAUF, N. Collocations in a Learner Corpus [M] Philadelphia: J. Benjamins Publishing House, 2005.
- [14] SHEI, C., PAIN, H. An ESL Writer's Collocational Aid [J]. Computer Assisted Language Learning (2000)13/2: 167-182.
- [15] STUBBS, M. Words and Phrase: Corpus Studies of Lexical Semantics. [M]. Blackwell Publishers, 2001.
- [16] YU-CHIH S, Li-Yuch, W. Concordancers in the EFL Classroom: Cognitive Approaches and Collocations Difficulty [J]. CALL, 2003(16): 83-94.
- [17] 桂诗春,杨惠中.中国学习者英语语料库[M].上海:上海外语教育出版社, 2003.
- [18] 卫乃兴.基于语料库的对比短语学研究[J].外国语,2011(4):32-42.
- [19] 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,2003(6):438-445.