

语料库在对外汉语教学中的应用

卢伟

摘要:以语料为基础的语言研究方法及其成果语料库,在对外汉语教学与研究方面可以得到广泛的应用。现代汉语语料库不仅可以为编制教学大纲提供定量依据,为编写教材提供真实的语言素材,为编纂教学专用辞书提供语料和语言信息,还有助于研究语言形式的语境意义和用法,有助于发现第二语言学习和习得的规律,有助于发现和修正语言规律,促进汉语研究和教学的发展。因此,有必要加强我国语料库建设及其在对外汉语教学与研究中的应用。

关键词:语料库;应用;对外汉语教学

收集真实语言环境中的语言材料,并将它们作为素材来解释和例证现存的语言现象,或者发现新的语言规律,这是语言研究的基本方法。在计算机尚未应用于语言研究之前,语言研究人员只能通过手工收集、记录和整理语言材料。近年来,由于计算机应用于语言研究,建立语料库所需的技术设备日臻完善,语料库的建设规模不断扩大,功能越来越强大,检索愈来愈方便,利用语料库进行语言研究的手段也随之得到改善。正如语言学家 M. A. K. Halliday 所言,“现代语料的巨大包容性及开发语料的种种手段的出现构成了深化我们对语言的认识和理解的强大力量。”

不断涌现的研究成果改变了人们对语言研究的固有观念和模式,也使得利用语料进行语言研究的优良传统得以发扬光大。如今,语料库的建设和利用它进行语言研究,正在发展成为一门新兴的语言学分支学科——语料库语言学。它对语言研究的许多领域产生愈来愈大的影响,并广泛地应用于语言的词汇、语法、语义、语用、语体等方面的研究、社会语言学研究、口语研究、词典编纂以及自然语言处理、人工智能、机器翻译、言语识别和合成等研究。

对外汉语教学是将汉语作为外语或第二语言的教学,属于语言学的应用学科,涉及多种相关学科的交叉研究,必须从相关的新兴学科中汲取研究方法和应用成果,以丰富自身的学科理论,促进学科的发展。以计算机存储的语料为基础的语言研究方法,是当今语言学任何分支学科研究的一种自然的、现代化的、甚至是不可或缺的辅助方法。因此,我国的对外汉语教学也必须从相关的语料库中提取真实、客观、丰富的语言素材,进行教学与学术方面的研究。笔者认为,语料库在对外汉语教学与研究方面应得到广泛的应用。

1. 为研制汉语词汇、语法和汉字教学大纲提供科学的参考依据

定量统计是研制任何标准应该遵循的原则。我国对外汉语教学界的两个指导性纲领《汉语水平词汇与汉字等级大纲》和《汉语水平等级与语法等级大纲》,也把“多学科定量统计”作为编制原则之一。如果我们建成一个学科覆盖范围广泛的大规模现代汉语语料库,那么就可以

为从事对外汉语教学和研究人员提供大量反映各种语体和语域的真实语料,便于对现代汉语的字、词和语法项目进行更加科学、客观、准确的定量统计分析。可见,汉语作为外语或第二语言教学的各种等级大纲,如果能够建立在语料库提供的汉字、词汇和语法项目使用频率的定量统计基础上,就更具有广泛性、客观性、科学性和权威性,因而对教学实践和相关的学术研究也更有指导意义。

2. 为编写对外汉语教材提供真实的语言素材

对外汉语教学的最终目的,就是培养留学生运用汉语进行言语交际的能力,因此必须强调用于教学的语言材料的真实性,也就是语言素材应该具有实际的言语交际价值。如果我们能从大量真实的口头与书面语料中选取素材,就能让学生学到自然的、地道的汉语。反之,如果用于第二语言教学的汉语教材中,关于词汇和语法教学的语言素材和例子是编者依靠经验、凭借语感或者按照语法规则演绎编造出来的,那么其中有些例子就可能与实际交际中的言语事实有差距,甚至可能失去交际价值。英国语言学家 Quirk 曾批评道:“从这些自己编写的例证出发”,“把材料仅仅当作‘例证’来使用,用来证明先验的、或者是语法学传统规定的、甚至是凭直觉认定的某些语法上的区别和结构,而不是从大量自然语言材料中归纳这些语法上的差别,这只能是一种很不令人满意的权宜之计。”

有鉴于此,我们首先应该利用语料库所提供的材料进行分析和统计,从而发现使用频率高的词语、表达方式和句型等,将它们编入教材并作为教学重点,这样就可以减少教学的盲目性。其次还应该根据语料库所提供的语料,解释和例证语法规则和词语用法。根据语料处理教学中的语言项目,具有量化的标准和统计学的依据,能够使语言描写更加全面客观,解释更加准确科学,例证更有说服力。

3. 有助于研究语言形式在语境中的意义和用法

如果没有语料库为我们提供大量的语言材料,仅仅依靠有限的语料对语言形式进行意义和用法研究,可能会因为材料不足导致研究结论不够准确全面。相反地,利用语料库的定位检索系统,比如“上下文中的关键词(KWIC)”索引形式,可以从大量的语料中检索出某个汉语单词、短语、语法结构等语言形式以及它们出现的上下文。现代化的手段使得我们有可能从宏观的角度,对语言形式及其语境进行全面的观察和分析。因此,利用语料库进行各种语言形式在语境中的语义和语用研究,可以避免分析和判断的失误,有助于更加全面准确地认识汉语语言形式在交际语境中的意义和用法,据此所进行的描写也更加可靠。其研究成果可以帮助学习者掌握汉语单词、短语和表达方式的正确用法,加深对词语的语境意义的理解,还有助于克服第二语言学习过程中的“语用失误”(pragmatic failure)。

4. 可提供第二语言学习者的“中介语”(interlanguage)形式

第二语言学习与习得理论重视对学生在学习过程中所犯的言语错误进行对比分析,特别是对“中介语”进行偏误分析。传统的作法是,教师通过人工收集学生的病句,但是由于受到种种客观原因的限制,所能搜集到的病句毕竟有限,而大规模的第二语言学习者语料库却能为教师提供现代化的研究工具。比如北京语言文化大学的“汉语中介语语料库系统”,收集了汉语中介语语料 350 多万字,其中核心语料 100 多万字作了断句、分词和词性标注。通过语料检索系统,实现了对语料在字、词、句、篇等不同层次上快捷方便的自由检索和浏览,可以提供各种单项的或综合的资料、数据和信息。它为中介语研究和偏误分析提供了大量宝贵的资源,为这

方面的研究提供了科学便捷的集成环境和先进技术手段。

与人工收集有限的学生病句卡片资料相比,“中介语”语料库能够更加真实详尽地反映第二语言学习者的汉语中介语本来面貌,帮助我们更加全面系统地观察他们学习和习得汉语的过程,了解影响学习和习得的各种因素,所发现的规律反过来可以为第二语言教学的总体设计、教材编写、课堂教学、测试等各个环节的研究提供依据。此外,建立在中介语语料库基础上的偏误分析,也有助于预测和纠正第二语言学习者的言语错误。

5. 为编写第二语言教学专用的辞书提供真实的语料和相关的语言信息

西方早期的某些语料库是为编纂词典收集语料而建立的。西方所谓的“语料库革命”给信息资料学带来了根本的变化,因此我们完全有理由说,大型的汉语语料库也必将引起汉语辞书编纂的变革。同时,也为编写对外汉语教学专用的辞书提供大量真实客观的语料和相关的语言信息,诸如词频统计、词汇分级、词性标注、词语搭配、语体特征、词语用法、语境中的例句,等等。此外,在编纂词典的时候,首先应该坚持“描写性”(descriptive)而不是“规定性”(prescriptive)的方法,也就是要客观地描述语言事实,而不是单凭编者的主观判断来规定词语的用法。因此,如果对外汉语教学专用的汉语词典能够以大型的汉语语料库为基础,那么,这样的词典必将更加具有统计学的定量依据,对词语的意义和用法的描述也会更加客观全面,其科学性和权威性也就更强,对教学与研究也就更有参考价值。

6. 有助于从汉语的实际使用情况中发现和修正语言规律

英国语言学家 Quirk 等人编写的《当代英语语法》之所以被公认为权威之作,很重要的原因在于这本语法专著是建立在“英语用法调查”语料库的基础上的。后来,他们又根据扩充了的语料合编了更权威性的《当代英语语法大全》。这本巨著发现了一些新的语法规律,补充或修正了一些原有的语法规则。“语料库可用来作为验证有关语言假说的手段”。西方语言学家利用语料库研究语言,特别是语法的经验和成果表明,对语料库提供的大量语料进行分析,有时可以发现一些过去被忽略的语言规律,比如语法书上没有或者没有详细描述过的语言规律。有时,以语料库的大量语料为基础所进行的语言研究,更有可能使语言研究者发现原有对某些语法现象的解释与自然语言中的实际情况不相符合,这就使得他们有机会去修正或补充前人的结论,促进语言研究不断发展。因此,语料库不但有助于汉语语言的应用研究,而且为高效、全面、科学地进行现代汉语的理论研究提供了强大的现代化技术保证,特别是在发现和修正语言规律方面大有可为。毫无疑问,语言规律的发现和修正对于汉语研究具有重要的意义,而对汉语本体的研究成果,也必将促进对外汉语教学与研究进一步发展。

我国语料库建设发展较慢,这对于一贯重视利用真实语料进行语言研究的我国语言研究传统来说,显得很不适应,尤其是在对外汉语教学的应用方面,还有待于进一步发展。比如我国的对外汉语教学界,虽然成功地研制了“汉语中介语语料库系统”和“现代汉语研究语料库系统”,为第二语言教学与研究提供了基于大规模真实语料的现代化研究环境和技术手段,但是仍然远远不能满足教学与研究的实际需求,况且还有很多理论和技术问题没有解决。诸如怎样汲取和应用西方语料库语言学的研究方法和建库经验,现有的语料库规模偏小,语料标注加工的广度和深度不够,语料检索软件还有待改进,等等。

针对这些不足,我们认为至少必须重视以下几个方面的工作。首先必须继续学习西方语料库语言学的理论和方法以及语料库建设的经验和教训,并结合汉语特点加以汲取和应用。

其次,继续收集样本语料扩充并标注现有的语料库,使之语料更丰富,覆盖面更广,标注更科学。可以利用光电扫描技术和语音录入软件提高语料输入的效率,也可充分利用现有的电子出版物和“因特网”上的中文信息作为语料来源。应该特别加强语义层面和语用层面的标注。第三,有关部门统一组织力量研制不同用途的新语料库,比如普通话口语语料库、“专用型(specialized)”语料库、双语“平行对照型(parallel)”语料库等,以利于普通话口语研究、汉语与外语的双语比较研究。第四,在技术方面,提高现有语料库的运行速度,引进国外的或者改进现有的标注和检索程序,开发新的工具软件,使语料检索更加方便快捷,功能更加强大。第五,尽最大可能地发挥现有语料库的作用。已建成的语料库不要只局限于小范围内的研究人员使用,而应该推向社会,至少应该对全国从事对外汉语教学与研究的人员开放。建议采用有偿服务的方式为有关人员提供检索查询的方便,或者通过“中国教育和科研计算机网”(CERNET)上网,提供(有偿)查询浏览,这样才能达到信息资源共享,更好地为全国对外汉语教学与研究服务。

注:

M. A. K. Halliday: Corpus studies and probabilistic grammar, in Aijmer and Altenberg (eds). 1991. English Corpus Linguistics: Studies in Honour of Jan Svartvik. London: Longman. pp. 41~42.

胡明扬译:《英语用法调查》,《国外语言学》1992年第4期,第39页。

David Crystal: *A Dictionary of Linguistics and Phonetics* (2nd Edition) Blackwell, 1985, p. 77.

参考文献:

Aarts, Jan and Willem Meijs (eds.), 1990: *Theory and Practice in Corpus Linguistics*. Amsterdam - Atlanta: Rodopi.

Aijmer, K. and B. Altenberg (eds), 1991: *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman.

Johansson, Stig and Anna - Brita Stenstrom (eds.), 1991: *English Computer Corpora: Selected Papers and Research Guide*. Berlin and New York: Mouton de Gruyter.

陈建生:《关于语料库语言学》,《国外语言学》1997年第1期。

陈小荷:《“汉语中介语料库系统”介绍》;孙宏林等:《“现代汉语研究语料库系统”概述》;北京语言文化大学教务处编:《北京语言文化大学入选第五届国际汉语教学讨论会论文汇编》,《北京语言文化大学入选第五届国际汉语教学讨论会论文汇编》,1996年。

丁信善:《语料库语言学的发展及研究现状》,《当代语言学》1998年第1期。

胡明扬:《现代汉语通用语料库的建库原则和设想》,《语言文字应用》1992年第3期。

刘连元:《现代汉语语料库研制》,《语言文字应用》1996年第3期。

作者:海外教育学院副教授 责任编辑:邵青莲