国家社会科学基金项目优秀成果
上海外国语大学重大科研项目成果

谭晶华 主编 毛文伟 著

中国日语学习者语料库的构建及应用

上海外语教育出版社
外教社 SHANGHAI FOREIGN LANGUAGE EDUCATION PRESS
www.sflep.com

国家社会科学基金项目优秀成果
上海外国语大学重大科研项目成果

中国日语学习者
语料库的构建及应用

谭晶华 主编 毛文伟 著

上海外语教育出版社
外教社 SHANGHAI FOREIGN LANGUAGE EDUCATION PRESS

图书在版编目(CIP)数据

中国日语学习者语料库的构建及应用/谭晶华主编；毛文伟著.
—上海：上海外语教育出版社，2012
ISBN 978-7-5446-2682-8

Ⅰ.①中… Ⅱ.①谭… ②毛… Ⅲ.①日语－自学参考资料
Ⅳ.①H36

中国版本图书馆 CIP 数据核字(2012)第 021063 号

出版发行：上海外语教育出版社
（上海外国语大学内） 邮编：200083
电　　话：021-65425300（总机）
电子邮箱：bookinfo@sflep.com.cn
网　　址：http://www.sflep.com.cn　http://www.sflep.com
责任编辑：王　俊

印　　刷：同济大学印刷厂
开　　本：890×1240　1/32　印张 8.625　字数 244千字
版　　次：2012 年 6月第 1版　2012 年 6月第 1次印刷
印　　数：1 000 册

书　　号：ISBN 978-7-5446-2682-8 / H · 1278
定　　价：28.00 元

前言

基于学习者语料库的应用研究始于20世纪80年代末，是二语习得研究中的一个新兴领域。尽管历史较短，但是该领域的研究以新颖的方法、丰富的语料在描述学习者语言特征、揭示第二语言发展规律方面取得了一系列丰硕成果，对二语习得研究的发展起到了积极的推动作用。

基于学习者语料库的研究通常可以分为两大类，即中介语对比分析(CIA)和计算机辅助错误分析(CEA)。前者是在母语使用者和非母语学习者或不同母语背景的学习者之间进行定性或定量的对比分析。后者则以中介语中的错误为研究对象，对其进行分析和归因。

二语习得是一个复杂的学习过程。中介语对比分析(CIA)和计算机辅助错误分析(CEA)对于探讨二语习得规律有着不可忽视的重要作用。通过对学习者语料以及本族语语料进行多维度的审视和比较，我们能够发现其中隐含的语言能力发展的客观规律，并对中介语形成过程进行系统、科学的描述。这将提高我们对于日语中介语形成规律的认识，为研究第二语言体系的特征以及与母语的内在联系提供坚实的基础，并为外语教学提供有益的启示，帮助学习者的产出向本族语使用者不断接近。

另一方面，考察不同阶段的学习者产出中出现的各类表达失误，分析成因，探讨对策，能够促使教师及学习者正确看待在第二语言形成过程中出现的各种问题，帮助学生有意识地避免表达失

误的发生。通过在教学过程中有的放矢地对这些问题加以纠正，还能够较好地避免或者延缓二语习得过程中误用、石化、磨损等现象的发生，提高语言教学的效率和效果。

纵观国内外研究现状，在考查欧美语言特别是英语学习者学习状况方面，学习者语料库得到了较为广泛的应用。而相对于英语学习者语料库研究的蓬勃发展，基于日语学习者语料库的研究不仅数量较少，而且在理论体系的构建以及研究的深度、广度等方面都存在着不小差距。这主要可以归结于以下两个原因。首先，从研究理念来看，迄今为止的国内外日语研究往往集中于语法、词汇等语言本体研究，日语中介语研究尚未得到应有的重视。

此外，日语学习者语料库基础建设的滞后也阻碍了研究的发展。纵观现有的日语学习者语料库可以发现，最突出的问题就在于语料数量少且不同质，不能很好地反映学习者的整体水平，无法满足二语习得研究对学习者语料库的需求。

2008年获准立项的国家社会科学基金项目“中国日语学习者语料库的建设与研究(08BYY075)”就是为了解决这个问题而作的探索。该项目的目标是创建一个起点高、规模大、覆盖面广、代表性强的日语学习者语料库，并在此基础上开展了一系列相关研究。通过全面、系统地收集我国高校日语专业学生的作文、翻译语料，该语料库较为客观、翔实、准确地反映了我国日语专业学生的语言发展水平和表达特征，为课程设置、教材编写、评价体系建设以及二语习得、中日对比语言学等领域的研究提供了可靠依据。

本书是该项目的结项专著。在撰写以及后期的修改、成书过程中，笔者还获得了上海外国语大学重大科研项目“基于中国日语学习者语料库的二语习得研究”以及上海外国语大学青年教师科研创新团队项目的资助。本书中汇集了在中国日语学习者语料库建设过程中

开展的各项先导性研究以及部分基于该语料库开展的二语习得研究的阶段性成果，体现了我们在日语学习者语料库平台建设和应用研究两个方面实现的一些突破。在全国哲学社会科学规划办公室组织的结项成果鉴定中，被评为优秀。

首先，日语二语习得研究史的回顾、二语习得量化研究中数据采集方法的对比研究以及日语语料库建设的现状综述等基础性研究系统地回顾和分析了目前相关领域的研究现状，对于二语习得以及语料库应用研究领域的研究者具有一定的启示作用。

其次，围绕着语料库平台体系构建以及自动赋码器精度和抗干扰性等方面开展的一系列先导性研究具有一定的开创性，有效地提高了语料库建设和应用的效率，也为同类语料库的建设提供了有益的借鉴。

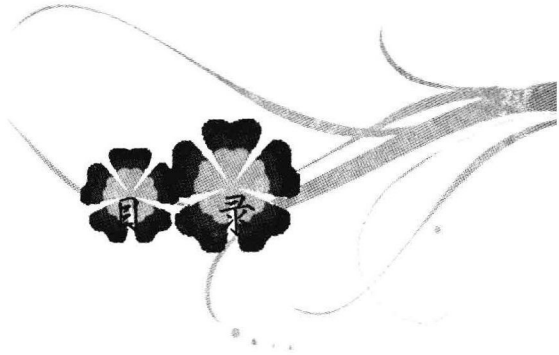
第三，基于学习者产出的一系列实证性研究使我们较为准确地掌握了学习者产出的特征、常用词目以及常见失误。在此基础上，我们可以对照本族语语料库反映出的语言实际使用状况，在教学过程中对部分语言现象进行有意识的强化、抑制或修正，帮助学习者的语言体系更加顺利地向目标语言靠拢，有效地避免或者延缓二语习得过程中误用、石化、磨损等现象的发生，改善语言教学的效率和效果。

当然，受学识和时间、精力所限，本书在具体的论述过程中，难免存在疏漏偏颇之处，恳请广大研究者不吝指正。

在本书的撰写过程中，得到了高等学校外语专业考试办公室、上海外国语大学相关部处、上海外国语大学日本文化经济学院以及各位前辈学者的大力支持。在此，谨表诚挚谢意。

著　者

2012 年 1 月



第一章

日语二语习得研究史回顾

1. 引　言

二语习得研究是一门新兴学科，目的在于考察人们在掌握母语后获得第二语言的过程及规律。它的兴起源于外语教育的需要。作为一门独立学科，二语习得研究大概形成于20世纪60年代末、70年代初(Block 2003)。与其他社会科学相比，它是一个崭新的领域。在研究过程中，研究者大多借用了母语研究、教育学研究或其他相关学科的方法(Larsen-Freeman & Long1991)。同时，又深受行为主义心理学、普遍语法、语言类型学等相关研究的影响。

在过去的50多年时间里，该领域研究者秉承的理念不断发生着变化。从研究方法来看，主要经历了语言对比研究、学习者失误研究和中介语研究三个阶段。研究对象从音韵、形态素、词汇、语法等语言本体层面的研究逐渐拓展到了对语体混淆、呼应不当、交际策略运用错误等语用层面的研究。从研究素材看，则从初期的个案分析逐步发展到当今利用大规模语料库开展实证性研究的阶段。研究的广度和深度不断得到拓展，对外语教学和语言学研究的发展作出了巨大贡献。

与欧美的相关研究相比，日语二语习得研究既有共性的一面，又呈现出不同的特点。本章旨在回顾日语二语习得研究的发展历程，考察研究方法演变的总体趋势以及最新进展，总结过去，寻求对将来的启示，以期推动日语二语习得研究不断走向深入。

2. 从语言对比研究到学习者失误分析

如上所述，从研究方法来看，二语习得研究大致经历了语言对比研究、学习者失误研究和中介语研究三个阶段。

最初，受到行为主义心理学影响，研究者着重运用结构语言学的分析性练习方式，主张通过反复加强刺激，促使学习者形成正确的语言反馈习惯，避免表达错误的发生。为了实现这一目标，就必须准确掌握学习者的母语和目标语言的差异。因此，语言对比研究受到了重视。在这一阶段，研究者普遍认为，表达失误由学习者的母语和目标语言的差异引起。因此，两者差异越大，习得越困难，反之则越容易。然而，基于该理论作出的许多预测都被证明与实际情况不符。例如，日语中很多句子省略了主语，因此，日本人在用英语写作或会话时理应容易出现主语的不当省略。但是，对在美国学习的日本大学生的英语作文进行调查后发现，9 成以上的句子中并未出现主语省略现象（Krahnke, Krahnke and Nishimura 1993）。此外，也有研究指出，在不同母语学习者的产出中发现了相同类型的失误。这说明，表达失误并非都由母语干扰引起。由于这些问题的存在，基于语言对比分析的二语习得研究的有效性受到了质疑，并逐渐式微。

由于不同母语学习者的产出中出现了同类型失误，研究者开始将目光聚焦在表达失误上，对学习者产出进行分析、标注，收集表达失误的实例，并探讨其成因（如 Buteau 1970，Dulay and Burt 1972，George 1972 等）。由此，二语习得研究进入了学习者失误研究阶段。研究者对学习者失误的认识也发生了显著变化。在语言对比研究阶段，研究旨在通过反复训练，形成正确的语言反馈，尽量避免失误的发生。而在学习者失误研究中，研究者认为表达失误是必然会产生的，在语言习得过程中，学习者尝试构建各种规则，并不断对照接触到的目标语语料加以检验。如果在这一过程中发生了规则构建错误，学习者的产出中就会出现相应的表达失误。例如，迫田久美子（2002）指出，学习者在表达过程中会注意到自己的错误，并不断加以修正。观察例 1 可以发现，在产出过程中，学习者不断尝试选择正确的表达形式，检验自身假定的各种规则并加以修正。正是通过这一方式，学习者逐渐掌握了这些语言知识。

(1) 父がいま、年がほんとに高いです、年が多いです。(迫田久美子 1998:326)

欧美的学习者失误研究以 Corder(1967)为开端,而能反映 20 世纪 70 年代日语学习者失误研究大致状况的当属《日本语教育》杂志第 34 号。该期杂志组织了学习者失误研究专题,共收录有铃木忍(1978)、吉川武时(1978)等 6 篇文章。从中可以观察到日语学习者失误研究的两大方向。一是研究为教材开发以及改善教学方法服务。一是以学习者失误为鉴,推动词汇、语法等日语本体研究。随后的研究,例如《日本语学》杂志上登载的一系列学习者失误研究论文(吉川武时 1982、森田良行 1983、水谷信子 1984 等)以及小林典子(1987)、佐治圭三(1991)、松本·Sturt 洋子(2003)等研究都沿袭了这样的方向(长友和彦 1993)。

纵观这一时期的日语学习者失误研究,一个共同特征在于,研究者通常以某特定国家的学习者为考察对象,观察这些学习者的母语与日语在音韵、词汇或语法等层面上存在哪些差异,并分析它们与各类产出失误之间的因果联系(如青木晴夫 1980、野泽素子 1985、田窪行则 1987、小林ミナ·Quackenbush 寛子·深田淳1991 等)。从方法论上来看,这些研究仍然沿袭着语言对比分析研究的基本做法,研究焦点集中在由母语干扰引起的各类失误上,而没有如同欧美学界那样,系统地观察二语习得过程中可能产生的各种表达失误。因此,从研究视野的广度来看,不能不说是有所欠缺的。

3. 学习者失误研究的局限和中介语研究的兴起

随着研究的不断深入,学习者失误研究的局限性也逐渐显现出来。遭遇的问题之一就是对于某些正误界限较为微妙的表达形式,研究者的判断往往很难取得一致。例如,对于例 2,Makino and Tsutsui (1995)认为是正确的。而在网页「日本語教師塾」的相关讨论中,Shuji 却指其为错句。此外,也有一些表达形式的问题并非存在于词汇或语法层面。例如,尽管将例 3 改为例 4 可能更为自然,究其本身却很难说是错句。在这种情况下,不同研究者所作的正误判断必然会出现差异。这给各类研究尤其是定量研究的信度和可比性带来了不利影响。

(2) 東京の夏は暑くてならない。

(3) 私は日本語の授業を楽しみました。(迫田久美子 2002 例 16)

(4) 日本語の授業を楽しく受けました。

问题之二就是,对于自己不熟悉的语言表达形式,学习者往往会尽量采取回避策略。这导致在学习者的产出中,常常出现某些表达形式使用频率偏低的现象。如果仅仅以学习者产出中的失误作为分析研究的对象,就会忽略这种情况,导致研究者无法准确认识和把握学习者产出中存在的各种隐性问题。为了客观、全面地观察二语习得的整个过程,我们的视野不能仅仅局限在学习者的失误上,对于产出中的正确表达以及词汇、语法构成等情况也必须进行体系性考察,并与母语使用者进行比较,从而把握各个语言发展阶段学习者语言能力的全貌。由此,二语习得研究进入了中介语研究阶段。

首先提出中介语(interlanguage)这一概念的是 Selinker(1972)。他认为,既然在不同母语学习者的产出中观察到了同类失误,说明在他们的头脑中存在着与母语无关的共通的语言体系。学习者或接受目标语输入,或借助母语知识,对目标语言的规则进行假定。随后,在实际使用中不断检验该规则正确与否并加以适当修正。由此在自己的头脑中构建起目标语言的体系。Selinker 将这种语言体系称为中介语。中介语具有开放性、动态性和系统性等特点。影响中介语形成的因素包括母语迁移、语言规则的过度归纳、训练迁移、交际策略以及二语习得策略的运用等。在整个二语习得过程中,这五个因素共同发生作用,导致学习者头脑中构建起的中介语体系各不相同,并不断发生着变化。自 20 世纪 60 年代末 70 年代初开始,Corder 等一些学者也开始将学习者的语言视为创造性产物。尽管命名各有不同[①],但是,在将学习者的语言看成是向目标语言不断发展的复杂而活跃的语言体系方面,他们是一致的。

随着中介语概念的确立,研究方法逐渐从量化法拓展到了质化法。研究对象从学习者失误拓展到了对中介语的体系性考察。研究内容也从单纯的学习者失误分析发展到了会话分析、谈话分析、语法正误测试和心理学实验等多种形式。典型的量化研究包括问卷调查和实验研究。典型的质化研究则包括个案研究和人种志研究。语篇分析中大多数是

质化分析，也有量化分析，这取决于研究是以文字分析为主，还是以数据统计为主。欧美的二语习得研究方法经历了两次大转变。第一次大转变发生在20世纪80年代中期，量化法进入了成熟期。第二次转变则发生在20世纪90年代后期，质化法进入了成熟期（文秋芳·王立非2004）。从下节分析可知，相比较而言，日语二语习得研究长期停留在语言对比研究和学习者失误分析阶段，在中介语研究方面落后于欧美研究界。

在本章的第4节，笔者将对《日本语教育》杂志从创刊至2011年期间刊登的所有论文进行梳理、统计，旨在考查过去50年间相关领域中各类研究的分布情况、研究理念的发展过程以及总体趋势。

4. 日语教育研究的发展历程和总体趋势

在本节，笔者将对《日本语教育》杂志从创刊至2011年期间刊登的所有论文进行梳理、统计。之所以选用该杂志作为研究对象，首先是因为该杂志是日本语教育学会的会刊，是日本日语教育研究界的权威杂志。该杂志刊登的研究成果都面向日语教育，尤其是面向以外国人为对象开展的各类日语教育。其次，该杂志创刊于1962年，与二语习得研究成为独立学科的时间较为接近。对该杂志刊登的各项研究成果进行统计分析，可以较好地掌握过去50年间日语教育研究特别是二语习得研究的发展过程和总体趋势，以便总结过去，更好地展望未来。

纵观该杂志自创刊以来刊登的各类文章，根据研究方向，大概可以归为8大类，即教学法、日语本体研究、二语习得研究、情况调查、测试评价、教材研究、辞书研究和研究工具。

教学法类论文主要探讨教学方法和课程设计等方面的内容（如黑田巍1963、北条淳子1973、缝部义宪1991等）。日语本体研究指那些以日语中某些表达形式为考察对象的论文（如村木新次郎1982、江田すみれ1991、武藤彩加2001等）。运用问卷调查、参与观察、访谈以及失误分析等方法，或是通过学习者语料观察学习者的习得状况和第二外语获得规律的论文归入二语习得研究类。如本章第2节所述，语言对比类研究，即比较不同语言中的近似表达，探索共同点和差异的研究，通常也服务于二语习得，属于二语习得研究的方法之一。因此，也计入二语习得研究类论文（如古藤友子1987、久保田美子1994、赵萍2008等）。

情况调查类论文包括那些就日语教育发展现状以及教学实施等情况开展的各类调查(如大野喜代治 1982、谷部弘子 1999、小川誉子美 2005等)。需要说明的是,从考察对象上看,部分论文与日语本体研究、教学活动、课程规划以及教科书编撰等内容相关,但并非是作者本人的实践,而是对其他人或组织当前或在历史上开展的相关活动或研究现状等进行的调查、描述,也归入本类。不过,如果论文论述的是作者本人的研究成果、教学实践、课程规划等,则归入日语本体研究或教学法等相应类别。测试评价类论文主要探讨如何对被试者语言方面的各种能力进行衡量(如舆水实 1964、吉川武时 1977、中岛和子·桶谷仁美·铃木美知子1994 等)。关于教材、辞书的呈现形式以及内容编纂等方面的论文分别归入教材研究和辞书研究类(如中西芳绘 1973、日向茂男·清田润1983 和武田祈 1972、平塚真理·副田惠理子2005)。尽管绝对数量不多,但是在《日本语教育》杂志上刊登的论文中,还有一些论文重点阐述的并非具体研究,而是新的研究工具,如第 54 期的个人电脑专题、第 78 期的 CAI 专题以及第 130 期的语料库专题等。在此,全部归入研究工具类。此外,还有部分非研究类文章,如感想、答疑、书评以及讣告等,统一归入其他类。

将《日本语教育》杂志第 1 期至第 144 期刊登的所有文章按以上标准进行分类,再以 10 年为单位[②]进行统计后获得下表。

表 1 《日本语教育》论文分类统计表[③]

	1962－1970	1971－1980	1981－1990	1991－2000	2001－2010
教学法	14(17.5)	52(19.3)	77(18.5)	77(17.0)	76(20.3)
本体研究	32(40.0)	57(21.2)	58(13.9)	142(31.3)	100(26.7)
二语习得	4(5.0)	30(11.2)	82(19.7)	111(24.4)	101(27.0)
情况调查	7(8.8)	74(27.5)	143(34.3)	87(19.2)	52(13.9)
测试评价	2(2.5)	7(2.6)	13(3.1)	12(2.6)	23(6.1)
教材研究	1(1.3)	17(6.3)	21(5.0)	10(2.2)	5(1.3)
辞书研究	1(1.3)	4(1.5)	0(0)	0(0)	1(0.3)
研究工具	0(0)	0(0)	6(1.4)	6(1.3)	9(2.4)
其他	19(23.8)	28(10.4)	17(4.1)	9(2.0)	7(1.9)
合计	80(100)	269(100)	417(100)	454(100)	374(100)

如上表所示，《日本语教育》杂志不同时期刊登的论文数量相差较为悬殊。为了使数据具有可比性，笔者不是以论文的绝对数量，而是以其在所属时间段全部论文中所占比例为标准衡量该类论文的发表情况。将上表中的相关数据制成折线图后得到下图。

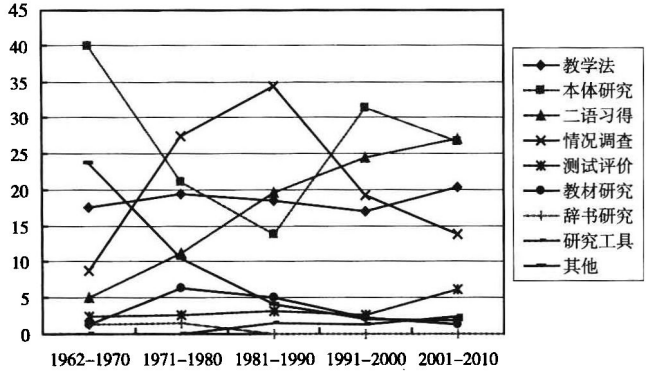


图 1 《日本语教育》论文分类统计图

从上图可以看出，在《日本语教育》问世以来的近 50 年里，教学法类论文所占比例一直较为稳定，始终保持在 18%左右，是《日本语教育》所刊论文的主要研究方向之一。除此之外，不同阶段在该杂志上占主导地位的论文类型则发生了较大变化。在该杂志创办之初，日语教育尚未受到足够重视。不仅投稿数量少，而且刊登的文章中有很多并非是研究性，而是感想类文章。同时，由于正值欧美二语习得理论的萌芽期，所以二语习得类论文也很少，日语本体研究以及教学法类的文章占主导地位。70 年代至 80 年代，随着日本国力的大幅度增强，为了满足对日经济、文化交流的需要，各国逐渐掀起了学习日语的热潮。日本政府也开始通过国际交流基金等组织在各国推广日语教育，以培养日语人才，提高国际影响力。日语教育由此开始受到重视。在这种背景下，不仅二语习得类论文不断增加，反映各国日语教育发展状况的情况调查类文章也异军突起，在该时期发表的论文中占了很大比重。这充分体现了学界关注重心的转移。进入 90 年代之后，日语教育界开始更加关注研究的深度以及新的研究方法、研究手段的运用。因此，情况调查类论文明