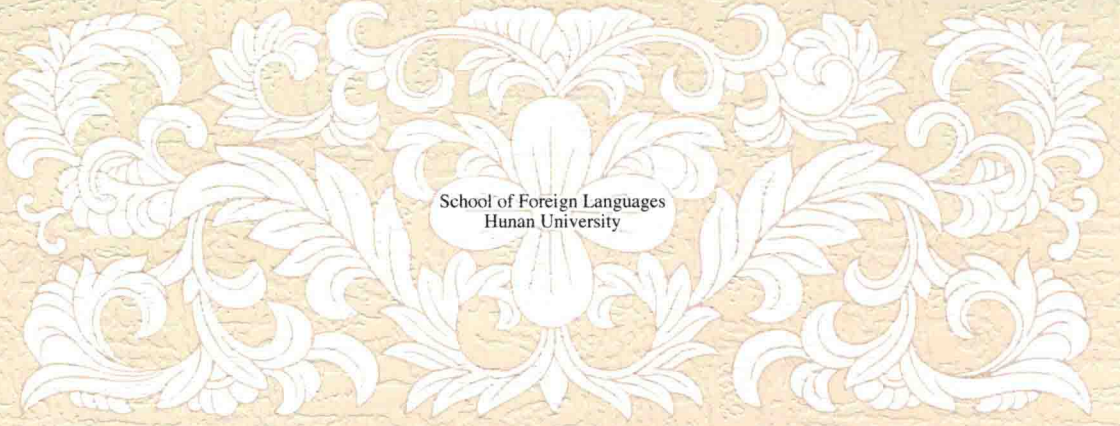




School of Foreign Languages
Hunan University

刘正光 主编

湖南大学外国语学院
新人文话语丛书

语料库与日语复句习得研究

基于用法建构的语言习得理论

苏 鹰 著

中国社会科学出版社

School of Foreign Languages
Hunan University

刘正光 主编

湖南大学外国语学院
新人文话语丛书

语料库与日语复句习得研究

基于用法建构的语言习得理论

苏 鹰 著

中国社会科学出版社

图书在版编目(CIP)数据

语料库与日语复句习得研究:基于用法建构的语言习得理论 / 苏鹰著.

—北京:中国社会科学出版社,2018. 10

ISBN 978-7-5203-3348-1

I. ①语… II. ①苏… III. ①日语-复句-研究 IV. ①H364. 3

中国版本图书馆 CIP 数据核字(2018)第 242763 号

出 版 人 赵剑英
责任编辑 曲弘梅
责任校对 郝阳洋
责任印制 戴 宽

出 版 中国社会科学出版社
社 址 北京鼓楼西大街甲 158 号
邮 编 100720
网 址 <http://www.csspw.cn>
发 行 部 010-84083685
门 市 部 010-84029450
经 销 新华书店及其他书店

印刷装订 北京君升印刷有限公司
版 次 2018 年 10 月第 1 版
印 次 2018 年 10 月第 1 次印刷

开 本 710×1000 1/16
印 张 12
插 页 2
字 数 181 千字
定 价 48.00 元

凡购买中国社会科学出版社图书,如有质量问题请与本社营销中心联系调换
电话:010-84083683
版权所有 侵权必究

本书系教育部人文社会科学研究项目

“汉日复句构式的跨语言认知研究——以条件、因果复句为中心”

(16YJC740062) 阶段性成果

丛书总序

“新人文话语”是湖南大学外国语言文学学科推出的系列开放性丛书。本丛书2007年推出第一批，由湖南人民出版社出版，共五部，即《隐喻的认知研究：理论与实践》（刘正光著）、《二语习得与外语教学》（肖云南著）、《翻译：跨文化解释》（朱健平著）、《华莱士·史蒂文斯诗学研究》（黄晓燕著）、《亨利·詹姆斯小说理论与实践研究》（王敏琴著）。推出这个丛书的最初想法是，鼓励老师们潜心学术研究，助力学院学科发展。

转眼到了2017年，过去的十年见证了本学科快速发展。2007年的时候，本学科的教授不到十名，具有博士学位的教师也不多。十年后的今天，本学科在2012年国务院学科评估排名中并列第十七位，2017年在“软科中国最好学科”排名中位列第十二，排在前10%的第一位。

湖南大学外国语学院有着悠久的办学历史，其最早可追溯到岳麓书院创建的译学会（1897）。1912—1917年，岳麓书院演进到时务学堂以及湖南高等师范学校后，正式设立英语预科和本科部。1926年创建湖南大学外国语言文学系，2000年组建外国语学院。陈逵、黎锦熙、杨树达、金克木、林汝昌、周炎辉、徐烈炯、宁春岩等许多知名学者先后执教于此。

经过几代人的不懈努力，本学科凝练成了理论语言学、应用语言学、文学与文化、翻译学四个稳定且颇具特色的研究方向。

理论语言学以当代认知科学理论为背景，以语言与认知的关系研究为重点，以认知与语言交叉研究为基本范式，寻找认知发展、语言

认知机制与语言本体之间的内在联系。

应用语言学以语言测试、二语习得与外语教学为研究重点,强调研究成果在学生能力培养中的实际应用。

文学与文化以小说与诗歌创作理论和生态诗学为重点,紧跟全球化语境下文学理论与文化批评理论的研究前沿,探索文学、文化、政治、历史话语之间的互动关系。

翻译学以哲学、文学理论、文化理论、认知科学、语言学等为理论基础,探索本土文化对异域文化的接受历程和中国文化在西方世界的旅行轨迹,阐释翻译与认知的内在关系以及翻译理论、翻译实践与翻译教学的互动关系。

学科进步的重要标志是人才培养质量和高水平的科学研究。本学科聚集了一大批学术能力很强、潜心研究的中青年学者。“70后”贺川生教授在 *Synthese* (当年哲学类排名第一)、*Syntax*、*Lingua* 等一批国际一流期刊上发表的系列论文引起了国内外学术界的高度关注;“80后”王莹莹教授在 *Language* 和 *Journal of Semantics* 等重要期刊上发表了一系列语义学论文;全国“百优”田臻(“80后”)在英汉语构式对比研究方面取得了令人瞩目的研究成果。

近年来,本学科中青年学者学术研究成果丰富,为此,学院决定继续定期推出第二批和第三批研究成果,做到成熟一批,推出一批。

第二批推出六部著作,分别是《汉语儿童早期范畴分类能力的发展研究》(曾涛著)、《违实条件句:哲学阐释及语义解读》(余小强著)、《语料库与日语复句习得研究》(苏鹰著)、《京都学派——青木正儿的中国文学研究》(曹莉著)、《〈楚辞〉英译研究——基于文化人类学整体论的视角》(张娴著)、《语言边界》(艾朝阳著)。这些著作中,除第六部外,其余的作者均为“70后”,体现出学科持续发展的坚实基础和潜力。

第三批著作也在酝酿之中。作者的主要群体也许将是“80后”了,他们承载着本学科的未来和希望。相信也还会有第四批、第五批……

湖南大学外国语言文学学科的快速、健康发展，得到了各兄弟院校和各界朋友的大力支持。为此，我们衷心感谢，同时也恳请继续呵护我们成长。

是为序。

刘正光

于湖南大学外国语学院双梧斋

2017年12月12日

目 录

第一章 语言研究中的日语语料库	(1)
第一节 语言研究中的语料库	(1)
第二节 语料库的分类	(2)
第三节 母语者书面语语料库	(3)
一 太阳语料库	(3)
二 青空文库	(4)
三 现代日语均衡语料库 BCCWJ	(5)
第四节 母语者口语语料库	(6)
第五节 学习者口语语料库	(7)
一 KY 口语语料库	(7)
二 C-JAS	(9)
三 鲁东大学会话语料库	(11)
四 日本国立国语研究所会话对照语料库	(12)
第六节 学习者书面语语料库	(12)
一 东吴大学作文语料库	(12)
二 华东政法大学作文语料库	(16)
三 日本国立国语研究所作文对译语料库	(17)
四 湖南大学中国人日语学习者中介语语料库 (HUN-CJS)	(18)
第七节 中日对译语料库	(20)
第八节 小结	(24)

第二章 湖南大学中国人日语学习者中介语语料库

(HUN-CJS)	(26)
第一节 语料库介绍	(26)
第二节 语料库内容	(26)
一 口语语料	(26)
二 书面语语料	(27)
第三节 抽样书面语语料特征	(27)
一 基于语料 A 的考察分析	(30)
二 基于语料 B 的考察分析	(34)
第四节 小结	(40)
第三章 HUN-CJS 语料库中学习者接续表现使用的调查	(42)
第一节 本章研究目的及内容	(42)
第二节 本章的分析数据	(43)
第三节 接续表现在语料库中的使用情况	(44)
一 未被使用的接续表现	(45)
二 被使用的接续表现	(47)
第四节 「ガ」「ト」「カラ」的使用情况分析	(48)
一 整体数据分析	(49)
二 分阶段考察分析	(50)
第五节 小结	(55)
第四章 在单句—复句—语篇连续体中存在的复句	(57)
第一节 单句到复句到语篇的连续性思考	(57)
第二节 复句内部的连续性	(59)
第三节 复句中的主从复句	(62)
第四节 理论基础	(64)
一 主体性意义和客体性意义	(64)
二 从事物关系到事件关系的认知研究	(65)
三 事件间意义连续的认知研究	(67)

第五节 复句的认知模型及其扩展过程	(70)
一 时间复句	(72)
二 因果复句	(75)
三 条件复句	(77)
四 转折复句	(80)
第六节 小结	(82)
第五章 复句习得的相关先行研究	(85)
第一节 复句的习得研究	(85)
第二节 日本的接续助词习得研究	(86)
一 研究数据	(86)
二 研究内容	(86)
三 研究方法	(88)
第三节 中国的日语接续助词习得研究	(91)
一 研究数据	(92)
二 研究内容	(92)
三 研究方法	(93)
第四节 先行研究的问题点	(94)
一 研究语料的问题点	(94)
二 研究内容的问题点	(95)
三 研究方法的问题点	(95)
第五节 小结	(96)
第六章 复句习得研究的研究对象、目的及理论背景	(98)
第一节 研究对象	(99)
一 学习者的中介语	(99)
二 成人 JFL 的语言习得	(100)
第二节 研究目的	(102)
第三节 理论背景 (之一)	(102)

一 基于用法建构的语言习得理论 (usage-based model)	(103)
二 基于用法建构的语言习得理论运用于二语习得的可能性	(104)
三 基于用法建构的语言习得理论的应用	(106)
第四节 理论背景 (之二)	(108)
一 图式	(108)
二 固定表达	(110)
三 固定表达的习得与复句的习得	(111)
第五节 小结	(113)
第七章 复句习得研究之一	
——典型条件表现「バ」的习得	(115)
第一节 以条件复句为中心的复句习得研究	(115)
第二节 典型条件表现「バ」的习得	(117)
一 教科书中的语法提示	(118)
二 「なければ」类	(120)
三 动词与接续助词「バ」的搭配	(125)
第三节 小结	(134)
第八章 复句习得研究之二	
——条件表现「ト・タラ・ナラ」的习得	(136)
第一节 条件表现「ト」的习得	(138)
一 教科书中的语法提示	(138)
二 によると	(138)
三 N/Vともなると~	(140)
四 「否定+と」类	(141)
五 动词与接续助词「ト」的搭配	(142)
六 小结	(152)
第二节 条件表现「タラ」的习得	(153)

一 教科书中的语法提示	(153)
二 动词与接续助词「タラ」的搭配	(153)
三 小结	(163)
第三节 条件表现「ナラ」的习得	(164)
一 教科书中的语法提示	(164)
二 名词与接续助词「ナラ」的搭配	(166)
三 动词与接续助词「ナラ」的搭配	(167)
四 小结	(168)
第九章 关于复句习得的几点思考	(169)
第一节 固定表达在复句习得中的作用	(169)
第二节 学习者如何到达复句构式的习得	(170)
第三节 其他影响因素	(170)
第四节 复句习得模式的假说	(171)
参考文献	(173)
后记	(181)

第一章

语言研究中的日语语料库

第一节 语言研究中的语料库

一般来说,语料库就是一定真实语言样本的集合(常宝宝,2009;苏金智,2012;何婷婷,2003等)。常(2009)、苏(2012)、何(2003)认为语料库还应该具备以下三个特征:

①语料库的语料应该具备一定规模;

②语料集合的构成应该保证均衡;

③语料库的语料应该是电子化处理过的语料具备能被计算机读取的能力。

此外,常(2009)认为语料库一个必备的特征是语言材料必须经过适当的标注和加工处理,例如词语切分、词类标注处理等。但笔者认为这并不是语料库一个必需的特征,因为目的不同决定了语料的处理方式和程度不同,很多语料并不需要上述词语切分等处理,例如全文语料库的每一个语料都是一篇文章的全文。

语料库在研究中起到的是辅助作用。具体例如,语料库可以使研究人员从手动搜集实例替换为用机器来搜集实例,或是从机器的已存数据中提取实例。这大大减少了研究人员在寻找实例中所消耗的精力,也正是语料库的优点所在。语料库所提供的实例经过有规则的整理后,在寻找速度和精确度上更有保障。提供的实例还可以单独处理,这样能够更充分有效地起到例证作用,使研究更为准确。简而言之,语料库的使用可以提高语言研究的质与量,极大程度上增加研究

成果的产出。

收集语料的另一种传统方式是问卷调查。问卷调查至今仍是人文社会科学研究的重要方法。问卷调查过程中所实施的都是结构化的调查问卷。所谓“结构化”，是指问卷中的问题及其备选答案都是事先设计好的。这些结构化了的问卷调查，在实施过程中难免会对答题人产生暗示，而事先设计好的备选答案也难免会给答题人一些主观引导，使得调查结果的有效性降低。此外，问卷调查通常收集的是一个时点上的语料，即要求所有被调查者在同一时间或者大体在同一时间回答问卷中的问题。这就很难为历时研究所用。相比而言，使用不同语料库收集语料可以克服上面的这些问题，使得抽取出来的语料更加客观、更能适用多种研究。

第二节 语料库的分类

语料库的核心是语料，根据语料库含有语料数量的不同，可分为单语言语料库和多语言语料库。

如果一个多语言语料库，它的语料是具有对译关系的两种或两种以上语言的话，这种语料库就是平行语料库。与之相对，不具有对译关系的不同语言的语料所构成的多语言语料库，则是非平行语料库。

如果语料库的语料是来源于母语者，那么这种语料库可称为母语者语料库；如果语料库的语料来源于学习者，这种语料库可称为学习者语料库。

语料具有时间的客观属性。如果语料是在一段跨度较长时期内收集的，该语料库可称为历时语料库。如果语料是在某个特定时期内收集的，该语料库可称为共时语料库。

语料的存储方式有两种，正如我们生活中使用书面语和口语两种交流方式一样，语料也分为书面语和口语两种形式。基于这两种语料的语料库分别叫作书面语语料库和口语语料库。

随着计算机技术的进步,能存储的数据的形式也越来越丰富。由此诞生了能存储图片、视频、音频及多种形式数据的语料库,这种语料库叫多媒体语料库。在语言研究越来越“视觉化”的同时,语料库也逐渐扩充内容,将说话人的手势、目光、脸部表情、情感等要素纳入其中,呈现一种“多模态化”的趋势,现在一般称之为多模态语料库(刘剑,2017)。这类语料库将在今后的人工智能、远程医疗诊断与康复等领域有巨大研究空间。

在语言研究领域,语料库为了满足研究者及其他用户的需求,也在不断向前发展,比如给语料加上词性、词频等标注,从而给用户更多的信息,这种语料库叫作标注语料库。相对于标注语料库,没有经过加工处理的语料库叫作未处理语料库。

接下来,我们将根据以上语料库的分类,结合现有的相关日语语料库分别进行说明。

第三节 母语者书面语语料库

一 太阳语料库

太阳语料库是日语母语者书面语历时语料库,也是日语近代语语料库。该语料库在2005年公开的时候,收录了从明治后期至大正期间的1895年(明治28年)、1901年(明治34年)、1909年(明治42年)、1917年(大正6年)、1925年(大正14年)刊登于博文馆发行的月刊『太陽』杂志上的3400篇作品,共计1450万字,作者人数达到1000人左右。所有语料都保存为xml文件形式,其中添加了各种语法信息。此外,还提供了相应的检索软件「ひまわり」。

太阳语料库是以CD-ROM形式由博文馆新社刊公开发售,内容包括CD-ROM一张,说明书一本。和太阳语料库一起储存在CD-ROM里的是旧版「ひまわり」(ver. 1.1)软件,旧版「ひまわり」不适用于Windows Vista、Windows7和Windows8,如果需要在上述操

作系统使用太阳语料库,还需要在网上下载免费公开的新版「ひまわり」(ver. 1.5)才能使用。

2016年10月26日开始,该语料库的数据经过更新,以及形态素解析,现在变更为「日本語歴史コーパス」的「明治・大正編」雑誌」的一部分,在后面所提到的现代日语均衡语料库BCCWJ的中纳言页面中公开。

二 青空文库

青空文库是1997年7月开始创建的母语者书面语历时语料库。文库中收录了大量版权失效的日文图书。这些图书通过大量志愿者的共同努力,被制作成电子版图书发布在网络上供所有人免费阅读和下载。志愿者们的这一工作目前仍在继续。

一般来说,作家享有其创作出来的文字作品的著作权。在日本,作家享有该著作权的时间从该文字作品被创作出来后开始计算,直到作者去世后50年。因此在青空文库中,明治时期到昭和初期丧失了著作权的作品占了藏书的大部分。其中也有部分翻译作品,以及作者自身允许公众无偿阅览的现代作品。制作文本文件时所使用的书籍,一部分带有插画或者图画,但因为文章作者与插画或图画作者并非一人,且著作权保护期限不同的缘故,几乎所有作品都不得不只公开文章。

青空文库中的作品从政治哲学到自然科学,收录作品类型范围广泛,其中比较文学作品(包括时代小说、侦探小说等娱乐作品)较多。虽不能说过去的有名的作品都有收集,但日语作品已经相当丰富。翻译作品因为译者著作权问题数量还比较少。2011年3月15日收录作品数达到1万部左右。截至2016年11月11日,收录作品数为13896部,其中著作权失效的有13626部,著作权有效的270部。在青空文库志愿者的努力下,现在这一数字还在持续增加中。

青空文库中收录的名家中,较早期的有森鸥外、夏目漱石、芥川龙之介等,近代的则有中岛敦(1942年逝世)、太宰治(1948年逝

世)、林芙美子(1951年逝世)、宫本百合子(1951年逝世)、堀辰雄(1953年逝世)、坂口安吾(1955年逝世)、高村光太郎(1956年逝世)、永井荷风(1959年逝世)等。

作品由志愿者们使用 JIS X 0208 字符集^①将作品电子化为青空文库文本文件或者 HTML 文件。正如『青空文庫収録ファイルの取り扱い基準』(《青空文库收录文件处理基准》)开篇提到的「人の思いや考えに、より多くの人が、より自由にふれられるようにと願って、青空文庫を開きました」(为了让更多的人更自由地接触到其他人的思考及想法,我们制作了青空文库),青空文库中的文件,在遵守该处理基准的前提下,不论是有偿还是无偿的,都可以自由复印、转发、共享。同时,也可以自由地以青空文库中的文件为基础进行登台表演、口述、改编等活动。

青空文库的文件下载后,使用起来非常方便,在电子计算机、PDA、移动电话等载体上都可以使用。也可以与阅读软件组合使用,或者还可以改造为适于盲人阅读的文件,等等。

青空文库语料库自带简单的检索功能,用户可以通过作家名、作品名或作品的种类来查找文章。青空文库的网址是: <http://www.aozora.gr.jp/>。

三 现代日语均衡语料库 BCCWJ

现代日语均衡语料库 BCCWJ 是日本「大学共同利用機関法人人間文化研究機構国立国語研究所」为主开发的日语最大型的语料库。设计制作现代日语均衡语料库最初的目的是把握现代(明治时期以后)日语书面语的整体情况。为了达到不同文体间的均衡性,语料库收录语料的种类很多,包括书籍、杂志、新闻、白皮书、博客、电子传告栏、教科书、法律等,共计一亿四百三十万个词。其中,图片依赖性较高的漫画和广告没有收录。这也可以说是该语料库的一个遗

① JIS X 0208 是计算机中的一种字符集。