NIHONGO NO GAKUSHU TO KENKYU 日語学習与研究
2020年 第2期 总207号

基于语料库的中国日语学习者有对自他动词偏误研究

上海交通大学 外国语学院　　于　杨

[摘　要]　本文以“中国日语学习者语料库”为基本素材，运用统计学方法，通过考察学习者常见偏误的出现频率、高频偏误词及偏误类型等数据来把握中国学习者习得有对自他动词的情况。考察结果发现：学习者在初级阶段输入的词汇，随着学习时间的推进习得水平有所提高。但是同时随着新的有对自他动词的大量输入，词汇使用幅度的拓宽，从整体来看在中高级阶段的学习者出现了更多的偏误。在偏误类型上，较为突出的问题是助词和动词的不匹配问题。这表明在学习者使用自动词表达事态结果还是使用他动词表达主观意图的意识之前，谓语中心词本身到底是自动词还是他动词，即学习者是否能正确区分有对自他动词本身成为了习得的巨大障碍。

[关键词]　有对自他动词　偏误频率　高频偏误词

引　言

日语中有诸多如「止まる－止める」「焼ける－焼く」「壊れる－壊す」这样词干相同、词尾不同、形成了自他对应的动词。某一组自动词和他动词在形态、意义和句法形式上都能够形成对应[1]，可以称之为“有对自他动词”。

有对自他动词在动词中占很大比重，越是基本的动词越容易出现成对的自他对应动词[2]。川濑（1986）调查了日语中350个常用自动词，其中170个，即接近半数的常用自动词有对应的他动词。岛田（1979）整理了《当用汉字音训表》中的1040个训读词，发现其中有580词（55.8%）形成了有对自他动词。由此可见掌握有对自他动词的重要性：可以说掌握了有对自他动词就是掌握了日常生活中最基础的动词。

但是由于汉语并没有自动词和他动词之分，日语有对自他动词在形式上的相似性无疑会令中国学习者感到困惑。另一方面，日语的自动词和他动词需要基于词性来搭配相应的格助词标志，而汉语中没有助词，这也为学习者习得有对自他动词增加了难度。

本研究对“中国日语学习者语料库”进行实证考察，运用统计学方法探析中国日语学习者在有对自他动词的实际语言产出中出现了何种偏误，其偏误频率、高频偏误词、偏误类型有何特点，以期明晰学习者在习得有对自他动词上的难点所在。

1　关于有对自他动词偏误研究的文献综述

相对于有对自他动词的本体研究，关于有对自他动词的习得研究起步较晚，近年来才开始受到学者们广泛的关注。迄今为止的研究根据研究方法大致可以分为偏误类型概括型研究、实验调查型研究和自然产出型研究三种类型。

1.1　偏误类型概括型研究

偏误类型概括型研究对学习者容易出现的一些偏误现象进行了概括总结。吴珺（2006）指出关于有对自他动词出现的偏误通常是与可能表现相

关的偏误，与使役表现相关的偏误以及由社会文化习惯的不同所引起的偏误。以及可能表现相关的偏误原因主要是自动词表示结果可能的使用。与使役表现相关的偏误主要是由日语的使役表现同汉语的“让”不完全对等导致的。由社会文化习惯不同引起的偏误属于语用范畴，主要体现在自动词描述的是一种自然结果，与当事人无关；他动词则表示一种人为的动作，是负有责任的行为。

虽然吴（2006）系统地指出了学习者在有对自他动词方面容易出现的偏误，但是这种分析只是基于内省，既没有学习者实例的数据支撑，也没有分析到底什么样的学习者在何种状况下会出现这种错误。更缺乏对习得阶段性的考察。

1.2　实验调查型研究

实验调查是研究有对自他动词最常见的方法，守屋三千代（1994）、王冠华（2012）、张鹏等（2012）是其中具有代表性的研究。

守屋（1994）的研究对象是完成了初级或中级前半阶段学习的日本大学1、2年级学生或具有同等日语程度的进修生，以助词和自他动词搭配选择的问卷形式分别在应该使用自动词、应该使用他动词和两者均可以使用的情况下考察学习者选择的倾向性。其结果认为自动词的选择是习得最困难的地方，并从事态的人为性、行为主体的明确性、对行为的实现或意图是否关心三个角度划分了自动词习得难易程度的层次。

王（2012）优化了守屋（1994）问卷调查的选项设置，增加了一些有可能态和被动态选项的题目。其结论是自动词习得困难主要体现在由外力引起的非人为事态、表示行为的结果状态以及表示事态实现的可能性三种场合。另外通过后续访谈指出除了母语干扰和对异文化的认知不足等因素之外，以及针对性的用例过少也是导致习得困难的原因。

张（2012）的研究主要基于有对自他动词的下位分类以统计学的方法进行数据分析，是实证性研究的代表。其证明了根据“派生模式”和“接辞类型”的不同，学习者有对自他动词的习得状况存在显著差异，其中他动化派生动词对以及由“(u)-er(u)”接辞构成的动词对是习得上的难点。

上述实验调查型研究各有特色，但存在以下几点问题。守屋（1994）的问卷测试虽然有120名参加者，但完成了初级和中级前半阶段学习的参加者接触有对自他动词的数量有限，对有对自他动词的认知尚处于基础阶段。王（2012）的问卷对象是某大学二三年级学生共91人，能够横向观察学习者的习得情况，但无法纵向考察习得的进步。张等（2012）的实验对象是某大学日语专业本科学生61人，对于统计学定量研究方法而言人数略少，无法突出定量研究的优势。

1.3　自然产出型研究

上述对在实验环境下收集的诱导性数据进行分析的研究数量众多，与之相比，鲜有研究涉及学习者自然产出的语言数据，管窥之见，主要是姚艳玲（2004）、中石ゆうこ（2005）和伊藤秀明（2012）的相关研究。这些研究着眼点各有不同，但研究对象都是“KY口语语料库”。

姚（2004）通过对比大学日语专业二年级29名日语学习者的88篇作文语料和KY语料库中以汉语为母语的学习者的口语语料，发现作文语料中在“导致事态的发生”这一情况下，由汉语的他动结构引发的直译偏误非常显著。而口语语料则不凸显该特点，更多问题出现在自他动词表达是否恰当。中石（2005）对KY语料库考察的结果表明：初级阶段的学习者几乎没有偏误，到了中级阶段开始出现偏误问题，而到了高级阶段以汉语为母语的学习者依然出现偏误，且形成了自他动词固定以某种活用形出现的倾向。伊藤（2012）则着重于KY语料库中中国学习者中级到超级这一阶段。其中中级学习者的主要问题在于形态，而高级和超级学习者则是在意义和语用范畴下的偏误。

上述研究都是基于学习者自然产出的实例进行分析，对于把握学习者真正的习得状况具有重要意义。但由于KY口语语料库中出现的有对自他动词偏误的例句仅有58处，而伊藤（2012）和姚

（2004）又只抽取了其中中国学习者的数据，从语料库统计的角度来说，未能发挥语料库在定量分析上的优势，其结果是这些研究仅停留在罗列了偏误例句，针对偏误例句进行性质的分析。也就是说，这些研究只能归类为以学习者的实际产出作为材料进行定性分析的研究，难以发现学习者的习得倾向性。

2 本论文的研究课题和研究方法

2.1 本研究的重点

文献综述的主要结论是自动词的习得难度大于他动词，因此本研究首先希望通过数据验证该结论是否正确。另外，还需要解决在有对自他动词的偏误中，哪种动词的偏误率更高，习得难度更大。而且需要明确随着学习的深入，偏误是否会有所减少，也就是习得是否会进步。综上所述，本研究主要期待解决的问题如下：

RQ1）有对自他动词的偏误是否会随着学习的深入而减少，习得是否有进步。其中有对自动词的偏误是否会比他动词更多。

RQ2）不同习得水平的学习者出现的高频偏误词的特点。哪些有对自他动词使用难度更高。

RQ3）学习者自然产出出现的偏误类型有哪些。

通过考察上述RQ，对学习者在有对自他动词方面自然产出的偏误情况进行描述性统计，以期综合把握学习者的习得状况。

2.2 本研究所使用的语料库

本研究所使用的学习者语料库是“中国日语学习者语料库”（以下简称CJLC）。CJLC采用随机抽样的方法，从2007-2009年全国日语专业四级（以下简称L4）和专业八级（以下简称L8）的所有命题作文中各抽取400篇作为语料。

表1是CJLC单词数一览表。如表1所示，CJLC单词总数为568，549词，各阶段各年度取词数量大体相同，可以说是现有学习者语料库中词汇数较多、选取规则较科学、具有均衡性的语料库。另外，CJLC中的作文语料均由日本老师进行批改和修正，其中的偏误有赋码标注。

表1 中国日语学习者语料库单词数一览表

语料	中国日语学习者语料库						总计
	2007-L4	2007-L8	2008-L4	2008-L8	2009-L4	2009-L8	
篇数	400	400	400	400	400	400	
单词数	87,004	107,943	83,591	102,428	82,338	105,245	568，549

资料来源：引自毛文伟，2012：74。

2.3 数据抽取与处理

由于CJLC是对偏误进行了赋码标注的语料库，从“检索-错误检索”的“文法-自立语-动词的自他”项目下，检索出所有与动词自他相关的偏误。将检索出的例句结果下载，再通过人工将例句与原文一一对照，检验所有偏误例句是否与有对自他动词相关，对偏误的标注和偏误判断是否正确统一等问题进行降噪操作。

3 结果与考察偏误频率

3.1 统计结果

经过2.3节所述数据处理方式后得到的日语学习者在有对动词自他方面的偏误共209例。按照级别和错误类型归类为表2。其中L4出现的偏误有71例，L8出现的偏误为138例。

表2 CJLC有对自他动词偏误数

级别＼类别	自动词→他动词	他动词→自动词	按级别总计
L4	45（63.4%）	26（36.6%）	71
L8	78（56.5%）	60（43.5%）	138
按词性总计	123	86	209

注：作者依据CJLC语料库数据绘制。

虽然在表2的原始偏误频数中L8的总体偏误数要高于L4的总体偏误数，但尚不能够说明L8学习者在有对自他动词的习得方面不如L4学习者。用KH Coder对CJLC进行形态素解析得到L4语料中动词的使用总数为29944词，L8语料中动词的使用总数为37539词。将偏误频数与动词使

用总数做频度调整后用SPSS21进行卡方检验，得到表3的结果。统计结果表明，在显著性水平α=0.01下，p值为0.002，低于显著性水平，可以说明L4和L8学习者在有对自他动词的偏误频率上存在显著差异。换句话说，L8学习者在有对自他动词上的偏误比L4学习者更多。

表3　偏误数与级别卡方检验结果

	值	df	渐进Sig.（双侧）	精确Sig.（双侧）	精确Sig.（单侧）
Pearson卡方	9.189	1	0.002		
连续校正[b]	8.771	1	0.003		
似然比	9.410	1	0.002		
Fisher的精确检验				0.003	0.002
线性和线性组合	9.189	1	0.002		
有效案例中的N	67483				

注：作者依据CJLC语料库数据绘制。

另外，表2中自动词的偏误数高于他动词的偏误数量。本应使用他动词而误用了自动词的偏误有123例，反之本应使用自动词而误用了他动词的偏误为86例。而且从不同学习阶段来看，在L4中应使用他动词而误用了自动词的有45句，应使用自动词而误用了他动词的有26句，自动词的偏误率较高。同样的倾向也显现在L8中，偏误数分别是78例和60例。无论是在整体还是不同级别上，从绝对偏误数量来看都是自动词偏误较多。但是以有对自动词实际使用数（2674词）和有对他动词实际使用数（1593词）在显著性水平α=0.01下对整体和L4、L8级别分别进行卡方检验得到的p值为0.433、0.166和0.566，均高于显著性水平。说明无论是从整体还是分级别来看，仅就现有的数据无法判断自动词的偏误明显比他动词多，也就是说尚无法得出自动词习得要比他动词习得更困难的结论。

3.2　分析

3.1节中得到的L8学习者偏误数比L4学习者偏误数高，而且该数据具有显著性差异说明随着学习时间的增加，有对自他动词的习得难度有所增大。在英语二语习得研究中，Kellerman(1979)曾提出二语习得的U-形发展模式。学习者在初级、中级、高级三个阶段的错误频率是开始低，然后升，之后又降。而精确度则与之相反，呈U字形。基于U-形习得发展框架，Oshita(2001)提出非宾格陷阱假说。其研究发现在二语习得的早期阶段，学习者不能区分非宾格不及物动词和非作格不及物动词，容易把两类不及物动词都看作是非作格动词，受母语迁移影响最大；到中期阶段，受母语迁移影响下降很多，而到高级阶段又回升不少。

本研究对非宾格、非作格动词不做深入讨论，但是影山（1996）认为非宾格动词相当于日语中「ナル型自動詞」，而有对自他动词对中的自动词大都属于「ナル型自動詞」。那么3.1节的数据可以验证中国学习者在有对自他动词中的习得上基本符合U-形习得发展模式。当然，这并不能说明L4学习者要比L8学习者对有对自他动词的掌握程度更好。

那么为何会出现学习年限越长的学习者偏误数会越多的结果呢？一个可以想象的因素是习得动词的数量之差。在初级阶段，学习者输入的有对自他动词的词汇量和学习者能够自由应用的有对自他动词的词汇量较中级阶段要少很多。虽然有对自他动词在词尾上的形态可以分为十一类之多[3]，但是在初级阶段输入的自他动词词尾种类多是“aru-eru”“ru-su”。对于L4的学习者来说，与其说是按照类型或是从语法意义上来习得有对自他动词，莫不如说是以词汇为单位进行习得的。到了中级阶段随着大量有对自他动词的输入，学习者自然会不自觉地意识到词尾的发音特点，从而形成对有对自他动词的分类归纳。而有对自他动词的词尾变化的丰富性导致这种前期简单的归类规则无法适用到更多动词时，对学习者来说有对自他动词的分类会越来越扑朔迷离。表层的词性再加上有对自他动词在深层的语义认知的区分，进一步增加了其使用难度，从而出现更多偏误。

4　结果与考察高频偏误词

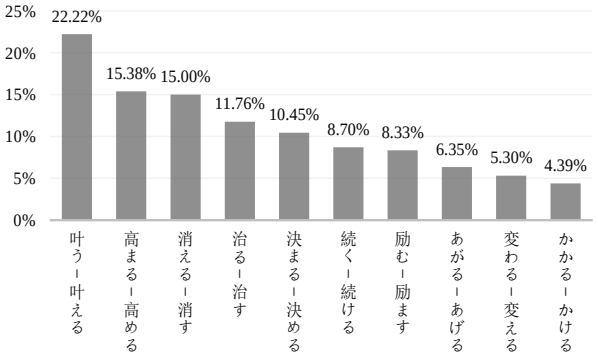
接下来我们要考察的是到底哪些有对自他动词在学习者的自然产出中更容易出错，也就是说

哪些有对自他动词是习得中的难点。

4.1 统计结果

将2.3节抽取的偏误例句以一组自他动词为单位进行偏误词目数量的统计。其中，L4学习者中出现偏误的有对自他动词词目数为35组，L8学习者的偏误词目数为44组。为避免偶然性，将仅偏误一例的有对自他动词组排除后，L4级别中出现偏误的有对自他动词词目数为15组，L8级别的偏误词目数为22组。L8出现的有对自他动词的偏误范围更广。按偏误数除以使用数得到的偏误率排序分别得出L4和L8的高频偏误词。

图1是把通过根据上述计算方法得出的偏误率进行排序而成的L4级别高频偏误词（前十组）。其中偏误率最高的是「叶う－叶える」，偏误率为22.22%。接下来的「高まる－高める」和「消える－消す」偏误率较为接近，分别是15.38%和15%。到排序第十的「かかる－かける」的偏误率降到4.39%。各有对自他动词组之间的偏误程度差别较大。

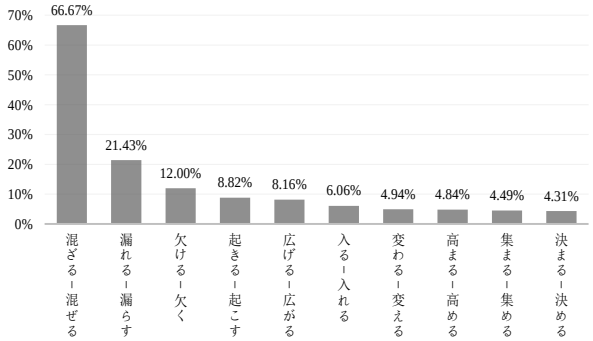


注：作者依据CJLC语料库数据绘制。

图1 按偏误率排序的L4高频偏误词

按照同样方法得出L8级别的高频偏误词如图2。其中偏误率最高的是「混ざる－混ぜる」，达到66.67%，远远高于其他词例。而其他组别从「漏れる－漏らす」到「決まる－決める」的偏误率水平与L4大体相同，偏误率在21.43%至4.31%之间。

此外，对比图1和图2的高频偏误词，除「変わる－変える」「高まる－高める」之外，L4学习者



注：作者依据CJLC语料库数据绘制。

图2 按偏误率排序的L8高频偏误词

和L8学习者容易出现偏误的词例大不相同，从一个侧面反映了L4和L8学习者在不同的阶段对不同的词汇习得程度是不同的。

4.2 分析

基于4.1节的数据可以推出假设：是不是在L4阶段出现的偏误到了L8阶段习得会有所进步呢？为证明这一假设，笔者对将L4和L8中出现的高频偏误词的偏误情况进行横向对比。采用的方法是对4.1节的高频偏误词进行再处理。基于L4中的高频偏误词在L8中的使用数和偏误数，用偏误数除以使用数分别计算出在L4和L8中的偏误比率，然后用L8偏误率减去L4偏误率得到两者的比率差，数据结果归纳为下页表4所示。

下页表4的结果非常明显。除「入る－入れる」这组词外，L8与L4的偏误比率差均为负数，说明L4的高频偏误词到了L8阶段偏误率都有所降低，在L4阶段偏误率较高的词汇组在L8阶段习得程度是有所提高的。也就是说在初级阶段输入的出现偏误的有对自他动词的习得会随着学习的深入有所进步。

但是到了L8阶段，随着更多有对自他动词的输入，学习者一方面需要在词汇方面区分有对自他动词，另一方面需要兼顾其语法意义，正确掌握这些有对自他动词变得更加困难。3.1节中关于L8学习者的偏误率高于L4学习者的偏误率这一事实并不是学习者在习得已知项目中没有进步，而是在习得新项目时受到更多的干扰，对于有对自他动词的认知产生了更多的偏差。

表4　高频偏误词在L4和L8中的横向对比（按L4偏误率排序）

有对自他动词	L4			L8				
	词数	偏误数	偏误比率	词数	偏误数	偏误比率	词数差	比率差
叶う－叶える	9	2	0.2222	49	0	0.0000	40	-0.2222
高まる－高める	13	2	0.1538	62	0	0.0000	49	-0.1538
消える－消す	20	3	0.1500	17	0	0.0000	-3	-0.1500
治る－治す	17	2	0.1176	4	0	0.0000	-13	-0.1176
決まる－決める	67	7	0.1045	116	5	0.0431	49	-0.0614
続く－続ける	69	6	0.0870	70	1	0.0143	1	-0.0727
励む－励ます	24	2	0.0833	24	0	0.0000	0	-0.0833
あがる－あげる	63	4	0.0635	92	3	0.0326	29	-0.0309
変わる－変える	132	7	0.0530	162	8	0.0494	30	-0.0036
かかる－かける	114	5	0.0439	172	6	0.0349	58	-0.0090
残る－残す	60	2	0.0333	37	0	0.0000	-23	-0.0333
増える－増やす	61	2	0.0328	45	0	0.0000	-16	-0.0328
過ぎる－過ごす	107	3	0.0280	158	1	0.0063	51	-0.0217
出る－出す	199	3	0.0151	351	4	0.0114	152	-0.0037
入る－入れる	419	2	0.0048	627	38	0.0606	208	0.0558

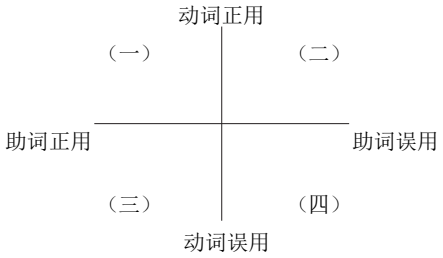
注：作者依据CJLC语料库数据绘制。

5　结果与考察偏误类型统计

5.1　偏误类型

有对自他动词首先作为词汇存在，具有自动词、他动词的基本特征。自动词、他动词的使用在句式结构上要求动词前的格标志与其搭配，他动词以ヲ格为标志，自动词以ガ格为标志。因此助词的搭配使用是考察有对自他动词习得情况的一个层面。而另一方面，在句中是否正确区分并恰当使用了有对自动词和有对他动词则属于动词本身使用层面的问题。

将助词层面和动词使用层面的正用和误用分别作为横轴和纵轴，可以得到图3中有对自他动词使用情况的四个象限。



注：作者依据CJLC语料库数据绘制。

图3　有对自他动词正用误用类型象限图

由于本文主要考察的是有对自他动词的偏误部分，因此本研究统计的对象是象限三和象限四。象限三是助词使用正确，但是用错了动词词性。它体现的是学习者能否意识到有对自他动词的存在以及对有对自他动词各词形的掌握是否准确。属于第三象限的偏误例句如：

（1）本を見てから私に一番印象を残った（→残した）のはある企業家に関する報道だ。（CJLC L4）

（2）違い情報とか、事実でない情報とかよく目に入るのはとても残念で私たちの生活を乱れる（→乱す）に違いない。（CJLC L8）

（3）しかし、日本語を習った後、この考えが変えた（変わった）。（CJLC L4）

（4）文を書く時も、指さきがすこし押すと、大量的な情報が現して（→現れて）くる。（CJLC L8）

在上述例句中，学习者对句式中意图表示事态发展的结果或者是表示动作进行的使用意识并没有偏差。在表示动作意图的地方，使用了ヲ格标志。例（1）和例（2）中的“留下了印象”、“扰乱了生活”均接续了对象语，正确使用了助词。例（3）和例（4）中表达“学习日语后，想法改变了”“手指

用力按下后，就（发现）会有大量信息呈现”。“想法”和“信息”出现在主格位置，正确使用了ガ格助词表述事态的结果变化。上述例句都是在有对自他动词的选择上出现了偏误，混淆了自动词和他动词的词性。

例（5）-（8）是属于第四象限的偏误，即助词和动词均属偏误用法。从词汇层面来看，例句中的日语表述「書く力を高めた」「個人的情報を漏らした」「社会経験が増える」「電話がかかって」的助词和动词的搭配并没有问题，有对自他动词都按照其句法要求搭配了正确的格助词。

（5）私は毎日単語を覚えたり、本文を読んだり、日本語を聞いたりした。それから、日本語がだんだんよくなった。たとえば、聞く力とか、話す力とか、読む力とか、書く力などを高めた（→が高まった）。（CJLC L4）

（6）この間、たくさんの名人の電話番号などの個人的な情報を（→が）インタネットによってもらした（→漏れた）ことが発生した。（CJLC L8）

（7）だから、今の私にとって、実力があがることと社会経験が増える（→を増やす）ことはもっとも重要なことだ。（CJLC L4）

（8）他人にあることを教えたい時、電話がかかって（→をかけて）、国および世界の大事を知りたい時、テレビを開いてすぐ分かる。（CJLC L8）

但是，例句（5）前句已经出现「日本語がだんだんよくなった」，明显要表述“学习水平得到提高”这一事态的结果，并非描述动作主体的意愿，使用他动词「高める」不合适。例（6）中描述的是“发生了个人信息泄露的事件”，虽然不像例（5）一样描述某个特定主体，而是描述非特定主体的动作。但是同样，由于聚焦的不是特定某个动作主体的主观意志性，而是事件本身，因此也应该使用自动词。与其相对，例（7）和例（8）的重点在个人主观意志的改变和行动本身。例（7）的焦点是特定主体主观“加强自身的实力和增加社会经验”使自身发生改变。而例（8）的焦点则是要向他人传达的时候所采取的方法和手段，以打电话的动作来实现目标。这两句均需使用他动词。

象限四的特点是虽然助词和动词的搭配没有问题，但是根据句子的焦点侧重不同，基于视点到底放在哪一层面使学习者产生了应该选择使用自动词还是他动词的犹豫。

5.2 数据统计

根据5.1小节中对偏误类型的总结，再次梳理CJLC中的偏误数据。统计出语料库中出现的助词正用动词误用（第三象限）和助词误用动词误用（第四象限）两种类型的偏误数，并按学习水平进行归类，得到基于偏误类型的数据统计如表5。

表5 按偏误类型统计的偏误数

类别 / 级别	助词正用动词误用	助词误用动词误用	按级别总计
L4	63（88.7%）	8（11.3%）	71
L8	120（87%）	18（13%）	138
按偏误类型总计	183（87.6%）	26（12.4%）	209

注：作者依据CJLC语料库数据绘制。

表5中无论从整体还是分级别来看，助词正用动词误用的偏误频率都要远远大于助词误用动词误用的频率。这说明动词与搭配的助词不匹配是更加突出的问题。学习者尚未分清一个词到底是自动词还是他动词，或者说即使在意识中对于是自动词还是他动词已经能够辨认，但是在选择与其相适应的格助词搭配时还存在较大的问题。

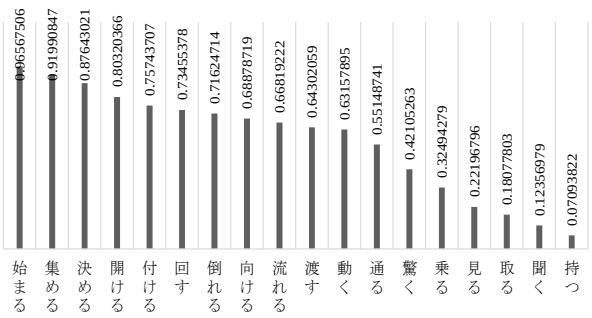
5.3 分析

有对自他动词既是语法项目，也作为独立的动词而存在[4]。相对于其他动词，学习者对有对自他动词的习得首先要解决的是词汇的认知问题，这也是先行研究中最容易被忽略掉的一点。特别对初级和中级学习者而言，在问卷测试前如果没有测试学习者对有对自他动词作为词汇本身的掌握程度，

在学习者尚未辨别有对自他动词的情况下对其进行语法选择性测试，只能是强迫学习者在已有的选项中选取看起来貌似更加正确的选项，而无法反映学习者的真正水平。

从数据统计的结果来看，学习者在有对自他动词作为词汇的这一认知本身还存在较大问题。与其他词汇或是语法项目的习得相比，有对自他动词在词性和发音上的类似性极易引起学习者的辨认混乱从而导致偏误。不仅是初级日语学习者容易出现这样的错误，即便到了高年级、日语水平提高到一定程度的学习者也可能会出现。

在笔者的另一项关于判断动词的词性和是否是有对自他动词的测试调查（有效回答438份）中，计算其中18个动词的认知正确率为下图4。只有前两个词「始まる」和「集める」正确率能达到90%以上。到了「取る」正确率降到18.08%，而「持つ」的正确率仅为7.09%，不足一成，通过调查可知，词间差异非常明显，与上述本研究通过语料库统计的结果相一致。由此可见如何解决学习者的词汇认知问题成为了最突出的问题。



注：作者依据CJLC语料库数据绘制。

图4　有对自他动词自他性和有对性认知正确率

6 结　语

本文主要通过抽取CJLC语料库中出现的有对自他动词的偏误例句，通过偏误频率、偏误高频词和偏误类型等的数据进行统计学分析，对学习者的偏误状况做“面”上的把握。

偏误频率数据表明：L8学习者在有对自他动词上的偏误总体比L4学习者更多，通过对照L4和L8学习者的动词总体使用数进行显著性p值检验，认定该差异具有显著意义。但是这并不能说明L4学习者的习得状况要优于L8学习者，通过对偏误高频词的对比发现在L4学习者偏误率较高的词汇组中，L8学习者的习得程度是有所提高的。但是随着学习的深入，到了L8阶段，由于有对自他动词的数量的大幅增加，学习者使用的有对自他动词数量更多，但是在掌握这些有对自他动词的正确率方面尚有欠缺。上述结果可以说明有对自他动词的习得具有阶段性，但仅就本研究的语料库数据来说，尚不能论证有对自他动词中的自动词的习得是落后的。

另外，在偏误类型上助词和动词的不匹配问题比较突出。表明学习者在出现应该使用自动词表达事态结果还是使用他动词表达主观意图的意识之前，是否能够有效区分自他词性本身成为了习得的巨大障碍。从有对自他动词的下位分类以及使用意识来具化学习者的习得状况是今后需要进一步解决的问题。

[本文为2019年国家社会科学基金“基于数据掘挖技术的中国日语学习者认知机制研究”（项目批准号：19BYY201）的阶段研究成果。项目主持人：毛文伟]

注

[1] 早津恵美子.対応する他動詞のある自動詞の意味的・統語的特徴[J].京都大学言語学研究会.言語学研究，1987（6）：81.

[2] 早津恵美子.有対他動詞と無対他動詞の違いについて—意味的な特徴を中心に[J].言語研究，1989（95）：250.

[3] 庵功雄, 高橋志乃, 中西久実子, 山田敏弘.初級を教える人のための日本語文法ハンドブック[M].東京：スリーエーネットワーク，2000：97.

[4] 中石ゆうこ.対のある自動詞・他動詞の第二言語習得研究—「つく－つける」、「きまる－きめる」、「かわる－かえる」の使用状況をもとに[J].日本語教育，2005（124）：24.

参考文献

毛文伟.中国日语学习者语料库的构建及应用[M].上海：上海外语教育出版社，2012.

王冠华.中国日语学习者有对自动词习得状况的实证研究[J].日语学习与研究，2012（5）.
何午.日语的非能格自动词与非对格自动词[J].外语研究，2005（3）.
吴珺.从日语自动词、他动词的误用探讨母语的干扰[J].日语学习与研究，2006（2）.
张鹏・朱虹・雷鸣.中国日语学习者“自-他对应动词”习得的实证研究[J].外语教学与研究,2012（1）.
川瀬生郎.日本語教科書『日本語初歩』の作成とその問題点[J].日本語教育，1986（59）.
小柳昇.自他動詞の派生対立の分類再考―自動詞と他動詞の両方に現れる「-er-」の位置づけ[J].拓殖大学院言語教育研究，2008（8）.
中石ゆうこ.日本語教育における対のある自動詞・他動詞の研究[J].県立広島大学人間文化学部紀要，2017（12）.
守屋三千代.日本語の自動詞・他動詞の選択条件―習得状況の分析を参考に[J].講座日本語教育，1994.
伊藤秀明.学習者は「対のある自他動詞」をどのように使っているか―中国人日本語学習者の中級から超級に注目して[J].国際日本研究，2012（4）.
島田昌彦.国語における自動詞と他動詞[M].東京：明治書院，1979.
姚艶玲.中国語母語話者の日本語自他動詞の使用実態―作文とKYコーパスの分析を通して[J].東アジア日本語教育・日本文化研究，2004（7）.
影山太郎.動詞意味論―言語と認知の接点[M].東京：くろしお出版，1996.
Kellerman, E. The problem with difficulty[J].Interlanguage Studies Bulletin, 1979(4).
Oshita, H. The Unaccusative Trap in Second Language Acquisition[J].Studies in Second Language Acquisition, 2001(23).

作者简介：于杨（1976—） 女 汉族 上海交通大学外国语学院讲师 上海外国语大学日语语言文学专业博士研究生
研究方向：日语语言研究、日语教育研究
联系方式：E-mail：uyou@sjtu.edu.cn

A Corpus-based Study of Errors with Corresponding Intransitive and Transitive Verb Pairs by Chinese Japanese Learners

Abstract: Based on the CJLC (Chinese Japanese Learner Corpus), this paper makes a comprehensive survey of the frequency of common errors, high frequency errors and types of errors by learners according to learning time using statistical methods. The results show that, on the whole, from the perspective of specific verbs, the vocabulary acquired at the initial stage will improve with the progress of learning and the level of acquisition. However, with the new input of corresponding intransitive and transitive verbs and the widening of learners' usage, learners with higher learning levels as a whole have more errors. The most prominent form of error in these cases is the mismatch of particle and verb. This shows that whether the predicate is the intransitive verb or the transitive verb, in other words, the correct distinction between corresponding intransitive verbs and transitive verbs has become a huge obstacle to acquisition before the intension of expressing the outcome of a situation using intransitive verbs or the formation of subjective intentions using transitive verbs.
Keywords: corresponding intransitive and transitive verbs; frequency of common errors; high frequency errors

Author's Information:
Yu Yang (Female) Year of birth: 1976
Lecturer at Shanghai Jiao Tong University, China
Japanese Linguistics, Japanese Language Education
E-mail: uyou@sjtu.edu.cn