

多国日语学习者中日同形词使用情况考察

程薇

(湖南大学, 湖南省、长沙市, 410012)

摘要: 本文以日语能力考试4级的中日同形词作为研究对象, 使用《多语言母语日语学习者横断语料库》中的口语语料, 通过量的分析, 考察多国日语学习者对中日同形词使用的情况。通过 SPSS 聚类分析进行的量的研究, 得出以下结果: 首先, 研究对象虽然是4级基本词汇, 但是在使用数量上, 学习者和母语者仍然存在差异, 学习者有自己的中介语系统; 其次, “汉字圈”, “非汉字圈”这种与汉字的亲疏和“亚洲圈”, “欧美圈”这种地理位置的远近, 在一定程度上与调查地点的分类存在联系; 最后, 通过对所考察同形词的分类的进一步分析, 得出第二类和第三类同形词学习者经常使用, 不会回避, 且词性差别对分类结果影响较小。

关键词: 中日同形词, 多国日语学习者, 口语语料库, 组间聚类分析, 使用状况

中图分类号: H0

文献标识码: A

1. 引言

众所周知, 在现代日语中, 除了和语词汇外, 还有大量从中国传入的汉语词。因此, 以中文为母语的日语学习者常常被认为在学习日语的汉字词时具备很多有利条件。

迄今为止, 同形词受到很多学者的关注, 也取得了巨大的成果。其中大部分采用的是对照分析的方法, 寻找中日同形词间的区别, 以此为基础, 推测同形词的习得情况。而且, 研究对象多是汉字圈里的学习者, 如中国, 韩国。以下从同形词的对照研究和习得研究两方面来看同形词的先行研究。

1) 同形词的对照研究

(1) 中日同形词的对照研究

对中日同形词的研究开始于20世纪70年代, 中国和日本几乎在同时期开始。《中国語と対応する漢語》(1978 日本文化庁)中, 从语义的角度将中日文对应的同形词分为S(same)、O(overlap)、D(different)、N(nothing)四种类型。此后, 以大河内(1997)为首, 上野(1995)、荒川(1979)等一系列学者对同形词进行了概述, 并对其意思做了区分。此外, 上野(1979)以中日同形词为对象, 从词义入手, 系统的阐述了两语言中的不同, 这些研究结果对中国的日语教育有着非常重要的作用。曲维(1995)根据《日本語教学基本語彙七種比較对照表》, 对二字音读的中日同形词(2063)进行了分类, 其中, 语意大致相同有1805个, 占87.5%。语意一部分相同有206个, 占9.95%, 语意完全不同有52个, 占2.52%。

(2) 韩日汉字词的研究

1972年, 韩国文教部将1800个汉字词定为汉字教育用的基础汉字词, 从汉字字数上比较, 并不比日本常用汉字词数量逊色。之后, 在2014年的中日韩贤人会议上, 发表了808字的《中日韩共同常用汉字表》, 在中日韩三国, 特别是日常不使用汉字的韩国引发大量关注。

在日语和韩语中, 通常认为, 同形词不仅在词语上, 在文法特征上也有很多相似性。但

是在韩语和日语的汉字词中，虽然都有基本的名词词性，但是另一个词性变成动词，形容词或副词的情况也时有发生（李.2002）。有时也存在同时拥有3个以上不同词性的情况。因此，虽然说韩语和日语在汉字词上存在类似性，但是在词性上仍然存在细微的差异。

2) 同形词的习得研究

以中国和多国学习者作为研究对象，对同形词的习得研究做一个简要的说明。

(1) 以中文母语者为研究对象

陈（2003），以问卷调查的形式，让学习者选出目标语（日语）词汇的对应词，以此考察学习者在习得过程中对词汇异同的把握。这种方法考察的是学习者对词汇的理解，并且采用的是调查问卷的形式，不是自然语料的产出，因此这种研究手段偏主观。

大和，玉岡（2009）将不同水平的中国人日语学习者分为上下位群，记录他们对“汉字词相关词的词语性判断题”反应的时间和正确率，以此分析学习者在处理汉字词时母语的影响。结果是，不论对日语词语知识掌握如何，以中文为母语的日语学习者，因为两语言的表记形态是共通的，能够迅速处理汉字。另一方面，日语词汇知识对正确的判断日语有着重要的作用。而且，母语是中文的学习者在对日语的判断时受到了母语的干涉。

以上对中日同形词的研究中，都以对照分析为手段，分析同形词的使用状态，把握学习者的使用情况。但是，母语干涉不是误用的唯一原因，学习者拥有自己的中介语。而且，多数都是对于书面语的研究，对于口语语料的研究很少。因此，本研究通过比较多国学习者和中国学习者的使用情况，探究中日同形词的使用状态和各国学习者的使用特征。

在数据分析方法的方面，在陈（2003）的研究的基础上，庄（2014）调查了词义，词性，词语难度这三个要素对学习者的影响，通过问卷调查，要求35名日语专业的中国学生将中文的同形词翻译成日语，在判断译文正误率的基础上，利用统计学的分散分析方法（One-Way ANOVA），发现要素间的差异在统计学上存在显著差异。通过统计学的手段对习得结果进行分析为本文提供了很大的线索。本研究也将采用统计学的方法，进行数据分析。

(2) 以多国学习者作为研究对象

大塚，林（2010）以汉字圈的中国、韩国；非汉字圈的泰国，印尼，瑞士的中上级日语学习者作为对象，收集学习者以相同话题为主题的作文。对作文中的词汇及汉字词进行分析。结果，以汉字圈（中国人、韩国人），非汉字圈（泰国、印尼、瑞士）的顺序，使用的词汇及汉字词都越来越简单。而且，在分析中国人的错误中，受到中文母语的干涉，直译成中文的情况多次出现。由此证明，学习者在作文中使用的汉字词，很大程度上受到母语的影响。

但是，以上的研究是以作业的形式在家中完成，可以使用电脑书写，因此作文中出现的汉字词可能只是理解、能使用的汉字词，不能客观的反应学习者汉字词的能力。而且学习者的日语能力为日语能力考试1~2级，不是同一个水平的学习者。各母语的研究对象也从30个到1个不等，各母语间学习者相互对比困难。

以上的先行研究，多以汉字圈的学习者为对象，选取部分同形词或者汉字词，利用学习者的作文数据，对学习者的使用情况进行调查。但是影响同形词使用情况的不仅仅是母语干涉，研究方法也不只有对照研究。本文以14个不同调查地点学习者为调查对象，进行量的分析，观察他们中日同形词的使用情况。本文使用日本国立国语研究所的《多言語母語の日本語学習者の横断コーパス：I-JAS》，收集同水平的学习者，4级中日同形词使用的口语数据。将使用情况用SPSS进行聚类分析，对各调查地点学习者和同形词分别进行分类。具体的研究目的有以下三个。

①从不同调查地点看同形词的使用倾向；

②从学习者的使用数量上给各同形词进行分类。

2. 研究概要

本章将对同形词进行定义，并明确研究的范围。在此基础上，说明数据的收集方法。

1) 本研究中同形词的定义和范围

对于“同形词”的定义，不同的研究者各有不同。大河内（1986）认为“同形词就是双方用相同汉字标记的词”，三浦（1997）认为“中日同形词就是中文和日语里都存在的，拥有相同词形的词”。在此基础上，潘均（1995）提出3个判断同形词的标准。①用相同汉字标记（忽略繁体和简体的差异）②相同的词源③中日两国都正在使用的。结合以上研究，从习得的角度来看，本稿将同形词定义为“用相同汉字书写的，中日两国现在都在使用的同形的词”。

本稿以日本国际交流基金、财团法人日本国际教育协会编写的“日语能力考试出题基准”中的679个四级词汇为对象，先从标记的立场，分为汉字标记词汇、汉字假名标记词汇，假名标记词汇。从汉字标记的164个词汇中选出与本文同形词定义相符合的103个词汇。

剔除其中21个量词以及15个数词，剩下67个词语，其中“まいつき（每月）”和“まいげつ（每月）”、“まいとし（每年）”和“まいねん（每年）”是用相同的汉字标记，“外国人”是“外国”+“人”，将“外国人”归属于“外国”，因此调查对象变成以下的64个同形词。

えいご（英語），えんぴつ（鉛筆），おんがく（音楽），がいこく（外国），がっこう（学校），かびん（花瓶），かんじ（漢字），ぎゅうにく（牛肉），ぎゅうにゅう（牛乳），きょねん（去年），ぎんこう（銀行），けっこん（結婚）・する，こんばん（今晚），ざっし（雑誌），さとう（砂糖），さんぽ（散歩）・する，じかん（時間），だいがく（大学），たいしかん（大使館），ちず（地図），ちゃいろ（茶色），でんわ（電話），どうぶつ（動物），としょかん（図書館），にく（肉），ぼうし（帽子），まいつき/まいげつ（毎月），まいしゅう（毎週），まいとし/まいねん（毎年），まいにち（毎日），まいばん（每晚），ゆうめい（有名），らいねん（来年），りゅうがくせい（留学生），りょこう（旅行）・する，れんしゅう（練習）・する，いちにち（一日），がくせい（学生），かぞく（家族），きょうしつ（教室），きょうだい（兄弟），げんき（元気），こうえん（公園），しょくどう（食堂），しんぶん（新聞），せんせい（先生），ぜんぶ（全部），せんたく（洗濯）・する，そうじ（掃除）・する，てんき（天気），でんしゃ（電車），べんり（便利），もん（門），もんだい（問題），かいだん（階段），やさい（野菜），けっこう（結構），りょうり（料理），さくぶん（作文），いみ（意味），しつもん（質問），じょうぶ（丈夫），だいじょうぶ（大丈夫），べんきょう（勉強）・する。

2) 数据收集方法

为了研究各国学习者同形词使用的实际情况，不可或缺的是以多国母语的日语学习者为目的的语料库。本文为了获得自然的产出，利用日本国立国语研究所公开的《多言語母語の日本語学習者横断コーパス：I-JAS》（后文简作I-JAS），以对话数据为研究对象，进行量的分析。本文研究的是能力考试4级词汇，作为基本词汇，在会话语料中也有大量出现。

本研究采用的 I-JAS¹ 语料库，是日本国立国语研究所开发的，包含日本和国外 17 个国家及地区的日语学习者的语料。此语料库包含了，12 种不同母语的学习者、日本国内的自然环境和教室环境的学习者、以及日文母语者的口语语料和文字语料。2016 年春，对 225 名对象的数据进行了第一次公开。本文使用的数据是 2016 年公开的口语语料的数据。以下对 I-JAS 语料库做一个简单介绍。

此语料库的调查对象按照分为以下四个类别：①海外教室环境学习者②国内教室环境学习者③国内自然环境学习者④日本母语者。对象是可以与母语者进行 30 分钟对话且在一定程度上能进行读写的学习者。①为海外的 17 个国家和地区，12 个母语种类。②是以初级水平来日本，在日本教育机构接受一年以上日语系统学习的学习者，主要是日语教学机构和日本大学。③是在日本生活的过程中自然习得日语的学习者 ④是从 20 岁到 50 岁，没有日语教学经历的日语母语者。

本文采用的 2016 年公开的数据，包括 12 种海外不同母语的海外日语学习者 15 名（180 名）、教室环境学习者 15 名、自然环境学习者 15 名，日语母语者 15 名，共 225 名。其中的 230 名日语学习者通过测试，被认为日语水平相当。数据种类有 6 种，共 12 个任务，本文对象是口语语料的 4 个类型，6 个任务（ストーリーテリング 1、ストーリーテリング 2、対話、ロールプレイ 1、ロールプレイ 2、絵描写）。因为“絵描写”没有完全公开，本文采用的的口语语料只有前 5 个任务，数据概况详见下表 1。²

表 1 本文利用的语料库概要

语言环境		调查地点	调查 ID	ストーリーテリング 1	ストーリーテリング 2	対話	ロールプレイ 1	ロールプレイ 2
海外	中文	中国（15 人）	CCM	○	○	○	○	○
	韩语	韩国（15 人）	KKD	○	○	○	○	○
	泰语	泰国（15 人）	TTH	○	○	○	○	○
	越南语	越南（15 人）	VVN	○	○	○	○	○
	印尼语	印度尼西亚（15 人）	IID	○	○	○	○	○
	英语	澳大利亚（15 人）	EAU	○	○	○	○	○
	德语	奥地利（15 人）	GAT	○	○	○	○	○
	法语	法国（15 人）	FFR	○	○	○	○	○
	西班牙语	西班牙（15 人）	SES	○	○	○	○	○
	俄语	俄罗斯（15 人）	RRS	○	○	○	○	○
	匈牙利语	匈牙利（15 人）	HHG	○	○	○	○	○
土耳其语	土耳其（15 人）	TTR	○	○	○	○	○	
日本国内	日本教室环境	国内教室（15 人）	JJC	○	○	○	○	○
	日本自然习得	日本自然环境（15 人）	JJN	○	○	○	○	○
	日语母语	日语母语（15 人）	JJJ	○	○	○	○	○

¹ 根据迫田久美子(2016)『I-JAS 構築に関する最終報告書』公开的信息。

² 根据迫田久美子(2016)『I-JAS 構築に関する最終報告書』公开的调查信息。

其中语料库中会话的内容，主题和基本流程详见表 2。

表 2 对话内容及主题

内容	主题	要求
ストーリーテリング	タスク 1 ピクニック	提示されたイラストのストーリーを話す
	タスク 2 鍵	
対話	全 15 項目で、前半は学習者本人に関する過去、現在、未来について、後半には、談話レベルの発話が引き出せるよう、意見陳述や反論を設定した。	自然な会話を 30 分程度行う
ロールプレイ	店長にアルバイトの出勤日数の変更を依頼する	設定された場面に応じて、与えられた役を演じて会話する
	店長からの仕事内容の変更を断る	
絵描写	「家の中のこと」と「家の外のこと」	イラストを見て、説明する

3 同形词的使用状况

64 个对象同形词在口语语料库中出现的总个数是 14353 个，也就是说 225 名调查者，平均每人使用了 63.79 个对象同形词，平均使用率较高。接下来的图 1 是使用次数前 20 位的同形词的使用次数以及各自次数占总使用次数的比率(下文称为使用率)。

图 1 前 20 位对象同形词的使用状况



从同形词前 20 位的图来看，前 20 位同形词的总使用次数高达 12406 次，占同形词整体的 86.05%。从以上的调查结果来看，对同形词的使用集中在前 20 位的同形词，在某种程度上可以看出学习者使用同形词的倾向，这种倾向性应该与语料库调查时的话题以及学习者的使用倾向都有关系。

之后的两部分,将根据不同调查地的每一个学习者对于能力考试4级同形词的使用次数和使用频率的数据,通过SPSS的组间聚类分析,研究不同母语学习者的使用特征,以及对不同同形词使用的倾向。

1) 基于同形词的使用次数对调查地分类

① 研究目的

不同母语的学习者在同形词的使用特征上应该拥有自己的特征。对于同形词使用的特征,先行研究中有“汉字圈”与“非汉字圈”这样基于与汉字关系远近的分类,也有基于地理位置远近的“亚洲圈”、“欧美圈”的分类。这样的分类,在长期的汉字研究中已经成为一种默认的方法,直觉上来看确实存在一定合理性,但是通过对于同形词实际使用的数据分析,结果到底会怎样呢?同时,本文的研究对象是4级同形词,作为最基础的词汇应该是学习者早已习得的项目,但是学习者实际的使用情况会与母语者存在差异吗?

以解开以上问题为目的,将通过SPSS的组间聚类分析,对不同调查地进行分类。

② 研究顺序

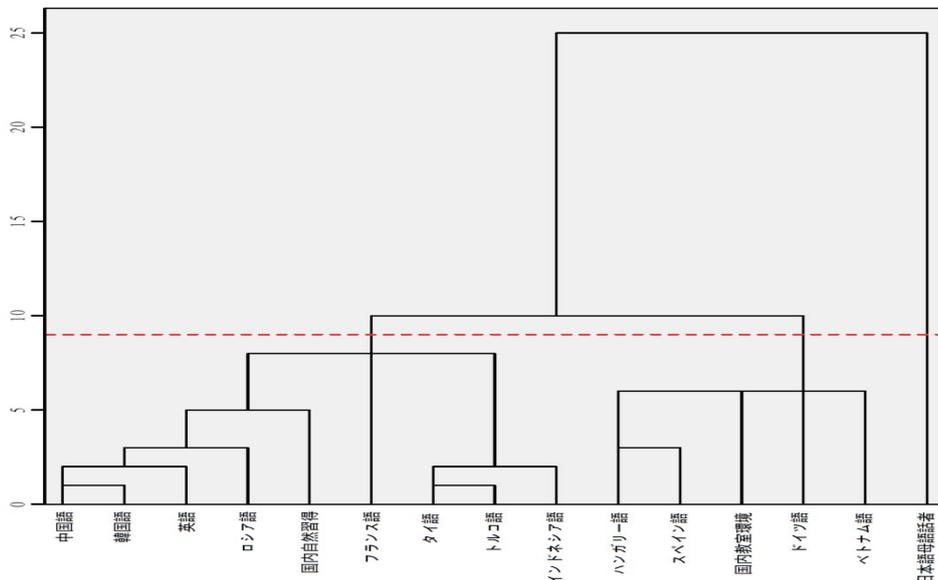
本部分使用的数据,是15个调查地的被调查者4级同形词的使用次数。调查地是表1的15个调查地点,各母语的日语学习者15名(共180名),日本教室学习者15名,日本自然环境习得15名,日语母语者15名,共225名调查者。

15个调查地的调查对象对64个同形词的使用总数为14353次,在此通过SPSS,以各调查地点的词语使用总数为变量,以调查地点为样本做组间聚类分析。

③ 结果及分析

通过数据分析得出以下树形图,结合研究目的,将调查地点分成了三类。

图2 15个调查地点分类树形图



如图2,将调查地点分成了三类,第一类调查地点的母语分别是“中文”、“韩语”、“英语”、“俄语”、“国内自然环境”、“法语”、“泰语”、“印尼语”;第二类调查地点的母语分别是“匈牙利语”、“西班牙语”、“国内教室环境”、“德语”、“越南语”;第三类只有日语母语者。

接下来,根据研究目的进行进一步分析。首先,从各国学习者的母语对使用频率的影响

来看。以上分类，一类和二类全部是日语学习者，藤堂（1971）将中国，日本，韩国，越南同时归为汉字文化圈。图2左端，同属汉字圈的中国和韩国基本融合在一起，右端的越南与中国，韩国相分离，与其他非汉字圈的国家相融合。分析其原因，越南成为了法国殖民地之后，在其支配下流入罗马字标记的国语字，除了一部分年长者和专家，以及日语和中文学习者以外，理解汉字的人很少。接下来，从“亚洲圈”“欧美圈”这样的地理远近的分类来看，第一类的调查地点中，除了国内自然习得，英语、法语，俄语属于欧美圈，其他的五个都属于亚洲圈。在第二类中，除了日本教室环境以外，三个都是欧美圈，只有越南是亚洲圈。从以上分类可以看出一定的与汉字圈的关联性。

此外，对于学习者与母语者使用情况的比较。图2最右端的日语母语者很明显与第一类和第二类的学习者分开了。从这个结果来看，日语母语者与日语学习者在同形词使用上还是存在明显差异的。由此可以得出结论，学习者拥有自己的中介语系统。本文的研究对象是4级的同形词，作为基本词汇，通常认为学习者已经习得了。但是从结果来看，使用尚未融合，从某种意义上来说，对于基础词汇的使用，尚不地道，与母语者仍然存在差异，还存在向母语表达靠近的空间。

2) 基于同形词的使用次数对同形词分类

本节部分基于同形词的使用次数的数据，对64个对象同形词进行分类，寻找学习者在使用同形词时的偏好。

① 研究目的

一般仅仅看学习者的语料很难明确看到词语使用的共通性，本文尝试用组间聚类分析对同形词进行分类。具体来说，有以下几个目标①同形词该如何分类？②同形词词性的差异对分类是否有影响？③同形词词义以及对话主题对分类是否有影响？

② 调查顺序

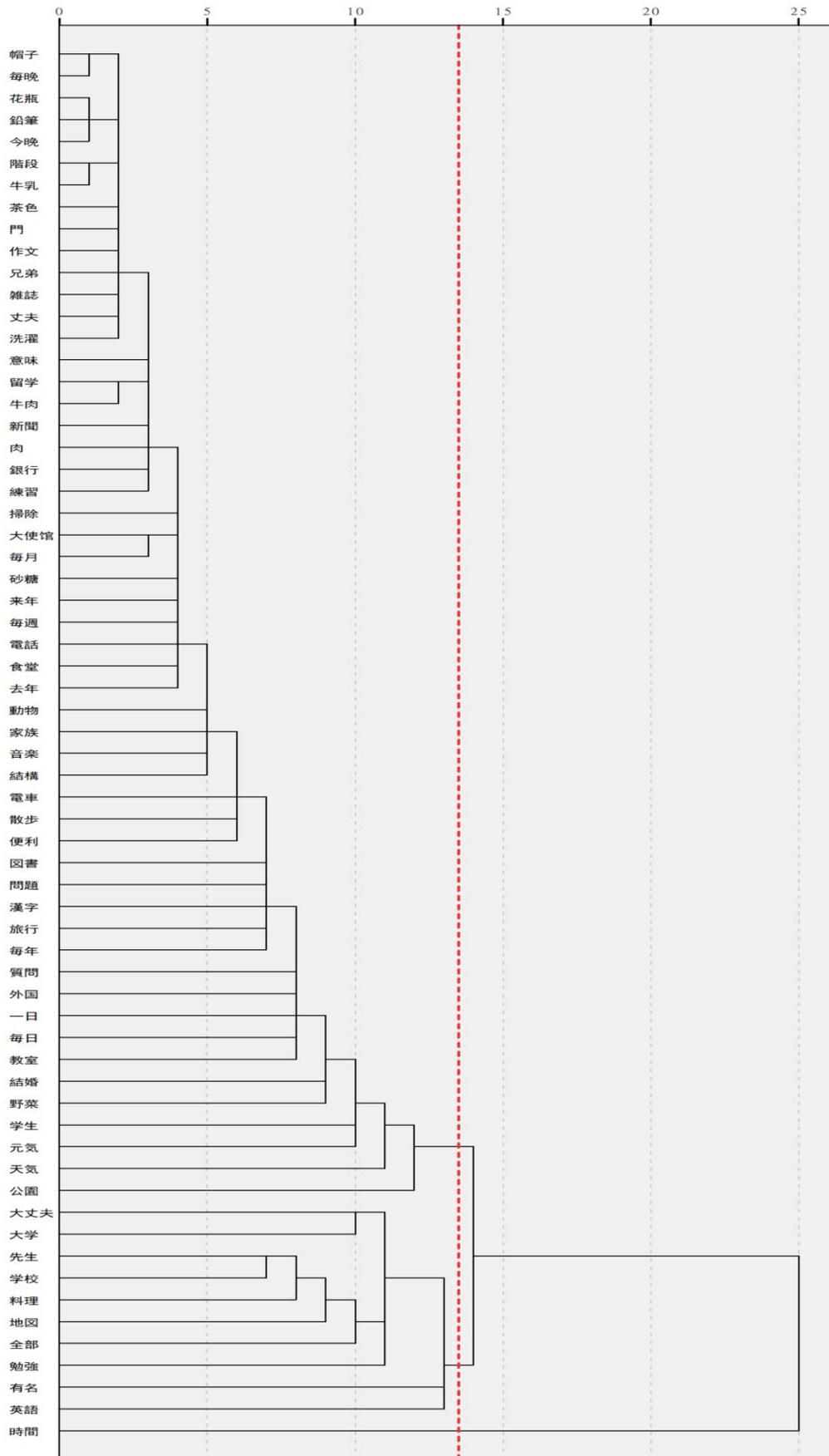
本节使用的数据与1)调查地点分类的数据相同，以15个调查地点的64个对象同形词的数据进行分析。以多国学习者每个人同形词使用的次数为变数，以64个同形词为样本，进行组间聚类分析。

③ 结果与分析

通过组间聚类分析得到图3的树形图，根据研究目的和图的情况，画出虚线，将图中的64个同形词分成三类。

第一类有：帽子、每晚、花瓶、铅笔、今晚、阶段、牛乳、茶色、門、作文、兄弟、雜誌、丈夫、洗濯、意味、留学、牛肉、新聞、肉、銀行、練習、掃除、大使馆、每月、砂糖、来年、每週、電話、食堂、去年、動物、家族、音樂、結構、電車、散步、便利、圖書、問題、漢字、旅行、每年、質問、外国、一日、每日、教室、結婚、野菜、学生、元氣、天氣、公園。第二类有：大丈夫、大学、先生、学校、料理、地図、全部、勉強、有名、英語。第三类有：時間。

图3 64个同形词分类树形图



接下来，分析词性对分类的影响。第三类只有唯一的一个名词；第二类的10个词语中，

两个形容词（大丈夫、有名），两个サ变动词（料理、勉強），一个副词（全部）。在第一类的 53 个词中，也含有各种词性。由此可以看到词性对于分类的影响并不大。

最后从词义与对话的主题来看对使用次数的影响。单看个别同形词很难看出倾向，要结合大类来研究词义及对话主题对学习者使用的影响。结合表 1 使用次数前 20 的同形词表来看，第二类和第三类的同形词全部都是使用次数前 15 的同形词。第三类的“時間”，在对话中有“バイト時間の調整”这样的话题，所以“時間”的使用数量异常多，单独成为了一类。而第二类的词语，与话题和学习者生活的关联度较高，成为学习者不回避且经常使用的词汇。通过以上结果可以看出，第二类词和第三类词是受学习者个人影响小，在某种程度上存在一些相似性的词。

4 结语

本文采用了 I-JAS 语料库中的口语语料，以 4 级的 64 个同形词为调查对象，考察 14 个调查地的学习者以及日语母语者中日同形词的使用情况，并用 SPSS 做了组间聚类分析，分析不同调查地点的学习者以及母语者同形词使用特征，以及学习者使用不同同形词的倾向。

研究结果发现，各国学习者的习得有其自身特点。首先，研究对象虽然是 4 级基本词汇词汇，但是在使用倾向上，学习者和母语者仍然存在差异，学习者有自己的中介语系统。其次，在同形词使用频率对调查地的分类中，“汉字圈”，“非汉字圈”这种与汉字的亲疏和“亚洲圈”，“欧美圈”这种地理位置的远近，在一定程度上与调查地分类存在联系。最后，通过同形词使用频率对同形词的分类，得出第二类和第三类是不被回避、经常使用的同形词，且词性差别对分类结果影响较小。

上述研究只考察了同水平学习者对日语能力考试 4 级同形词的使用情况。今后有机会将进一步扩大考察范围，对各水平学习者进行纵向研究，研究对象也可以扩展到其他各级别的词汇。

参考文献

- [1] 曲 维, 1995, 「中日同形词的比较研究」, 辽宁师范大学学报(社科版)(6), 34-37.
- [2] 潘钧, 2000, 『关于中日同形词语法差异的一次考察』, 北京大学外国语学院日语系北京大学日本文化研究所, 日本语言文化论集 2[C]. 北京: 北京出版社.
- [3] 庄倩, 2014, 「中日同形词的产出难易度及其影响因素」, 日语学习与研究, 52-58.
- [4] 大河内康憲, 1992, 「日本語と中国語の同形語」. 『日本語と中国語の対照研究』, くろしお出版
- [5] 陳毓敏, 2002, 「中国語を母語とする日本語学習者における漢語習得—同形同義語の文法的ずれに焦点を当てて」, 2002年日本語教育大会秋季大会予稿集[R], 143-149.
- [6] 陳毓敏, 2003, 「中国語を母語とする日本語学習者の漢語習得について—同義語・類義語・異義語・脱落語の4タイプからの検討—」, 2003年度日本語教育学会秋季大会予稿集, 日本語教育学会, 174-179
- 大和裕子、玉岡賀津雄, 2009, 『中国人日本語学習者の日本語漢字語の処理における母語の影響』, ことばの科学(22), 117-135.
- [7] 李善姬, 2009, 「韓国人日本語学習者の漢字学習に対するピリーフとストラテジー」, 『日本學 研究』第 27 輯, 韓国日本学会, 韓国, 485-503 .
- [8] 大塚薫、林翠芳, 2010, 「中上級レベルの日本語学習者の作文教育—意見文にみる語彙・漢字使用及び誤用の分析結果を踏まえて」高知大学総合教育センター修学, 留学生支援部門紀要, 47—66
- [9] 張麟声, 2014, 「中国語話者のための日本語教育研究」. 日中言語文化出版社, 17-30

The Investigation and Analysis of Japanese Kanji Words of the Japanese Learners from Various Countries

Cheng Wei

(Hunan University, Changsha in Hunan Province, 410012)

Abstract: By taking Japanese kanji words of *The Japanese-Language Proficiency Test* level4 as the research objects,the thesis is aimed to examine the usage of Japanese kanji words between multinational Japanese learners on the basis of analyzing,organizing and summarizing a great quantity of spoken corpus in *International Corpus of Japanese As a Second Language*.The research turns out that multinational Japanese learners has their own characteristics in the usage of Japanese kanji words.All the conclusion is drawn by the SPSS cluster analysis.First,although the research objects are level4 basic vocabulary,there are still differences between learners and native speakers in terms of usage of quantity,and learners have their own inter language phonology.Then,in terms of the frequency of usage of Japanese kanji words and the classification of survey location,the intimacy of Chinese characters such as “Chinese characters circle”and “non Chinese characters circle”and the geographical distance such as “Asian circle”and “Europe and America circle”are related to the classification of survey location to some extend.Finally,through the frequency of usage of Japanese kanji words and the classification of these words,the conclusion is drawn that the third and second categories are frequently used and the part of speech has little influence on the classification.

Keywords: Japanese kanji words, Multinational Japanese learners, oral corpus, cluster analysis ,usage