



语言学论丛

国家社会科学基金项目成果（立项批准号12BYY138）

A STUDY OF HOMOGRAPHS IN
MODERN CHINESE AND
JAPANESE

中日现代语言 同形词汇研究

施建军 —— 著



北京大学出版社
PEKING UNIVERSITY PRESS

A STUDY OF HOMOGRAPHS IN
MODERN CHINESE AND
JAPANESE

中日现代语言 同形词汇研究

日本是中国以外大量使用汉字的国家，现代日语中使用着丰富的汉字词汇。汉语和日语哪些汉字词汇是通用的，即有多少同形词？使用情况如何？这不仅是语言学界长期关注的学术问题，也是中日两国人民普遍关心的社会问题。本书引入大数据的理念和自然语言处理技术，以中日两国的主流报纸近10年的数据为对象，开展了中日同形词的大规模调查，所处理的数据规模超过了3.5亿字。共收集中日同形词16000多条，调查整理了中日两国现代语言生活中正在使用的同形词的条目、数量以及使用频度。同时，利用中日对译语料库和统计学方法计算出常用同形词中日之间意义用法的差距。这些成果为学界深入开展中日语言比较研究提供了可靠的学术资料。



“北大外文学堂”
微信公众号



“北京大学出版社”
微信公众号

ISBN 978-7-301-30740-3



9 787301 307403 >

定价：106.00元



语言学论丛

A STUDY OF HOMOGRAPHS IN
MODERN CHINESE AND
JAPANESE

中日现代语言 同形词汇研究

施建军 —— 著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目 (CIP) 数据

中日现代语言同形词汇研究 / 施建军著. —北京: 北京大学出版社,
2019.10

(语言学论丛)

ISBN 978-7-301-30740-3

I. ①中… II. ①施… III. ①汉语—同形词—对比研究—日语 IV. ①H13 ②H363

中国版本图书馆 CIP 数据核字 (2019) 第 194102 号

- 书 名 中日现代语言同形词汇研究
ZHONGRI XIANDAI YUYAN TONGXING CIHUI YANJIU
- 著作责任者 施建军 著
- 责任编辑 兰 婷
- 标准书号 ISBN 978-7-301-30740-3
- 出版发行 北京大学出版社
- 地 址 北京市海淀区成府路 205 号 100871
- 网 址 <http://www.pup.com.cn> 新浪微博: @北京大学出版社
- 电子信箱 lanting371@163.com
- 电 话 邮购部 010-62752015 发行部 010-62750672 编辑部 010-62759634
- 印 刷 者 北京虎彩文化传播有限公司
- 经 销 者 新华书店
- 720 毫米 × 1020 毫米 16 开本 32.25 印张 510 千字
- 2019 年 10 月第 1 版 2019 年 10 月第 1 次印刷
- 定 价 106.00 元

未经许可, 不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有, 侵权必究

举报电话: 010-62752024 电子信箱: fd@pup.pku.edu.cn

图书如有印装质量问题, 请与出版部联系, 电话: 010-62756370



施建军 上海外国语大学语言研究院教授、博士生导师、中国外语战略研究中心研究员。历任解放军外国语学院教授、北京外国语大学北京日本学研究中心教授、博士生导师。曾获解放军科技进步奖、解放军教学成果奖、宋庆龄基金会“孙平化日本学学术奖励基金”专著奖等多项学术奖项。入选2010年度教育部“新世纪优秀人才支持计划”。教育部重大攻关项目“东亚国家语言中汉字词汇使用现状研究”首席专家。汉日对比语言学会第四届理事会会长、中国日语教学研究会第八届理事会秘书长、汉日对比语言学会第三届理事会副会长兼秘书长。

主要从事日语语言学、日文信息计算机智能处理、计量文体学、汉字文化圈国家语言文字比较等领域的研究。在国内外学术期刊发表论文数十篇，出版专著2部，主编和参编学术著作近10部，主持和承担了包括国家社会科学基金重大项目、教育部哲学社会科学重大课题攻关项目、国家社会科学基金一般项目、教育部人文社会科学研究项目在内的10多项纵向和横向科研课题。

前 言

35年前开始学习日语，看到日语里有很多汉字和汉字词，认为日语比较容易。但是，大学毕业时，这个观念彻底改变：日语很难。现代日语遵循着很强的语法规律，格助词、助动词等语法功能词把日语的句子成分标记得非常清楚，但是，在学习过程中却发现日语越学越难。其中一个难点正是一直被中国学生看做学习日语捷径的汉字。

首先，汉字的日语读音，虽说日语的音读汉字词和古代汉语的发音有着千丝万缕的联系，但是，现代汉语的发音已经和古代汉语相去甚远，试图寻找现代汉语和日语之间的发音联系并非易事。加上同一个日语汉字有多种发音，而且有些文章中的汉字甚至是使用者自己决定其发音的，碰到这种情况即便日本人都不敢随便读出。其次，日语中有很多和现代汉语一样或者相似的词汇，我们称之为中日同形词。初学日语时，我们毫无顾忌，将这些词在汉语中的意义用法也沿用到日语同形词上面。后来在和日本人交流时发现，这些同形词的意思汉语和日语并不一样。除了少数意义完全相同或完全不同的之外，绝大多数同形词中日之间的意义用法都存在微妙差别，似是而非，同形词的误用有时甚至影响了中国人和日本人之间的交际。

众所周知，汉语对日语语言文字的形成产生了巨大影响。日本人不仅借用汉字创造了日语的文字，日语中还使用着大量汉字词汇。据日本国立国语研究所的调查，汉字词占日语总词汇量的47.5%。在近代，日语又反过来对汉语产生了影响。明治维新以后，日本人为了翻译西方先进文化的概念创造了许多汉字词，这些词有相当一部分也传到中国，融入了汉语。汉语和日语

的这种相互影响形成了大量的中日同形词。

中日两国学界很早就关注汉语和日语的汉字词汇问题。中日邦交正常化后，随着中日交流的日益频繁，日语教学和汉语教学事业在中国和日本蓬勃发展，活跃在教学第一线的日语教师和汉语教师，都从教学的角度积极探讨日语中的汉字词问题。因为这个问题涉及留学生的培养，所以中日同形词问题，也引起了中日两国政府的关注。日本政府文化厅曾于20世纪70年代初，给早稻田大学下达了研究汉语和日语汉字词汇的课题。经过长期探索，学界在这一领域取得了丰硕的研究成果。

但是，不可否认中日汉字词汇的比较是一个广泛的研究领域，存在无数值得研究的课题。虽然学界做了大量探索，但是，很多问题并没有得到真正解决。有的甚至是这一领域最基础的问题。直到现在，我们还一直在纠结，我们应该把这些在汉语和日语中大量使用的、具有相同形态的词叫做中日同形词还是叫做汉日同形词呢？后来因为这是中日两国语言的特有现象，学界一直使用“中日同形词”这个说法，所以这本书里就将“中日同形词”作为一个专业术语来使用。另外，还有一些非常基础、也是大家普遍关心的问题，如中日两国现代语言生活中使用着多少中日同形词等。对于这些问题，学界并没有给出一个看法一致的结论。

带着这些疑问，我们在国家社会科学基金的资助下，对中日同形词问题开展了大规模的调查研究。由于我们的目标是要摸清中日同形词在现代语言生活中的使用情况，必须处理大量的现实语言材料，因此在研究过程中，使用了近10年的报纸数据，规模达到几亿字，同时使用了前沿的自然语言处理技术。正是因为数据和技术的保障，使得我们的研究得以顺利开展，并且保证了成果的质量和科学性。在研究期间，我们将所获得的阶段性研究成果同国内以及日本的同行进行交流，也得到了同领域专家的高度评价。有的学者认为这是汉字词汇研究的珍贵资料，希望能够尽早出版、供学界共享。

该项目执行过程中，得到了北京外国语大学谯燕教授以及南京邮电大学

洪洁老师、南京工业大学许雪华老师的支持。特别是在同形词的形态、比较方法、近似度测定等研究中三位老师分别做出了贡献，借此机会向这三位老师表示衷心感谢。另外，北京外国语大学的朱京伟教授、辽宁师范大学的刘凡夫教授、陕西师范大学的彭广陆教授长期以来对本研究给予了高度关注和热情鼓励，在此对三位教授表示感谢！同时，该书出版过程中，得到了北京大学出版社的大力支持，编辑兰婷老师付出了很多努力和汗水，提出了许多宝贵建议，也一并表示感谢！

由于时间仓促、水平有限，书中一定存在不少疏漏和不足，恳望各位读者批评指正！

上海外国语大学语言研究院

中国外语战略研究中心

施建军

2019年6月

目 录

第一章 绪论·····	1
第二章 中日两国现代语言同形词比较研究的现状及存在的课题·····	9
第一节 中日同形词的界定问题及其研究概况·····	10
第二节 关于中日同形词数量和条目问题的研究·····	13
第三节 关于中日同形词分类问题的研究·····	16
第四节 关于中日同形词意义用法差别描述方法的研究·····	20
第五节 中日两国现代语言同形词研究的课题和方法·····	23
第三章 中日两国现代语言同形词诸态·····	26
第一节 关于同形词概念和范围的再思考·····	28
第二节 标准中日同形词·····	31
2.1 简化字与中日同形词·····	32
2.2 日语代用字与中日同形词·····	32
2.3 源自日文训读词的中日同形词·····	34
2.4 汉语新的日语借词·····	35
第三节 广义中日同形词·····	36
3.1 汉语成语被缩略的广义同形词·····	36
3.2 错位四字广义同形词·····	37
3.3 汉语的词组和日语的词·····	38

3.4 中日同素逆序词	39
3.5 部分词素同形的词汇	39
3.6 中日无关联的同形词	40
第四节 中日同形词诸态	41
第四章 中日两国现代语言同形词的大规模调查	43
第一节 基于语料库的现代汉语词汇和现代日语	
汉字词汇的获取	44
第二节 中日同形词计算机自动获取的方法和过程	49
2.1 中日同形词的形式特征	50
2.2 中日同形词自动抽取用异体字表的建立	51
2.3 中日同形词的计算机自动抽取	52
第三节 中日同形词的手工甄别	57
3.1 机器获取同形词的剔重处理	58
3.2 其他不规则同形词的手工确认	64
第五章 基于意义用法差别的中日同形词分类问题	67
第一节 中日同形词意义用法分类的复杂性	68
第二节 关于同形近义词差异大小的问题	72
第三节 中日同形近义词精细描述的必要性	77
第六章 中日同形词意义用法距离及其计量	81
第一节 中日同形词意义用法差异研究的计量视角	83
第二节 对译比和中日同形词意义用法距离	84
第三节 对译系数和中日同形词意义用法距离的计量	89

第七章	中日同形词意义用法对比方法·····	96
第一节	中日同形词意义用法差异描写方法相关研究·····	97
第二节	中日同形词意义用法差异情况·····	99
第三节	中日同形词对比的方法论·····	101
	3.1 意义层面 ·····	102
	3.2 词性层面 ·····	103
	3.3 句法功能层面 ·····	104
	3.4 搭配关系层面 ·····	109
	3.5 中日同形词意义用法差异分析的普遍视角 ·····	116
第八章	中日两国现代语言同形词使用情况总表·····	119
第九章	常用中日同形词意义用法距离·····	471
参考文献	·····	506

第一章 绪论

众所周知，汉语对日语语言文字的形成产生了巨大影响。日本人不仅借用汉字创造了日语的文字，日语中还使用着大量汉字词汇。据日本国立国语研究所的调查，汉字词占日语总词汇量的47.5%。在近代，日语又反过来对汉语产生了影响。明治维新以后，日本人为了翻译西方先进文化的概念创造了许多汉字词，这些词有相当一部分也传到中国，进入了汉语。汉语和日语的这种相互影响形成了大量的中日同形词汇。

有关中日同形词的研究在中日两国学界有很长的历史，取得了丰富成果。但是不可否认，该领域研究仍然存在着许多尚未彻底解决的问题，有的甚至是有关中日同形词的最基础的问题。本课题计划的目标是以中国和日本正在发行的主流报纸为调查对象，系统考察现代汉语和现代日语中正在使用的同形词汇的体系和同形词的实际使用情况。并在此基础上，遴选常用汉日语同形近义词，用语料库语言学的方法调查其意义用法差别。同时，在同形词的计算机自动抽取以及同形同义词、同形异义词、同形近义词的计算机辅助归类、同形词意义用法的甄别等方面进行理论探索。重点解决以下几大问题：1. 中日同形词的界定；2. 基于语料库的中日同形词的抽取和同形词使用情况调查；3. 同形同义词、同形近义词、同形异义词的甄别方法探索；4. 常用同形近义词的遴选；5. 常用同形近义词意义用法的描述。经过3年多的努力和攻关，我们圆满解决了上述问题，达成了预期的研究目标，完成了包括《中日两国现代语言同形词使用情况总表》在内的近60万字的最终研究

成果。本研究成果重点研究了以下几方面的内容：

第一，在中日同形词的界定问题上，我们从中国和日本正在发行的两种主流报纸上收集了大量的中日同形词实例，根据这些语言事实对中日同形词进行了清晰的界定。我们认为判断中日同形词的第一要素应该是词形相同。但是由于中日同形词的情况非常复杂，简单地用词形相同并不能够概括中日同形词这一复杂的语言现象。我们综合考虑了词形、词义、词源等因素，将中日同形词分成了词形相同的标准同形词和词形相似且中日之间有渊源关系的广义同形词两大类。在中日同形词的界定问题上提出了我们的观点，发表了论文《再论中日两国语言中的同形词问题》^①。

第二，在对中日同形词进行了界定后，我们根据中日同形词的性质和形态特征，以近10年内发行的《光明日报》（约1.7亿字）和日本《每日新闻》（约1.65亿字）以及日本“新潮文库”等为语料，利用计算机和手工相结合的方法，就中日两国现代语言生活中正在使用的同形词汇有多少、有哪些条目、如何将中日两国语言生活中正在使用的这些同形词汇调查清楚等问题，对中日同形词汇进行了大规模调查，共获取中日同形词16000多条。基本搞清了中日两国现代语言生活中同形词汇的使用情况。

第三，根据同形词中日之间意义用法差别对同形词进行归类是长期困扰学界的一个难题，特别是大规模地对中日同形词进行同形同义、同形近义、同形异义的甄别归类是一项庞大而复杂的工程。针对这个问题，我们认为以中日对译语料库为知识库，可以用计算机实现中日同形词意义用法差距的计算，并以此为依据对中日同形词的意义用法差别进行甄别。利用这个方法我们成功解决了1000多条核心同形词的分类问题。

第四，我们根据中日同形词的使用情况，结合《HSK词汇表》和《日语能力考试分级词汇表》，分别对面向中国人学习日语的核心同形词和面向日

^① 施建军、许雪华：《再论中日两国语言中的同形词问题》，《解放军外国语学院学报》2014年第6期，第132-139页。

本人学习汉语的核心同形词进行了遴选。共选取中国人学习日语的核心同形词1964条，日本人学习汉语的核心同形词3138条。

第五，中日同形词意义用法差别的描述方法对研究中日同形词用法区别具有重要的学术价值，在日语教学和对日汉语教学中具有重要的应用价值。我们对此开展了理论研究，发表了论文《汉日同形词意义用法的对比方法研究》^①，该论文被2013年10月人民大学复印报刊资料《语言文字学》全文转载。利用论文提出的方法，我们对1500多条中日核心同形词进行了意义用法的辨析和描述，计划出版《常用中日同形词意义用法辨析词典》。

在本项目实施过程中，作为阶段性成果我们出版了著作1部；共撰写论文7篇，已经公开发表6篇，其中CSSCI期刊（含扩展版）发表论文3篇，人民大学复印报刊资料《语言文字学》全文转载1篇。圆满地完成了本项目预定的各项研究计划。

本项目在中日同形词的界定、同形词的调查方法、同形词的分类方法、同形词中日之间意义用法对比方法等方面均有创新。

第一，在同形词界定方面的创新。日本文化厅1978年出版的『中国語と対応する漢語』以“漢字音読語”作为中日同形词的研究对象，而将“取消、取締、立場、手続”等中日之间具有相似形态且有渊源关系的词汇排除在外，缩小了同形词的研究范围。上述成果的这个观点一经提出就在学界引起了广泛的争论。日本学者荒川清秀提出中日同形词的判断应该注重中日同形词的“视觉同一性”。但是，由于汉字在古代日语中使用非常广泛，以至于现代日语的绝大多数词汇都可以全部或部分用汉字书写，因此，“视觉同一性”的观点提出后，许多学者又认为“新しい、走る”等日语固有词汇也应该纳入中日同形词研究，这又过分夸大了中日同形词的研究范围。我们认为，中日同形词的界定既要注重形态特征，也要注重汉日语之间词汇交流的

^① 施建军、洪洁：《汉日同形词意义用法的对比方法研究》，《外语教学与研究》2013年第4期，第531-542页。

历史，中日同形词应该是中日两国语言中具有相同形态的词汇。但是考虑到中日两国语言中汉字字形的差异以及同形词形成的历史复杂性，判断同形词时应该遵循以下三条原则：①历史上曾经使用相同汉字构成的同形词汇；②中日同形词中包括由日语进入到汉语的训读词汇；③因语言接触而进入对方语言中、虽词形发生变化但中日之间仍存在严格对应关系的词汇。在此基础上，我们将中日同形词划分成两大类，即标准同形词和广义同形词。标准同形词是指中日之间历史上词形完全相同的词汇，广义同形词是指那些中日之间因语言接触而产生的词形相似、有严格对应关系的词汇。

第二，在调查对象和内容上我们有很大的突破，调查清楚了中日同形词系统的全貌。传统的同形词研究主要是以词典为对象，进行小规模尝试性探索，无法描述整个同形词系统，更不能了解同形词的使用情况。我们利用大规模语料库和计算机技术对中日同形词体系进行了系统的调查，调查清楚了中日两国现代语言生活中正在使用的同形词词条及其使用情况，并给出了中日同形词词汇表。我们使用了近十年的《光明日报》和日本《每日新闻》等资源作为语料，调查语料规模超过3.5亿字。根据大数据的理念，我们利用自然语言处理技术，对这些语料进行了全数调查，获取中日两国现代语言生活中正在使用的同形词汇16000多条，其中标准同形词近15000条，广义同形词1000多条。日本国立国语研究所著名的语言计量研究专家山崎诚教授认为这是该领域研究从未有过的重大成果，是研究东亚国家文化过去和现在的宝贵数据。

第三，我们开发了独创的中日同形词抽取技术。过去有关同形词的调查研究一般使用手工对照词典的方法获取中日同形词。这种方法不可能进行大规模作业，也不可能对中日同形词在现代语言生活中的使用情况进行研究。我们在清晰地界定了中日同形词后，根据同形词的性质特征，找出了中国汉字和日本汉字（包括日本代用字）的对应规律，研究了中国汉字和日本汉字的归一化算法，在对中日汉字进行归一化处理的基础上进行同形词抽取。这

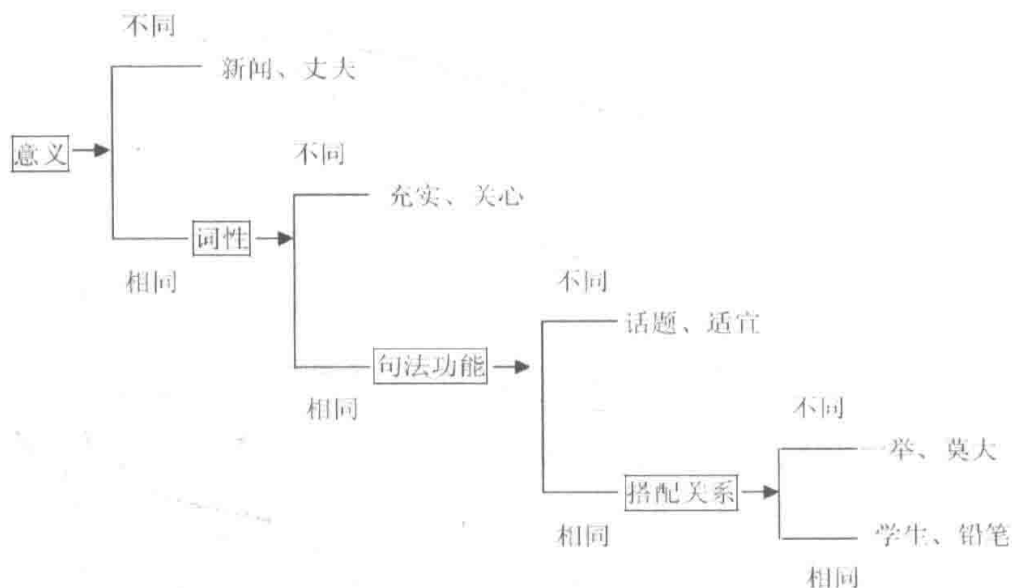
种方法对语料库中标准中日同形词的抽取准确率几乎达到了100%。

第四，首次提出了衡量同形词中日之间意义用法差距的计量方法。学界一直认为同形词中日之间的意义用法差别可用分类的办法来衡量，即可以将中日同形词分为同形同义、同形近义和同形异义三大类别。但是，实际上同形词中同形同义和同形异义的词汇很少，特别是中日之间意义用法完全相同的词汇更少，绝大多数都是同形近义的。而同样是同形近义词，其中日之间的意义用法千差万别。因此，用同形同义、同形异义、同形近义三个类别来区分中日同形词的意义用法差别显然是不科学的。我们认为，中日同形词的互译情况能够客观地反映同形词中日之间的意义用法差别。意义用法基本相同的同形词其互译的可能性就比较大，反之互译的可能性则很小。基于此，我们利用中日同形词的互译比构建了衡量同形词中日之间意义用法距离的系数F：

$$F = \frac{2 \times \text{汉译日比例} \times \text{日译汉比例}}{\text{汉译日比例} + \text{日译汉比例}}$$

F值越大则中日之间的差距越小，F值越小则中日之间的差距越大，同形同义词F值接近1，同形异义词F值接近或等于0。我们利用中日平行语料库计算出了一部分同形词的中日之间意义用法距离的F系数，取得了非常好的效果。

第五，提出了中日同形词意义用法差别的结构化描述方法。中日同形词的意义用法差别表现在什么地方，如何将同形词中日之间的这些差别描述清楚，是我国日语教学和对日汉语教学的一个重要课题。过去针对同形词中日之间差别的个案研究非常多，但是，这些研究所提出的方法往往针对特定的同形词，缺乏普遍性和系统性。我们认为中日同形词意义用法的差别总是表现在以下四个方面，即词汇义、词性、语法功能、搭配。中日同形词表现在这几个方面的差别并不是杂乱无章的，而是有主次之分的，它们形成了如下有序的层级关系：



根据同形词中日间意义用法差别所表现的这四个方面的层级关系，我们按照“意义→词性→句法功能→搭配关系”顺序就可以系统地描述清楚每一对同形词中日差别之所在。

中日同形词是中日两国现代语言生活中大量存在的语言现象，但是，正在使用的中日同形词到底有多少？有哪些条目？等基础性问题迄今为止并没有得到圆满解决。我们通过对大量现实语言材料进行全数调查，不但搞清了中日两国现代语言中正在使用的同形词的数量，而且列出了具体的中日同形词表，使得中日同形词的全貌得以呈现，这在中日词汇比较研究的历史上属首次，为今后进一步深入开展中日词汇交流研究打下了坚实的基础。日本国立国语研究所著名计量语言学专家山崎诚教授认为，获得16082对中日同形词，只此一项即可认为取得重大成果。迄今为止学界从未有研究如此清楚地调查中日同形词系统，该项目所得数据对同时开展东亚文化圈的去和现在的研究极为重要，期待项目竣工之时能够广泛公开数据，使其发展成为吸引全球学者参与的研究领域。

本课题研究成果的学术价值和应用价值具体体现在以下五个方面：

第一，调查现代语言中所使用的中日同形词汇，对面向日本人的汉语教