

中国语言生活绿皮书
国家语言文字工作委员会发布



A002

三本汉字的汉语拼音规范（草案）

文语转换与语音识别系统语言文字评测规范（草案）

机器翻译系统语言文字评测规范（草案）

语料库系统语言文字评测规范（草案）

现代汉语语料库元数据规范（草案）

基于概念层次的语句概念结构语料库标注规范（草案）

教育部语言文字信息管理司 组编

中国语言生活绿皮书



语文出版社

日本汉字的汉语读音规范（草案）

文语转换与语音识别系统语言文字评测规范（草案）

机器翻译系统语言文字评测规范（草案）

语料库系统语言文字评测规范（草案）

现代汉语语料库元数据规范（草案）

基于概念层次的语句概念结构语料库标注规范（草案）

教育部语言文字信息管理司 组编

图书在版编目(CIP)数据

日本汉字的汉语读音规范(草案)·文语转换与语言识别系统语言文字评测规范(草案)·机器翻译系统语言文字评测规范(草案)·语料库系统语言文字评测规范(草案)·现代汉语语料库元数据规范(草案)·基于概念层次的语句概念结构语料库标注规范(草案)/教育部语用司编. —北京:语文出版社, 2009. 7
ISBN 978-7-80241-129-6

I. 日… II. 教… III. 汉语-规范-汇编 IV. H102-65

中国版本图书馆CIP数据核字(2009)第102075号

中国语言生活绿皮书

国家语言文字工作委员会发布

日本汉字的汉语读音规范(草案)

文语转换与语音识别系统语言文字评测规范(草案)

机器翻译系统语言文字评测规范(草案)

语料库系统语言文字评测规范(草案)

现代汉语语料库元数据规范(草案)

基于概念层次的语句概念结构语料库标注规范(草案)

教育部语言文字信息管理司 组编

*

语文出版社出版

100010 北京朝阳门南小街51号

E-mail: ywp@ywcbs.com

新华书店经销 北京市联华印刷厂印刷

*

787毫米×1092毫米 16开本 8.75印张 151千字

2009年10月第1版 2009年10月第1次印刷

印数: 1—1,000 定价: 22.00元

本书如有缺页、倒页、脱页, 请寄本社发行部调换。

《中国语言生活绿皮书》A 系列说明

《中国语言生活绿皮书》由国家语言文字工作委员会发布，旨在落实《中华人民共和国国家通用语言文字法》精神，提倡“语言服务”理念，方便社会共享政府拥有的信息，积极引导社会语言生活和谐发展，为构建和谐社会做贡献。

《中国语言生活绿皮书》分 A、B 两系列，各自连续编号发布出版。B 系列是中国语言生活的状况与分析，主要发布语言生活的各种调查报告和实态数据，为国家决策提供参考，为社会提供服务。

A 系列是引导语言生活的“软性”规范。语言文字规范及其制定有各种不同情况：语言文字规范制定难度大、涉及面广，常需较长的试行试用过程；许多语言文字现象具有弹性，不易在短期内形成共识或不宜有“硬性”规定；有些热点领域急需引导但一时又难以制定出妥贴的规范；语言文字信息处理等领域需要一些超前性的引导规范。以《中国语言生活绿皮书》的形式发布一些“软性”规范，就是为了适应语言文字规范制定的复杂情况，满足社会语言生活的多种需求，保证规范制定的科学性和时效性。

“软性”规范的作用是向社会提供参考，并鼓励需要者采用。规范加上“草案”字样，以与国家正式的语言文字规范有所区别。有些规范通过试用完善，可以升格为国家语言文字的正式规范。

《中国语言生活绿皮书》的发布内容是开放的，不局限于国家语言文字工作委员会的科研项目，也吸纳其他基金项目和企业、社会研究者的优秀成果。《中国语言生活绿皮书》的出版者也是开放的，欢迎各家出版社加入出版行列。《中国语言生活绿皮书》的发布内容和出版单位的选定，都遵循一定的遴选程序。

许嘉璐先生为《中国语言生活绿皮书》题字；教育部副部长袁贵仁同志，

教育部副部长、国家语委主任赵沁平同志，国家语委咨询委员会副主任朱新均同志，都很关心《中国语言生活绿皮书》的编制发布工作。相关课题组为《中国语言生活绿皮书》的研制做出了贡献，一些出版单位和社会人士对《中国语言生活绿皮书》编辑出版给以关心和支持。在此特致谢忱！

教育部语言文字信息管理司



顾 问 许嘉璐 赵沁平

组 编 教育部语言文字信息管理司

编委会

主 任 李宇明

副主任 王铁琨

编 委 陈双新 陈淑梅 戴红亮 王翠叶 王丹卉 王东海

王 敏 张万彬 周洪波

日本汉字的汉语读音规范（草案）

研制人：费锦昌 松冈荣志

文语转换与语音识别系统语言文字评测规范（草案）

机器翻译系统语言文字评测规范（草案）

语料库系统语言文字评测规范（草案）

研制人：冯志伟 肖 航 富 丽 章云帆 王仁华

现代汉语语料库元数据规范（草案）

研制人：何婷婷 沈辉宇 徐晓琦

基于概念层次的语句概念结构语料库标注规范（草案）

研制人：张 全 黄曾阳 张运良 缪建明 韦向峰 池毓焕 郝惠宁 雒自清

谭 露 孙雄勇 吴 晨 袁 毅 谢法奎 贾 宁 熊 亮



目 录

日本汉字的汉语读音规范（草案）	（ 1 ）
《日本汉字的汉语读音规范（草案）》研制报告	（ 6 ）
文语转换与语音识别系统语言文字评测规范（草案）	（ 17 ）
机器翻译系统语言文字评测规范（草案）	（ 33 ）
语料库系统语言文字评测规范（草案）	（ 42 ）
《文语转换与语音识别系统语言文字评测规范（草案）》《机 器翻译系统语言文字评测规范（草案）》《语料库系统语言 文字评测规范（草案）》研制报告	（ 54 ）
现代汉语语料库元数据规范（草案）	（ 73 ）
《现代汉语语料库元数据规范（草案）》研制报告	（ 89 ）
基于概念层次的语句概念结构语料库标注规范（草案）	（ 123 ）

日本汉字的汉语读音规范（草案）

1 范围

本规范（草案）规定了汉语文献中日本汉字的普通话读音的拟音规则，并为 32 个使用频率较高的日本汉字拟定了普通话读音。

本规范（草案）可作为汉语报刊、电视、广播、互联网和其他文献称读日本汉字的读音依据，并可为今后拟定汉语文献中其他日本汉字的读音提供参考。

2 规范性引用文件

第一批异体字整理表（1955 年 12 月 22 日）

汉语拼音方案（1958 年 2 月 11 日）

简化字总表（1986 年新版）

现代汉语通用字表（1988 年 3 月 25 日）

3 术语和定义

3.1 日本汉字

日文中使用的汉字。包括：日语文工具书通常认定的仿照中国汉字造字原理和结构体式自造的日本汉字，如“辻”“畑”等；中日同形字，如“杣”“霰”等；日语文工具书认定为日本自造汉字而实为中国古字或俗字的字，如“丼”“碯”等；日本使用的不同于中国简化汉字字形的简体字，如“仏（佛）”“缶（罐）”等。

3.2 中日同形字

日本自造汉字中跟中国已有的某一汉字字形相同,但字义没有关联或相去甚远的汉字,如“杣”;或字形虽为汉语字词书所收录,但无义项,无法进行字义比较的汉字,如“雫”。

3.3 略字

日文所用汉字中的简体字。包括日本自造的简体字和取自中国古字或俗字的简体字,如“对(對)”“缶(罐)”“庀(廣)”“竜(龍)”等。

3.4 古字

跟今字记录同一语词,但为古时通行的字形。如“袴”是“裤”的古字,“袴”“袴”是“裤(褲)”的古字。

3.5 俗字

跟正字记录同一语词,但没有取得规范地位的字形。如“梅”是“梅”的俗字,“榨”是“榨”的俗字。

4 拟音原则

4.1 沿用汉语字词书中已有读音

汉语字词书中已有读音,包括不特别生僻的读音,尽可能沿用。严格控制拟订新音项的数量。新音项的拟订要符合汉语普通话的语音系统。

4.2 通过“声旁”认读形声字

尽量采用通过“声旁”认读形声字的方法。

4.3 分类拟音

本规范(草案)将日本汉字分为日本自造汉字、中日同形字、中国古字或俗

字、日本略字四大类，分别确定拟音规则并拟音。

5 拟音规则

5.1 日本自造汉字

5.1.1 凡可视为形声或会意兼形声的，取“声旁”读音：如“畑”取 tián（田）音，“辻”取 shí（十）音，“込”取 rù（入）音；“嶋”取 tián（田）音；

5.1.2 凡不属或不能按形声或会意兼形声拟音的，取形似字或形似声旁的读音，如“磨”取形似字“磨”的字音 mó；“峠”，右边“声旁”跟“卡”形似，取 kǎ 音。

5.2 中日同形字

5.2.1 凡中国汉字有读音的，直接取同形的中国汉字读音：如日本汉字“匂”，与中国汉字“匀”的早期字形相同，就取 yún 音；日本汉字“榎”，与中国汉字中释为“木顶”义的“楨”的旧字形相同，就取 zhēn 音。

5.2.2 凡汉语字词书无注音或该注音作为日本汉字的汉语普通话读音容易造成汉语使用者误读的，可另定读音。拟定字音时，凡可视为形声或会意兼形声的，一般取“声旁”读音。如“杣”，《汉语大字典》注“弥羶切”，读音当同“棉”（mián）。若把 mián 作为日本汉字“杣”的汉语普通话读音，容易造成汉语使用者误读，就按形声字处理，取 shān（山）音。

5.3 中国古字或俗字

5.3.1 凡依据现有字词书判定为中国古字或俗字的，按中国汉字读音。如“俣”是中国古字，读 yǔ；“榨”是中国汉字“榨”的异体字，读 zhà；

5.3.2 凡汉语字词书无注音或该注音作为日本汉字的汉语普通话读音容易造成汉语使用者误读的，可另定读音。如“刳”，《汉语大字典》注 ní 音。若把 ní 作为日本汉字“刳”的汉语普通话读音，容易造成汉语使用者误读，就按形声字处理，取 rèn（刃）音。

5.4 日本略字（日本简体字）

5.4.1 凡判定日本简体字形的繁体与中国内地简化字形的繁体为同一汉字字

形的，按相应的中国简化字读音，如“広”（日）“广”（中）的繁体同为“廣”，就读“广（guǎng）”音；“桜”（日）“樱”（中）的繁体同为“櫻”，就读“樱（yīng）”音；“対”（日）“对”（中）的繁体同为“對”，就读“对（duì）”音；“実”（日）“实”（中）的繁体同为“實”，就读“实（shí）”音；

5.4.2 凡判定为同一汉字字形日本简化而中国内地没有简化的，按相应的中国传承字读音，如“拂”，日本简化为“払”，中国不简，日本汉字“払”就读“拂（fú）”音；“酔”，日本简化为“醉”，中国不简，日本汉字“酔”就读“醉（zuì）”音。

6 《日本汉字的汉语普通话规范读音表》说明

本规范（草案）对汉语文献中使用频率较高的 32 个日本汉字，依据上述拟音原则和规则，拟定了汉语普通话读音，编制成读音表。该表按汉字字头的笔画数排列先后次序。

7 日本汉字的汉语普通话规范读音表

序号	日本汉字字形	日语读音	汉语普通话规范读音
1	匂	nio - u	yún
2	井	don	dǎn
3	辻	tsuji	shí
4	込	ko - mu	rù
5	凧	tako	jīn
6	杣	soma	shān
7	柶	waku	zá
8	畑	hata hatake	tián
9	栃	tochi	lì
10	柶	toga	méi
11	峠	touge	kǎ
12	俣	mata	yǔ

续表

序号	日本汉字字形	日语读音	汉语普通话规范读音
13	朶	momi	rèn
14	畠	hata hatake	tián
15	雫	shizuku	xià
16	笹	sasa	shì
17	塀	hei	píng
18	楯	sugi	chǎng
19	砦	hazama	yù
20	蛸	ebi	lǎo
21	喰	ku - u	cǎn
22	搾	shibo - ru saku	zhà
23	榊	sakaki	shén
24	働	hatara - ku dou	dòng
25	糗	kouji	huā
26	鞆	tomo	bǐng
27	榎	maki	zhēn
28	榎	kashi	jiān
29	鳴	shigi	tián
30	嘶	hanashi	xīn
31	築	yana	liáng
32	磨	maro	mó

《日本汉字的汉语读音规范（草案）》研制报告

一、研制背景

中国文献翻译外国人名、地名时，凡用非汉字系统文字记录的，多用音译法，如把美国的 Mark Twain 译为“马克·吐温”；把俄国的 Москва 译为“莫斯科”，而中国文献翻译日本的人名、地名等专名时，用的主要是形译法。如人名“長島近夫”（“长岛近夫”），日本人读为“Nagashima Chikao”，移译到中国文献中，汉语的使用者不是像音译其他国家的人名那样读为“那嘎西马·奇卡欧”，而是按中国汉字的字音读成“Chángdǎo Jìnfū”；地名“鎌倉”（“鎌倉”“镰仓”），日本人读为“Kamakura”，移译到中国文献中，汉语的使用者不读为“卡马库拉”，而是按中国汉字的字音读成“Liáncāng”。道理很简单，由于两国都在使用同源的汉字，而记写“長島近夫”（“长岛近夫”）和“鎌倉”（“鎌倉”“镰仓”）的符号正是汉字。当日本人读出“Nagashima Chikao”和“Kamakura”时，中国人不知道指谁和哪儿；当中国人读出“Chángdǎo Jìnfū”和“Liáncāng”时，日本人也不知道指谁和哪儿，但中国人和日本人却通过“長島近夫”（“长岛近夫”）和“鎌倉”（“鎌倉”“镰仓”）的汉字字形达到了一致的指认。这是中日文字独有的现象，长期以来为两国所共同习用。但在日本记写人名、地名用的汉字中，有少量是日本仿照中国汉字的造字原理和结构体式自造的汉字，在日本称之为“国字”（或“和字”“邦字”等）。这些汉字在中国汉字中找不到跟它们相对应的形与音。比如：人名——畠山重忠、近衛文麿，東畑精一；地名——大畠，駒込、札の辻；机构名——椙山女学園（学校名）。其中的“畠”“麿”“畑”“込”“辻”“椙”都是日本自造的汉字。当它们在中国文献中出现时，使用和学习汉语的读者不知道应该怎样读它们的字音。

还有一些字，日本字词书认为是日本自造的汉字，但在中国汉字中也有同样的字形，其中有的在字义上没有共同点，可以视为中日同形（异义）字，有的则

是中国的古字或俗字。这后一类字在中国汉字中有字音，且为一般读者所习知，它们不会给汉语使用者造成读音上的障碍，如“碰”“碗”等，但其中也有一些字在中国汉字中虽有字音，然而对于通用层面的汉语使用者来说，或者字音过于生僻，或者其字形在中国通用层面已经罕用，甚至不用，如“零”“榨”，它们也会给汉语使用者造成读音上的障碍。

我们查阅了几本常见的汉语字词书，为日本汉字注明音义的有：《新华字典》（商务印书馆 2004 年第 10 版）、《现代汉语词典》（商务印书馆 2005 年第 5 版）、《现代汉语规范字典》（语文出版社 1998 年版）、《汉语大字典》（湖北辞书出版社、四川辞书出版社 1986 年版）等，但累加起来，它们收录的日本汉字只有“辻”“畑”“畠”“磨”等几个，为之拟定的读音也缺乏明确的规律性，不便掌握。请见下表：

	新华字典	现代汉语词典	现代汉语规范字典	汉语大字典
辻		shí	shí	
畑	tián	tián	tián	tián
畠		tián	tián	
磨			mǒ	

随着中日两国交流日益增多，日本人名、地名等专名在中国文献中出现的机会将越来越多，为这些日本汉字拟订汉语普通话规范读音就显得更加重要和迫切。《日本汉字的汉语读音规范（草案）》的研制和发布将为汉语普通话口语和书面语的交际，如口头传媒、语文教学（含作为第二语言的汉语教学）、新闻出版、辞书编纂、信息处理等方面的工作提供方便。

二、研制过程

（一）立项、研讨和征求意见

2003 年，教育部语言文字信息管理司立项，由教育部语言文字应用研究所研究员费锦昌和日本东京学艺大学教授松冈荣志组成课题组，承担课题研制任务。

2004 年 10 月 18 日，课题组与上海师范大学语言研究所在上海师范大学联合召开研讨会。出席研讨会的有上海、浙江等地从事语言文字研究、语言文字教学、汉语口头传媒和图书报刊编辑工作的专家学者潘悟云、詹鄞鑫、陆晓光、徐祖友、何华珍等二十多位。

2004 年 11 月 1 日，课题组与北京师范大学文学院在北京师范大学联合召开研

讨会。出席研讨会的有北京从事语言文字研究、语言文字教学、汉语和日语图书编辑、中文信息处理工作的专家学者王宁、李国英、王铁琨、张书岩、厉兵、张轴材、韩秀英等十多位。

两次研讨会研讨的主要问题有：为日本汉字研制汉语普通话读音规范的必要性、制订读音规范的范围和重点、日本汉字的界定、日本汉字的分类、制订读音规范的基本原则和规则、若干日本汉字的定音。

2005 年课题组写出初稿，以论文形式在《语言文字应用》2005 年第 3 期发表，公开征求各界意见，并请中国社会科学院语言研究所研究员张振兴和晁继周审读，同时征询了日本明海大学教授史有为的意见。

（二）鉴定

2008 年 5 月 13 日召开了《日本汉字的汉语读音规范（草案）》鉴定会。专家组一致认为：在中日交流日益频繁的情况下，该项目为缺失汉语读音的日本专名所用汉字拟定读音，对外交、传媒及双语教育等领域是十分必要的，有利于社会语言生活的规范化。该项目所拟读音及规则是科学合理的，有利于应用，达到了课题原定的目标。专家组鉴定通过了该项目，建议在对规范文本进一步完善后，按绿皮书发布使用。

鉴定专家组组长为王宁（北京师范大学），成员有张振兴（中国社会科学院语言研究所）、李国英（北京师范大学）、张轴材（北京书同文数字化技术有限公司）和韩秀英（商务印书馆）。

三、研制原则

为日本汉字拟定汉语普通话规范读音时，课题组始终遵循的主要原则有三条：

（一）这一规范的服务对象是汉语的使用者和学习者。在拟定汉语普通话规范读音时不必过多考虑这些汉字在日语中原有的读音。比如“丼”，在日语中记录的是一种类似中国盖浇饭的食品，读 don。但汉语中同形的“丼”，有两个读音，一是 jǐng，同“井”；二是 dǎn。我们决定不取日语的 don 音，而在汉语已有读音中选取 dǎn 音，因为它既是汉语原有的音读，又跟日语读音较为接近。

（二）中国字词书已有且不过分生僻的读音，尽量沿用。严格控制拟订新音项的数量。

（三）拟定的读音一定要尽可能方便汉语的使用者和学习者。实现这一条的前

提条件是这一规范的拟音具有人为性。具体的方法就是尽量使这些拟音符合中国读者早就建立起来的、平时用得最多的字形与字音之间的习惯联系方式，这就是形声字通过声旁表示字音的方法，有的学者称之为“读音形声化”。如“畑”“辻”“込”“嶋”等字，能视为形声字或会意兼形声字的，就按形声字的原理定音，“畑”就取 tián（田）音，“辻”就取 shí（十）音，“込”就取 rù（入）音，“嶋”就取 tián（田）音。有的字在中国字词书中的注音过于生僻，容易造成汉语使用者误读的，可根据字形另定读音。如“𪛗”，《汉语大字典》注 ní，假若作为日本汉字的汉语普通话读音，很容易造成汉语使用者误读，给学习和使用带来不便，不如干脆按形声字处理，取 rèn（刃）音。有些“声旁”是中国现行汉字中没有的部件，就仿取形似部件的读音，如“𪛗”，可视为形声字，但右边“声旁”是中国现行汉字中没有的部件。考虑到这一部件跟“卡”形似，中国读者见到这个部件时一般都会往“卡”字上套，定音时就取 kǎ 音。有的先生提出，把“上”“下”重叠起来的部件读成“卡”音，会不会造成字形上的混淆。我们认为，这个提示很重要，但考虑到这一拟音只为使用汉语的读者服务，而通用汉字中没有“上”“下”重叠起来构成的部件，只要在中国读者学习日文时予以提醒，就完全可以避免混淆。再如“𪛗”，有的学者考证右旁是中国汉字“卒”的俗写，但现在中国通用层面的读者往往会把这个字的右旁往简化字“杂”上靠，我们决定不取 zú（卒）音，而取 zá（杂）音。

四、研制方法和内容

（一）确定拟音对象

现行日本文字包含着两种符号系统，一种是假名，一种是汉字。凡是日本用来记录日语的汉字都是“日本汉字”。比如“松”，尽管它是从中国引进的，字义也跟汉语中的“松”大致相同，但它记录的是日语中“マツ”这个词，它的身份已经是日本文字。李思敬曾经打过一个比方，说日本用的大多数汉字，就好像一个加入外国籍的中国人，尽管他仍然是汉民族的血统，但已经变成外国人，不能再行使中国的公民权了。（李思敬《日本人名、地名汉字的汉语读音》，《语文建设》1992年第9期）这里的“日本汉字”是泛指日本记录日语时使用的所有汉字，而本规范要为之拟订汉语普通话规范读音的则仅是其中的一小部分，主要是日本仿照中国汉字的造字原理和结构体式自创的汉字，在日本称为“国字”（也有称为“和字”“邦字”的），它们的形音义都是中国汉字系统中没有的，如“𪛗”“𪛗”“辻”“𪛗”“𪛗”等。还有一些在中国汉字中有字音，但对于中国通

