中国语言生活绿皮书
国家语言文字工作委员会发布

B009

中国 Language Situation in China: 2009

语言生活状况报告

下编

国家语言资源监测与研究中心 编

ZHONGGUO YUYAN SHENGHUO ZHUANGKUANG BAOGAO (2009)

2009

商務印書館 THE COMMERCIAL PRESS
创于1897

中国语言生活绿皮书

国家语言文字工作委员会发布

YW B009

中国语言生活状况报告

(2009)

下　编

商務印書館

2010年·北京

图书在版编目（CIP）数据

中国语言生活状况报告. 2009. 下编/国家语言资源监测与研究中心编. —北京：商务印书馆，2010. 10
（中国语言生活绿皮书）
ISBN 978-7-100-07281-6

Ⅰ. ①中… Ⅱ. ①国… Ⅲ. ①社会语言学—研究报告—中国—2009 Ⅳ. ①H1

中国版本图书馆 CIP 数据核字（2010）第 129447 号

ZHŌNGGUÓ YǓYÁN SHĒNGHUÓ ZHUÀNGKUÀNG BÀOGÀO（2009）

中国语言生活状况报告（2009）

下　编

国家语言资源监测与研究中心　编

商　务　印　书　馆　出　版
（北京王府井大街36号　邮政编码 100710）
商　务　印　书　馆　发　行
北京瑞古冠中印刷厂印刷
ISBN 978-7-100-07281-6

2010 年 10 月第 1 版　　开本 787×1092　1/16
2010 年 10 月北京第 1 次印刷　　印张 34½

定价：56.00 元

顾　问　许嘉璐　赵沁平　郝　平　李卫红
策　划　教育部语言文字信息管理司
主　编　李宇明
审　订　陈章太　戴庆厦　陆俭明　邢福义

上　编

主　编　周庆生
副主编　郭　熙　周洪波
作　者　（按音序排列）

白　娟	蔡长虹	陈　慧	陈章太	程祥徽	崔晓飞
邸宇维	高建平	高陆洋	郭济修	郭　熙	郝阿庆
何　瑞	和丽峰	黄锦章	汲传波	孔江平	李　俏
李晓华	李旭练	李　阳	李宇明	林　涛	刘　畅
骆　峰	毛力群	倪　兰	曲彦斌	孙晓先	田　鹏
汪　锋	汪　磊	王翠叶	王　奇	王志娟	魏　丹
谢俊英	余桂林	战　菊	张　军	张日培	张　艳
张映川	赵守辉	赵小兵	赵晓敏	郑梦娟	周道娟
周洪波	周庆生	祝晓宏	邹玉华		

下　编

主　编　王铁琨
第一副主编　侯　敏
副主编　杨尔弘　苏新春　何婷婷　赵小兵
作　者　（按音序排列）

阿不都热依木·沙力	阿力木·木拉提	艾孜古丽·玉素甫		
曹　晖	陈　敏	陈　琪	陈　雪	崔　乐	高　璐
郭曙纶	何婷婷	何　伟	侯　敏	江格瓦尔·地汗	
阚明刚	李　安	李　钢	李艳娇	李永宏	刘　华
刘　佳	刘　俊	刘　薇	刘欣斐	潘伟民	齐向卫
祁坤钰	苏小康	苏新春	滕永林	通拉嘎	涂新辉
王华英	王　磊	王　宁	王　奇	王秋萍	王铁琨
王　燕	王宇波	魏　励	吴继媛	杨尔弘	杨　江
于洪志	玉素甫·艾白都拉	曾青青	曾小兵	詹祥妹	
张红春	张金爽	张　蕾	张　勇	张志平	赵小兵
庄晓云	邹　煜				

目　　录

Contents

报纸、广播电视、网络(新闻)用字用语调查

调 查 报 告

国家语言资源监测与研究中心利用国家语言资源监测语料库(包括平面媒体、有声媒体、网络媒体)的年度语料,已经连续发布了2005年度、2006年度、2007年度、2008年度的语言生活状况报告。其中,报纸、广播电视、网络(新闻)的用字用语调查是每年都进行的调查项目。报纸、广播电视、网络(新闻)的年度用字用语状况,可以反映媒体年度的语言使用实态,也可以透过这些字词语的使用状况看到年度的社会热点、重大事件等。2009年度的用字用语调查是在国家语言资源监测语料库2009年度的语料上进行的。

历时五年的语料积累,既可以用来反映五年共时的语言生活,也可以通过对比分析来反映五年历时的语言变化。在本报告的第四部分,对五年的字词语调查数据进行了比较,以期从共时、历时两个角度对五年的字词语使用状况进行分析。

一 调查使用的语料及调查内容

(一)调查使用的语料

本年度的调查语料涵盖平面媒体、有声媒体、网络媒体三种,共计1 249 387个文本文件,1 237 492 014字符次(包括标点、符号及西文字母、数字等出现的次数),其中汉字出现1 007 019 960字次。

2009年度语料采集的依据及选择过程与往年一致。在进行此项调查时,为了使年度间的数据规模基本保持在同一个数量级上,总字符次保持在12亿左右,总汉字次保持在10亿左右。平面、有声、网络媒体的语料量仍按5∶1∶4的比例进行选取。

1. 报纸

平面媒体选择了2009年度15种报纸作为调查语料，选择时综合考虑了“发行量、发行地域、发行周期、媒体价值”等因素，同时考虑了语料的可获得性因素。发行量参考了2008年6月2日召开的第61届世界报业大会（瑞典·哥德堡）发布的“2008年世界日报发行量前100名排行榜”（中国部分）；媒体价值参考了由2008世界品牌大会（中国·北京）于2008年6月2日发布的“2008年《中国500最具价值品牌》排行榜”。

选定的15种报纸是（按音序排列）：《北京青年报》《北京日报》《北京晚报》《法制日报》《光明日报》《广州日报》《华西都市报》《今晚报》《南方周末》《钱江晚报》《人民日报》《深圳特区报》《羊城晚报》《扬子晚报》《中国青年报》。

报纸语料共计795 046个文本，633 705 267字符次，其中汉字出现511 924 302字次。

2. 广播电视

广播电视语料是根据播出的录音或录像转写的文本，选取的主要依据是节目收视率，并且综合考虑了传播媒介（广播、电视）、媒体级别（中央、地方）、传播广度（是否上星）、播出时间（是否黄金时段）、节目样态（独白、对话、综合）、文本现存（是否有转写好的文本）等因素。

2009年度选取的广播电视语料包括：中央电视台、北京电视台、上海文广新闻传媒集团（电视）、天津电视台、重庆电视台、广州电视台、山东电视台、山西电视台、安徽电视台、河南电视台、哈尔滨电视台、深圳电视台、石家庄电视台、洛阳电视台等14家电视台，以及中央人民广播电台、北京人民广播电台、上海文广新闻传媒集团（广播）、天津人民广播电台、重庆人民广播电台、山东人民广播电台、深圳人民广播电台、石家庄人民广播电台、洛阳人民广播电台、包头人民广播电台等10家广播电台，总计169个栏目，20 219个文本。

广播电视语料量为124 884 433字符次，其中汉字出现102 707 631字次。

3. 网络（新闻）

根据年度调查所确定的三种媒体语料量的比例，网络媒体语料只选取了新浪、腾讯两个网站的部分新闻语料。选取的方式是，对已采集的两个网站的全部语料分别以两个网站每一天的所有语料为单位，在一天的语料中按既定比例随机抽取文本，然后将以天为单位抽取得到的所有文本集合在一起，形成网络（新闻）的样本语料。

由此获得的语料共计 434 122 个文本,478 902 314 字符次,其中汉字出现 392 388 027 字次。

4. 语料说明

报纸语料是网络版的。广播电视语料是由广播电视节目转写的文本,与原始有声语料之间存在些许差异。网络(新闻)语料来自新浪、腾讯 2009 年度的新闻页面,这些语料全部从网络下载而得。用于调查的语料是纯文本的格式,是由相应的页面文件转换而成的,转换过程中去除 HTML 标签信息、广告信息,报纸语料中去除转版信息、图片标题等短文本。

(二) 调查内容

本次调查的对象是汉字和词语。词语是计算机自动分词产生的分词单位,既包括语文词,也包括专名(人名、地名、组织机构名、其他专名)、时间表达式(如"2009 年 11 月、5 月、15 点 30 分")以及结合紧密、使用稳定的短语(如"经济危机、世博会")。调查项目主要有频次、频率、累加频率、出现文本数、使用率、累加使用率等,调查内容既包括 2009 年度内三种媒体语言的共时考察,也含有 2005—2009 五个年度间的历时考察。其中,频次、频率、累加频率、出现文本数和使用率的含义及计算方法与 2005—2008 年度的《报纸、广播电视、网络(新闻)用字用语调查》[①]相同,也可参考《中国语言生活状况报告(2008)》下编的"语言资源监测与研究相关术语"。[②]

二 汉字使用情况

(一) 说明

1. 本次统计没有甄别文本中的别字。
2. 本次统计不包括汉字部件、乱码、无法显示的字符等。

① 见国家语言资源监测与研究中心编《中国语言生活状况报告(2006)》下编第 2 页,商务印书馆 2007 年版。

② 见国家语言资源监测与研究中心编《中国语言生活状况报告(2008)》下编第 493—501 页,商务印书馆 2009 年版。

（二）基本情况

1. 字符总数：指全部语料中汉字、标点、符号等的总量，计 1 237 492 014 字符次。

2. 汉字总数：指全部语料中汉字出现的总字次，计 1 007 019 960 字次。

3. 字种数：指字形不同的汉字种数，计 10 204 个。

4. 共用字种数：指报纸、广播电视、网络三种媒体中都出现的汉字，计 6 238个。

5. 独用字种数：指只在报纸、广播电视、网络某一种媒体中出现的汉字，计 2 359个。

汉字使用情况的具体数据见表 1-1。三种媒体用字之间的关系除了表中列出的共用、独用字种之外，还有部分共用字种的情况。部分共用字种是指只出现在任意两种媒体中，而没有出现在第三种媒体中的汉字种数，比如报纸与广播电视的部分共用字种计 165 个，报纸与网络（新闻）的部分共用字种计 1 357 个，这两部分的并集形成了报纸的部分共用字种，共计 1 522 个。为了使数据更加简洁、直观，表中没有列出部分共用字种数。

本次调查的全部语料用字（共计 10 204 个汉字）形成 2009 年度用字总表[①]（简称“用字总表”），表中包括了汉字在语料中出现的频次和文本数两个最基础的数据，所有汉字根据使用频率由大到小排序。

表 1-1　2009 年度汉字使用情况

媒体	总字次	字种数	共用字种数	独用字种数
报纸	511 924 302	9 317	6 238	1 557
广播电视	102 707 631	6 597		109
网络（新闻）	392 388 027	8 373		693
全部语料	1 007 019 960	10 204		2 359

（三）汉字的覆盖率

汉字的覆盖率是调查的重要项目之一。它是反映汉字常用与否的重要指标，同时反映了在整个调查语料中汉字使用的分布情况。其统计结果见表 1-2。

① 见国家语言资源监测与研究中心编《中国语言生活状况报告（2009）》下编第 32－170 页，商务印书馆 2010 年版。

表 1-2　2009 年度汉字对语料的覆盖情况

覆盖率 / 语料	达到 80%		达到 90%		达到 99%		达到 100%
	字种数	字种比例(%)	字种数	字种比例(%)	字种数	字种比例(%)	字种数
报纸	609	6.54	982	10.54	2 460	26.40	9 317
广播电视	546	8.28	893	13.54	2 298	34.83	6 597
网络(新闻)	584	6.97	943	11.26	2 294	27.40	8 373
全部语料	602	5.90	970	9.51	2 400	23.52	10 204

(四) 与现行规范字表的比较

1. 用字总表前 2 500 字与一级常用字比较

用字总表前 2 500 字与《现代汉语常用字表》[①]一级常用字(2 500 字)比较，用字总表中有 342 字是一级常用字中所没有的。将用字总表按照前 500 字、501 至 1 500 字、1 501 至 2 500 字分为三段，每一段中没有出现在一级常用字中的汉字列于表 1-3。

表 1-3　用字总表前 2 500 字与一级常用字比较

范围	一级常用字之外的字
前 500 字	尔(1 个)
501—1 500 字	媒 圳 频 韩 伊 诺 措 迪 辑 综 俄 萨 澳 伦 曼 菲 姆 莱 洛 郭 聘 署 杭 蒂 姚 账 谓 咨 凌 拓 卢 屏 娜 埃 拟 邓 沪 弗 艾 鹏 琳 蔡 邦 贾 冯 浦 翔 斌 谐 逊(50 个)
1 501—2 500 字	兹 帕 曹 穆 敦 胎 赫 潘 秦 肖 募 玲 曝 蒋 徽 聊 莞 霍 彭 粤 颁 颇 氛 韦 墅 兑 戈 奈 涵 诈 莉 吕 袁 枚 颖 鸿 卦 履 玛 魏 讼 崔 涯 谭 癌 硕 卓 侯 廷 铭 郁 吁 肇 憾 匿 寓 耶 砸 怡 旭 婴 淮 茨 厢 詹 晖 碳 峻 汰 磊 辖 庞 抑 妮 妆 莎 契 鑫 镑 翰 瓷 魅 舆 柯 苑 蕾 岳 晰 雯 铝 湘 娟 琼 逻 逸 嘛 鼎 逾 汶 薇 歧 坤 昊 骏 邱 婷 滞 芯 赋 淑 杉 绎 邵 蓉 彰 弥 罕 雇 尹 幽 琪 薛 勃 坠 彦 弘 轴 淀 衷 揽 仲 巢 侠 奢 暨 挫 瘾 溢 萧 楠 鲍 浏 赁 擅 尬 尴 瞬 飙 豫 菱 亨 撰 彬 坪 宠 渝 卿 殷 廖 奠 睿 澄 爵 澜 馨 珊 佐 瘤 坎 邹 勘 擎 祭 裸 轩 熙 玮 啤 凰 矶 悍 淫 霆 奎 啥 呵 陌 蔓 靓 槛 萎 遂 啸 撼 腺 馈 赎 倩 坞 凸 逛 刹 俞 睐 龚 冕 焕 嫒 奕 殴 蕴 潭 玫 蔚 吻 喻 咖 窦 瑜 莹 侣 伽 郝 芭 缅 谍 衍 辐 骚 鹤 腕 啡 昔 现 扳 幢 甸 瑟 仕 靖 邢 隧 麟 迄 钦 崛 函 贬 倪 玄 惟 埔 噪 晤 溃 闫 遏 娅 崩 粹 堰 秉 媳 吾 勋 寂 懈 碟 睹 瞩 腻 缉 枢 湛 藤 聂 蟹 炫 皓 歹 禹 豹 潇 讶 喀 硅 裔 炜 帷 荔 儒(291 个)

① 国家语言文字工作委员会、国家教育委员会 1988 年联合发布。

用字总表的前 2 500 字覆盖了全部语料的 99.15%，是媒体的年度高频用字，这些字与一级常用字之间的差异，反映了年度语言生活和媒体用字的一些特点。比如，“杭、拓、婴、碳、汶、飙”等字的使用频率升高，与当年的社会生活息息相关，如 2009 年大众关注的“汶川地震”的灾后重建，“婴幼儿奶粉、杭州飙车案、力拓”等事件，全世界都在热议的“低碳”话题；媒体中有大量的专名被使用，像“尔、诺、迪、萨、伦、曼、菲、姆、莱、洛、郭、蒂、娜”等多是人名、译名用字；此外，“姚、卢、翔、逊、兹、彦”等汉字，则是由于“姚明”受伤后成为老板、“卢武铉”政治风波、“刘翔”复出、“迈克尔·杰克逊”辞世、“伍兹”大起大落、学习“王彦生”等成为媒体关注的热点事件，因而其使用频率也提高了许多。

2. 用字总表前 3 500 字与《现代汉语常用字表》比较

用字总表前 3 500 字与《现代汉语常用字表》(3 500 字)比较，用字总表中有 401 字是《现代汉语常用字表》中所没有的，表 1-4 分段列出了用字总表中有，而没有出现在《现代汉语常用字表》中的汉字。

表 1-4　用字总表前 3 500 字与《现代汉语常用字表》比较

范围	《现代汉语常用字表》之外的字
前 1 000 字	圳 迪(2 个)
1 001—1 500 字	弗 蔡 斌(3 个)
1 501—2 500 字	兹 曝 莞 韦 肇 耶 怡 茨 詹 晖 磊 妮 莎 鑫 镑 魅 柯 苑 雯 娟 嘛 汶 昊 邱 婷 邵 尹 琪 彦 弘 暨 楠 鲍 浏 尬 尴 飙 亨 渝 廖 睿 馨 佐 邹 玮 矶 霆 奎 靓 槛 馈 倩 俞 睐 龚 嫒 窦 瑜 伽 郝 仕 麟 崛 倪 惟 埔 闫 娅 吾 瞩 缉 湛 皓 禹 潇 喀 裔 炜 帷(79 个)
2 501—3 500 字	涅 汕 禅 韶 洙 淇 铉 昕 瑶 弈 琦 舜 妍 巅 霖 磋 朔 哦 侃 闵 绯 鄂 炳 盎 璐 诠 珀 厄 裴 卉 茜 胺 迭 甄 郡 轶 烨 尧 阮 牟 晏 殡 黛 墟 淄 禄 滕 驭 阜 妃 岚 鲨 佟 羌 璇 瑛 跻 圭 祁 沂 咋 菁 颐 萃 茹 酉 俑 曦 莓 戛 荫 拎 籁 坍 翟 粽 禹 锂 汀 喆 冉 羁 臻 孜 汝 姬 芸 慑 霄 麒 庚 娥 襄 嵩 汾 钰 嘘 铂 峨 泸 狄 虞 铮 翡 斐 氰 潍 瀚 聆 佼 郅 峥 婕 亟 雍 辗 绚 郜 钊 婧 丫 岐 俨 缪 鹭 隋 豚 赣 猝 峙 皖 陇 蓓 栩 佬 梓 煲 袂 渎 陀 孚 岑 黏 嫖 焱 霾 炙 匡 煽 曰 辄 镍 骸 绮 沁 铀 臧 麓 娴 藉 苯 拽 荟 哇 禧 憬 蹿 阪 葆 憧 漪 淼 韬 鳌 黯 瑕 愣 嘟 琶 毋 寅 钰 姗 蹊 竺 铬 驿 郴 踝 钛 妞 隽 邯 媲 嘻 炯 溧 璀 熠 炖 晟 烯 匮 璨 碁 兮 髦 琛 狙 缮 恪 暧 峪 毗 镖 邸 桓 咎 鄞 抉 亢 釜 惬 嘎 沓 荼 瞿 敖 篾 萱 祀 酶 浒 醛 祺 睫 祛 抨 诟 忡 瑾 逍 余 煜 嵘 俪 泓 桐 跤 噱 骥 罄 裘 铎 辍 榻 悖 汲 纶 氟 邡 濮 滇 镁 蓟 疵 癫 忻 漳 甬 遴 赟 坂 膺 戎 邬 眩 踵 覃 枭 啪 珑 獬 诧 跆 靳 迥 玺 罹 嗯 呱 瑙 锵 嫣 矣 骼 杞 赈 悸 孰 宕 泗 哉 烽 嫦 厮 栾 渲 绥 珂 踞 冥 骁 皋 槟 伎 饽(317 个)

3. 用字总表前 7 000 字与《现代汉语通用字表》①比较

用字总表前 7 000 字与《现代汉语通用字表》(7 000 字)比较,用字总表中有 725 字是《现代汉语通用字表》中所没有的。具体情况见表 1-5。

表 1-5　用字总表前 7 000 字与《现代汉语通用字表》比较

范围	《现代汉语通用字表》之外的汉字
前 3 000 字	闫 喆(2 个)
3 001—5 000 字	淼 碁 飚 珺 囧 濛 婳 玟 堃 玘 垚 翀 镕 瘀 晳 哒 芃 珮 槃 琍 时 崧 钜 昇 崚 焗 磡 弢 祐 噁 仝 孖 鲶 邨 珅 唰 暐 吒 甹 捱 後 埗 巿 琊 劼 岫 朮 塚 囍 楳 啫 曝 戉 犇(54 个)
5 001—7 000 字	锺 嘢 椏 玹 徳 嫣 迳 瓅 睒 簕 劵 炆 龑 峯 咁 谘 駮 氹 甦 這 玭 鋮 嘅 堀 暎 罂 冇 積 炤 诶 咗 咋 個 烎 霑 甯 湧 還 會 勢 崑 旼 呎 吵 點 炘 筦 姮 沒 撐 飏 苾 椠 姵 妳 賛 瑗 穌 塱 蹚 陞 娭 黃 馳 內 負 畊 査 栢 澔 澋 來 粿 郎 線 誌 闇 瑱 詠 潡 洸 廼 艸 砦 艄 壓 娯 現 槤 辻 謝 別 時 燚 椹 褚 係 頫 埇 孅 冪 蘋 髙 厶 纮 蚵 迺 峬 塝 佢 鹟 姳 尓 咲 孃 腁 浕 為 呲 媸 搧 凃 牤 盤 福 國 铓 槑 關 調 吶 問 茈 價 訢 佡 噺 續 臺 師 媞 話 鮰 呯 炣 哨 啟 鸥 將 迴 嶺 壆 硷 虒 過 衹 買 沓 給 孈 跩 驩 岕 椗 趨 絜 鸻 坣 遖 鍂 叡 觀 繼 壘 剎 侶 粄 菴 璨 徹 珪 進 萬 鑑 晧 睩 畑 煖 業 汎 啓 嚟 麼 潮 爀 鹛 乓 漲 堺 畝 艉 蓢 溋 長 穂 說 機 舵 搵 菈 題 熺 硃 場 愛 恵 奕 虓 醜 經 叐 宮 牲 抛 醣 丐 啱 栞 覺 礡 礅 佺 兒 蒨 開 冏 裡 鲋 嫈 偊 吖 噏 強 棐 腘 鲉 昳 資 彝 樸 請 兩 態 夆 宧 幾 蘭 沂 彥 該 鳌 較 冚 燦 勳 皀 譞 珖 謦 髈 薗 甗 曚 郃 變 發 灊 傢 對 見 亖 嗎 礽 渼 昉 禛 秈 卍 倮 垴 糬 晄 魟 崀 梶 區 們 澂 牠 彤 鹛 嘉 媗 問 韡 騄 鹋 製 倖 涉 貐 並 鐏 嵒 瑩 徵 奧 動 攵 抲 鏇 裝 熙 浬 藥 昡 佈 晢 滉 鷊 圖 蛯 挌 預 黒 珵 錸 滾 芶 頭 濕 枱 明 芤 汇 蓮 欷 顓 嗦 獻 競 岽 榊 暉 應 顏 淯 耑 東 暎 荖 勣 蠔 暫 弨 遾 楢 潑 鄗 璧 數 沋 錞 橫 鶤 鷗 曽 脩 説 卲 孃 嶋 澎 姫 閈 睍 濬 塊 奻 孮 龍 靚 翃 芇 鉉 冴 畬 篠 遊 結 雲 沖 罖 嵒 菉 枬 珞 砵 娒 昮 荾 況 劻 婥 吔 範 績 赀 廻 風 鍊 荊 彈 涼 莊 夗 眛 録 慮 蓧 剋 騋 塭 鬱 谞 髎 畠 児 岺 蒐 杕 砢 紘 囗 谿 兲 炏 柿 坵 出 篧 缋 屌 鷰 妺 議 纟 樋 怗 蹬 栊 鵟 饄 骕 痖 損 讚 盨 霂 牞 槅 詷 險 歴 皕 眪 瑨 語 銀 沬 閤 骉 廸 锠 瑅 無 億 繆 馬 瑯 盪 記 傑 嫑 車 疋 從 囯 毬 镙 侘 犽 栃 埤 認 書 暨 產 鳎 穀 苟 澙 絊 霙 劄 墻 堽 報 泲 囘 費 徬 烔 暇 孫 玆 鱤 島 唠 餘 晅 錯 彫 樂 獎 佷 槙 匐 鸮 樣 學 暸 碻 螠 罥 週 曼 箇 箝 稅 舲 蔻 鬆 參 猇 岡 皞 華 夲 實 墫 勝 騉 準 崐 钖 菸 俤 祺 螃 达 枏 茲 嬛 測 荳 連 瓴 湑 篪 嗯 蟳 篺 饾 榎 肶 昪 烃 钸 浀 義 伲 忞 當 許 咡 約 碁 鱓 瀼 貟 錢 卬 級 杩 姈 抹 瀨 睍 陜 陸 僜 凱 刓 崙 豊 懔 儁 體 陽 窻 畠 泭 泱 採 線 甥 賣 增 隻 扞 殭 電 崐 貌 啫 両 隨 實 堞 優 繻 嬲 響 丟 鈿 晛 喬 倉 煙 門 總 湳 滈 鏸 難 戸 棫 絶 銘 偦 鈜 沕 張 潦 緤 禅 漼 鎔(669 个)

① 国家语言文字工作委员会、中华人民共和国新闻出版署 1988 年联合发布。

在历年的年度用字总表前 7 000 字与《现代汉语通用字表》的比较中，2009 年度用字与《现代汉语通用字表》的差别是最大的。[①] 探究其中的原因，可归为如下几个方面：

（1）年度语言生活由于汉字的“繁简之争”“《通用规范汉字表》（征求意见稿）公开征求意见”而受到社会的普遍关注。众多语言文字工作者和广大民众对这些问题提出了自己的建议，在媒体中有许多讨论，其中不乏对文字规范工作计深虑远的建言献策，这使 2009 年度的用字总数增加，繁体字、生僻字增加。

（2）网络语言中汉字的使用标新立异，且传播速度快。2008 年度“囧、冏、槑”等汉字在网络上的流行，2009 年度众多生僻字也在网络上被发掘，尤其一些字形比较奇特的合体汉字，如“烎、嘂、嘦、王、奣、曼、嫑、巭、嫑、怹、兲、烎、忈、烪、勥、嘞”。在网络世界里，这些汉字的本义已经不重要了，字形是其在现代社会表义的重要依据。如“兲”不再表示“天”的本义，而是“王八”义；“烪”不再表示“光”，而是“火化”义。但有些仍保持原义，如“嫑”表示“不要”义，字音则由“不”的声母与“要”的韵母组成。此外，由于汉字“槑”在网上的流行，似乎在追随这样一种汉字结构带来的效果，一些由两个或三个相同的汉字成字部件构成的汉字在网络中也悄然流行开来，如“甡、囍、孖、屾、皕、垚、骉、犇、嚞、畾、燚”等。这些现象值得进一步关注与探究。

这 725 个汉字包括：不规范的简化字（“镟、钖、线、钜、椏、锺、迳”）7 个、繁体字（如“噁、後、瓅、龑、這、會、點”等）211 个、异体字（如“陞、廼、堃、喆”等）113 个、旧印刷字形（如“曁、茍、弑、玆、稅、沒、別”等）22 个、日本汉字（如“禅、畑、沢、麼、恵”等）21 个、旧计量字（“浬、吋、呎”）3 个。

4.《现代汉语通用字表》与用字总表比较

用字总表中未出现的《现代汉语通用字表》中的字有 162 个，具体如下：

锿 揞 鞁 鞞 鞴 鐾 琫 鳊 沘 诐 瘭 瀌 袯 皤 猹 蹅 镵 幝 疢 牚 瓻 犨 脞 忉 阽
刭 弹 饳 墦 荆 锇 钹 艴 馉 诖 鳍 涫 瓯 渱 轩 穰 萑 澴 埌 锽 锪 姞 鲇 檵 䂇
椟 戋 葛 淈 贶 祲 鼱 刭 僦 腒 溴 醵 锩 桊 觖 潏 肶 菁 莰 阌 硁 纩 裈 悢 饕
膁 裣 菽 辌 飕 僇 稆 朒 芼 钔 幪 醾 敉 萛 礳 镎 臑 柈 堋 狉 鲏 庀 澼 氕 帡
醭 鲯 尯 蠷 鳈 蘘 灏 澈 胂 鸤 螋 嗾 瞍 蹜 谇 阘 啴 绹 薮 龆 柊 酴 鳂 辒 邰
阢 侬 绤 墈 婞 滃 痃 芫 枸 庡 榼 繄 庈 酏 勚 鹝 蝓 憖 莸 溷 蜎 驵 篰 漾 瘵

---

① 见国家语言资源监测与研究中心编《中国语言生活状况报告（2009）》下编第 30 页，商务印书馆 2010 年版。

絷 蹢 痔 湦 锧 鸼 瘃 藡 腙 瘢 躜 槁

(五) 汉字使用的其他情况

用字总表中的繁体字、异体字、不规范的类推简化字、旧计量单位用字、日本汉字等，出现在用字总表的 2 874 位以后，即覆盖率达到 99.51%之后，与去年的 3 118 位相比有所提前。2009 年度汉字的使用情况相对来说比较活跃，与该年度大众积极关注并参与语言文字工作有较大的关系。具体情况见表 1-6。

表 1-6　汉字使用的其他情况统计

类型	全部	报纸	广播电视	网络(新闻)
旧印刷字形	86	65	9	65
繁体字	1 217	769	234	762
不规范的简化字	17	16	6	10
旧计量用字	4	3	1	3
异体字	396	335	64	241
日本汉字	84	61	13	60

(六) 报纸、广播电视、网络(新闻)语料与全部语料的汉字频率比值

计算报纸、广播电视、网络(新闻)各媒体前 2 500 个年度高频字与全部语料中相应汉字的频率比值[①]，并按照频率比值的降序排列比较结果，可以观察三种媒体语料的用字特点。表 1-7 分别列出了报纸、广播电视、网络(新闻)三种媒体语料中频率比值排在前 20 位的汉字。

表 1-7　报纸、广播电视、网络(新闻)汉字频率比值分析

媒体	前 2 500 字中频率比值在前的 20 个汉字
报纸	墅 龄 粤 腺 圳 荔 践 禺 塘 废 瘤 绣 穗 尿 晴 肠 幢 厨 胃 奴
广播电视	伽 窦 咱 嘛 呢 扁 您 殖 饲 啊 邢 呀 它 么 旱 猪 吗 怎 你 徽
网络(新闻)	卦 浪 匿 描 页 寸 蒂 娱 踢 篮 弗 浏 球 芯 杆 霆 扳 兹 玮 棋

1. 从表 1-7 可以看出，广播电视语料的口语化特征最为显著，记录语气词、称谓词的汉字居多；报纸语料中“粤、圳、禺”等显示了地域性报纸的特点；网络媒

① 见国家语言资源监测与研究中心编《中国语言生活状况报告(2008)》下编第 498 页，商务印书馆 2009 年版。

体中的“浪、页、浏”等表明了网络媒体语料采集来源及形式的特点。

2. 媒体的用字特色也体现了内容上的不同侧重点，报纸语料的“墅、幢、奴”等与社会生活中的房地产有关；网络媒体上“娱、踢、篮、球、兹、棋”等则反映了网络媒体关注文体娱乐的特质。这些汉字在用字总表以及各媒体用字表中的频率、排序情况见附表1。

三　词语使用情况

本次调查仍然使用中国科学院自动化研究所研制的分词标注系统。

（一）基本情况

1. 分词单位总数：指由分词软件对语料切分得到的字符串的总数，计717 321 946次。其中标点、符号等出现123 795 923次，其他分词单位出现593 526 023次。

2. 总词语数：即不包含标点、符号、纯西文、纯阿拉伯数字、数字与西文混合式、网址等的分词单位，共计592 414 821词次。

3. 词种数：2 348 100个。

4. 共用词种数：193 416个。共用词种是指报纸、广播电视、网络（新闻）三种媒体语料库都用到的词语。基本数据见表1-8。与汉字的调查相同，这里只列出了媒体间的独用、共用情况，未列部分共用的情况。

表1-8　2009年度词语使用情况

媒体	总词语数	词种数	共用		独用	
			词种数	比例（%）	词种数	比例（%）
报纸	298 944 198	1 579 054	193 416	12.25	1 049 055	66.44
广播电视	60 964 722	432 014		44.77	165 638	38.34
网络（新闻）	232 505 901	1 084 828		17.83	579 032	53.38
全部语料	592 414 821	2 348 100		8.24	1 793 725	76.39

（二）词语的覆盖率

表1-9列出了覆盖率为10%到90%和91%到100%各段的词种情况。从表中可以看出，词种数的明显上升是在覆盖率为90%之后，覆盖率99%到100%