

中国语言生活绿皮书
国家语言文字工作委员会发布

 B007

中国

Language Situation in China: 2008

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 (2008)

2008



商务印书馆
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



中国语言生活状况报告 (2008)

下 编

国家语言资源监测与研究中心 编

商务印书馆

2009年·北京

图书在版编目 (CIP) 数据

中国语言生活状况报告. 2008. 下编/国家语言资源监测与研究
中心编. —北京: 商务印书馆, 2009
(中国语言生活绿皮书)
ISBN 978-7-100-06700-3

I. 中… II. 国… III. 社会语言学—研究报告—中国—2008
IV. H1

中国版本图书馆 CIP 数据核字(2009)第 111918 号

所有权利保留。

未经许可, 不得以任何方式使用。

ZHŌNGGUÓ YǔYÁN SHĒNGHUÓ ZHUÀNGKUÀNG BÀOGÀO (2008)

中国语言生活状况报告 (2008)

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 978-7-100-06700-3

2009年9月第1版

开本 787 × 1092 1/16

2009年9月北京第1次印刷

印张 37

定价: 62.00 元

目 录

报纸、广播电视、网络(新闻)用字用语调查	001
调查报告	001
用字总表	030
高频词语表	162
年度新词语调查	320
调查报告	320
新词语表	331
中国媒体年度流行语	397
中文网络用字用语专项调查	407
基础教育新课标历史、地理教材用字用语调查	440
语言资源监测与研究相关术语(2009 版)	493
附录	502
东南亚华文媒体用字用语调查	502
调查报告	502
东南亚华文媒体特色词语表	528
图表目录	569
术语索引	576
后记	580

Contents

Survey of the Chinese Words and Expressions in Newspaper, Radio,	
Television and Internet (News)	001
A Research Report	001
A List of General Words	030
A List of High-frequency Words	162
Survey of New Words of the Year	320
A Research Report	320
A List of New Words	331
Popular Words and Phrases of the Year Used by Chinese Media	397
A Special Survey of Words and Expressions Used in Chinese Websites	407
Survey of the Use of the Chinese Language in the History and Geography	
Textbooks (New Curriculum Standards) for Elementary Education	440
Technical Terms Related to the Monitoring and Research on the National	
Language Resources (2009 Edition)	493
Appendices	502
Survey of Words and Phrases Used by the Chinese-language Media in	
Southeast Asia	502
A Research Report	502
A List of Special Chinese Words and Phrases Used in Southeast Asia	528
Index of Figures and Tables	569
Index of Technical Terms	576
Postscript	580

报纸、广播电视、网络(新闻)用字用语调查

调查报告

报纸、广播电视、网络(新闻)的用字用语状况,可以反映每年语言生活的实态。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、2006年、2007年的报纸、广播电视、网络(新闻)的用字用语调查报告。本次调查是在2008年的语料上进行的。

一 调查使用的语料及调查内容

(一) 调查使用的语料

本年度的调查语料包括平面媒体、有声媒体、网络媒体三种,共计1 374 157个文本文件,1 209 676 776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991 717 791字次。

2008年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,为了使年度数据的规模具有连续性和可比较性,平面、有声、网络媒体的语料量仍大约按照5:1:4的比例分配,总字符次保持在12亿左右,总汉字次保持在10亿左右。

1. 报纸

平面媒体选择了2008年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。发行量参考了2007年6月4日召开的第60届世界报业大会(南非开普敦)发布的“2007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2007年6月22日由世界品牌实验室(中国北京)发布的“2007年《中国500最具价值品牌》排行榜”。

这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 949 819 个文本,678 842 901 字符次,其中汉字出现 548 899 104 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,广播电视语料选取的主要依据是节目收视率,并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(黄金时段、非黄金时段)、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。

2008 年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、天津电视台、重庆电视台、广东电视台、新华电视台、安徽电视台等八家电视台,中央人民广播电台、北京人民广播电台、天津人民广播电台、山东人民广播电台、深圳人民广播电台、四川人民广播电台、重庆人民广播电台等七家广播电台的 228 个栏目的 19 961 个文本。

广播电视语料量为 120 318 319 字符次,其中汉字出现 98 787 148 字次。

3. 网络(新闻)

网络媒体语料选取了新浪、腾讯两个网站的全部新闻语料,共计 404 377 个文本,410 515 556 字符次,其中汉字出现 344 031 539 字次。

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某些差异。网络(新闻)语料来自新浪、腾讯 2008 年的新闻页面。上述语料均做了去除 HTML 标签信息和广告信息的处理。

(二) 调查内容

本次的调查对象依旧是汉字和词语。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“2008 年、5 月、15 点 30 分”)以及结合紧密使用稳定的短语(如“抗震救灾”、“发展中国家”)。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 2008 年度内的共时考察,也含有 2006—2008 年三个年度间的历时考察。其中,频次、频率、累加频率、出现文本数及使用率的含义

及计算方法与2005年^①、2006年^②的《报纸、广播电视、网络(新闻)用字用词调查》相同。亦可参考《中国语言生活状况报告(2007)》(下编)的“语言资源监测与研究相关术语”。^③

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、无法显示的字符以及仅在“火星文”中出现的书写符号。

(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1 209 676 776字符次。
2. 汉字总数:指全部语料中汉字出现的总字次,计991 717 791字次。
3. 字种数:9 271个。指字形不同的汉字种数。
4. 共用字种数:6 099个。指报纸、广播电视、网络三种媒体中都出现的汉字。
5. 部分共用字种数:1 240个。指只在报纸、广播电视、网络某两种媒体中出现的汉字。
6. 独用字种数:1 932个。指只在报纸、广播电视、网络某一种媒体中出现的汉字。

汉字使用情况的具体数据见表1-1。本次调查的全部语料用字情况形成2008年用字总表(简称用字总表)。

表1-1中的部分共用字种数,以报纸媒体为例,报纸媒体的部分共用字种数

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(2005)》下编,第003—016页,商务印书馆2006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(2006)》下编,第002页,商务印书馆2007年版。

③ 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(2007)》下编,第516—523页,商务印书馆2008年版。

是指报纸媒体与广播电视媒体共用的字种数(这些字种没有出现在网络媒体中)、报纸媒体与网络媒体共用的字种数(这些字种没有出现在广播电视媒体中)两者之和。全部语料的部分共用字种数指三种媒体中任意两个媒体部分共用字种数之和。

表 1-1 2008 年汉字使用情况

媒 体	总字次	字种数	共用字种数	部分共用字种数	独用字种数
报纸	548 899 104	8 156	6 099	1 112	945
广播电视	98 787 148	6 569		311	159
网络(新闻)	344 031 539	7 984		1 057	828
全部语料	991 717 791	9 271		1 240	1 932

(三) 汉字的覆盖率

汉字的覆盖率一方面反映了汉字在整个语料库中的分布情况,另一方面也是衡量汉字是否常用的重要依据之一。其统计结果见表 1-2。

表 1-2 2008 年汉字对语料的覆盖情况

媒 体	覆盖 率	达到 80% 的 字种数	达到 90% 的 字种数	达到 99% 的 字种数	达到 100% 的 字种数
报纸		609	979	2 431	8 156
广播电视		546	898	2 302	6 569
网络(新闻)		587	940	2 270	7 984
全部语料		604	971	2 384	9 271

(四) 与现行规范字表的比较

1. 用字总表前 2500 字与一级常用字比较

用字总表前 2500 字与《现代汉语常用字表》^①一级常用字(2500 字)比较,用字总表中有 344 字是一级常用字中所没有的。将用字总表按照前 500 字、501 至 1500 字、1501 字至 2500 字分为三段,具体情况见表 1-3。

^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。

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	圳媒伊频俄韩措萨诺迪综伦辑曼莱姆澳郭菲洛蒂账炬姚杭署咨聘曝谓凌娜沪颇埃汶鹏莞拟谐邦翔邓粤弗艾冯帕浦贾屏(51个)
1 501—2 500 字	拓卢琳敦穆韦颂枚蔡秦募赫氛斌玲霍肖潘曹聊彭莉羌莎痘胎祛兹蒋玛戈徽廷奈巢履兑虐寞厢堰袁吕颖婴涵契魏耶憾悴憔媿吁滞癌谭逊抑讼鸿硕茨墅崔涯诈郁娟怡挫砸卓旭翰晖篷寓庞坤妮琼柯汰魅鼎峻侯逾瓷仲铭辖铝萧逸鑫蕾妆晰婷磊嘛墟苑湘薇邱镑罕淮彦赋岳楠詹骏逻崩爵鲍杉瞬勃弥奥雯彰衷淑坪廖缉亨赎溢绎邵歧霆淀轴彬幽祭雇坎赁肇侠弘馨坞芯豫邹蓉逛琪凰熙啤冕坠昊胺尹擅薛溃馈菱暨撼撰宠矾佐揽奢尬馐珊飙裸澜玮扳殷瑜瘤轩斡刹擎埔磔勘玫奠菱奎甸陌贬靛呵氲函啸澄碳骚钦咖帷莹禺瘾炫瑰赈啡吻沽惟倪秉隧喻渝蔓皓鹤卿瑟窠潭遂蕊眯凸荔昔媛迄缅瞩湛汕蕴奕腺睹侣倩豹郝崛芭槛啥衍助幢粹禅腕懈仕祈浏焕韶枢裔讶麟毳淇喀噪媳昭玄睿靖寂蟹熬俞晤(292个)

由上表可以看出,2008年所使用的一级常用字范围外的汉字,多与人名、地名或年度大事件密切相关。“圳、伊、俄、韩、澳”等主要表示地名;“萨、诺、迪、伦、曼、莱、姆、郭、菲、洛、蒂、姚、娜”等多表示中国人的姓名或外国人译名,法国的“萨科奇”、奥运会的“菲尔普斯、郭晶晶、姚明、莎拉布莱曼”、金融风暴中的“莱曼兄弟”等对这些汉字的排序靠前有突出贡献;而“汶、赈、婴、炬、氲”等则与年度大事件中的“汶川地震、婴幼儿奶粉、奥运火炬、三聚氰胺”等有很大的关系,所以,高频字也会随着年度的不同而相应地发生一些细微的变化。

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 397 字是《现代汉语常用字表》中所没有的,具体情况见表 1-4。

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	曝汶莞弗(4个)
1 051—2 500 字	韦蔡斌羌莎祛兹耶媿茨娟怡晖妮柯魅鑫婷磊嘛墟苑邱镑彦楠詹鲍雯廖缉亨邵霆肇弘馨邹琪昊胺尹馈暨矾佐尬馐飙玮瑜斡埔奎靛氲帷禺瘾惟倪渝皓窠眯媛瞩湛汕倩郝崛槛禅仕浏韶裔麟淇喀睿俞(83个)

(续表)

2 051—3 500 字	伽龚涅妍晏璐阜弈娅诠厄瑶巛佟琦滕炜潇磋翟霖籓卉 尧朔炳姬黛绯盎侃闫茜昕坍鄂吾阮祁娥哦郝郡洙迭钰 翡茹郴珀颐甄峨菁戛璇牟圭跻汾岚狄禄裴庚荫郅宾拎 汀焯铮曦铂酋鲨芸跋粽襄陇咋萃赣俪禧岑妃倏冈孜岐 颞焱禹冉汧踝缪泸绮汝斐貽辗绚孚羈虞隋道鞘佼馭沁 鏢栩婧钦俨亟鹭瑛嘘袂煽婕钜黠悒豚蓓雍日瀚霾聆黏 嵩佬梓藉森臧嘻丫琛莓憬霄懂姻舜缮漪兮淩愣峙炙荟 葆麒咎骋阪嫦岷祀拽燮任韬姗猝毋鳌辄陀哇抉濮纶匡 晟钊滇熠洗匱萱哪毗漳炖皖骥瑕沂镁忡邸琶桔亢淮簾 矜臻抨躅麓俑煜髦泓余璀恪轶讴榻峥蹊骸荼铍蓟烽嘎 臻睫衷驿铀笨愜釜隼寅烯昱踵峪璞炯宕栾汲恺妞桐嗯 哉珂洱醛岌渲迥淄坂芮郇葵笃哽罹邯魃呱论喆釉悖疵 甬奚暖瑾沓膺槟寰覃矣熹懿婪悸鲤祺骼邬瞿眩靳腱邝 厮辍竺啪戎酶涪癩渭(308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 609 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》之外的字
前 3 000 字	闫(1个)
3 001—5 000 字	森飏喆珺濛玫碁哒婁玳堃罔峻皙瘵鯨琨崧钜仝谿吋鍾 吒岫垚喇槩捱尔翀昇迳市堀後畊斡熾劬玆鏗炊(43个)
5 001—7 000 字	祜礪芑珮瑯涪塚戮甯祐桎铖坤珮膈渠稹櫟槿孑焱崑暎 俟迥內瀋邗鯛德呎椹菱筐佳攸彝邕負欷蹉陸囍姦馭穌 滹說嘍蚵鴉郎埔焯苾虢褚甦缸苾罌禪疋卷颯燦船脫涅 圖嬌暉牠券黃埜啟談辻這廐响罔荳栢鞣湧咲菡壘譽洩 擱蘭駮曝給臆胤口炤煊龔霏紘驩頰醅腦塢別裡璫寔沒 徹陸嫫瑩戍脬說宥甞麼厶鼻屣裹析鉉熿偶迺祗野并幾 瓊費嗜拋鞣胤脩江宦獠蕨樸吁為姪錄砦佢勳璣瑯奧譬 狍嫩吞嘉烟刹茺闌亾濬峇鵠瓊廻聶菸詠準磨颯錫檐袴 樂昉弑沃贊題孃呼髮福嶋噓翊惠靚鴛將鄴時初鏃搖鷗 絮蠟箝氣坤鸚倪肥欄映麟來紘鑣矇窠苧姬誌椒泔吳瀾 賽砵峯告查鸕窻睽尅鰓姪洸國丁長磚症堯涂榭製準邗 增咋左欵媿鎔爭鏐栁岡緋鄺墩裸咁膈鄯禍枇點翠愍忪 銀託讚綸跣裝巳甸天眩強丰糸抹風斐檔采佈塋灯教珪

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。

(续表)

鏹勳鴻穗昨張莊鄭讓葉鷺實卻保鏞宮眈貨柯黑愷磴鰈 馳弧並愛嶺枏孃仝觀顛彥倅恣係嫉餘燊茲週簞資奴怙 畠邢洸錡畧牲癩籀巧妹儼懷曷鬢丟個復汎忝紀鑑灑扞 吟岷崑嘅決靜傑囹眾鯨屋对頤寶燮愨納鏹暇現澎佻晚 維雲與眾種遊炯獠訶咆請裯浦柑滿邁杓馬沫瀟激出 步俾盪島堙還筭枷間進開丐助豐筵壘窩潛堺泐啤皓該 登車搥錄髻彤尙偃企谿侶淳亾萬玗哲害鉞麥霍搵鍊驢 體陝誠篠响駟瓶標峯施呂們勅場場報沂東噉鷓曷掛鶻 箴憂獎蔣輝績璫揭況鸚蓀桂垠祖癩鶴規混級間榷傢戶 佻滴關閉璋幹给啲調鸚傍採寶忙凌盤鮓榮師玃碣癩研 薰糖勢諉陲無舢魚阜奮顛嫫圓喬涓踰齋佔織最鎮漲誅 庄糶稅蘇襪選顛陽閻綱坻柚登(565个)

这 609 个字中包括:不规范简化字(“鍾、迳、讠、钜、榷、鏹”)6 个,繁体字(如“後、蘭、給、張、現、迴、岡、愛”)97 个,异体字(如“犇、荳、峯、減”)81 个,旧字形(如“刹、奧、茲、丟、研”)21 个,日本汉字(如“辻、笹、畑、沢、腦、栃、磨、恵、麼”)20 个,旧计量字(“涇、吋、呎”)3 个。

2008 年,“囧”字使用频次突然上升,不再是表达其“光明”的本义,多数情况下要观其形,悟其义。其字形生动传神地刻画了人的面部表情,或悲伤、或惊讶、或无奈、或尴尬,这些情感体验记录了 2008 年中不平凡的点点滴滴。“囧”字被一些人称为“21 世纪的象形文字”,反映了一种网络文化现象。

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 197 个,具体如下:

鍍破揸鞞鞞琫瓠泚詖馨纒瀟被嶠嶺狹饒韞韶秒甌泝犇
 烜苙劬踟刁憇鞞蝕苾播荆鑽漢稈拔艷筭噴筍膏餈夬鰓
 洎媿蜆胛撤罟苾軻穰墟總嚙姑踏貳壑襟屬愨焦潛泚殪
 僦踞溟倦秦濤昔闕矧鮓棟恨鏘榜鞞蹠颯嶠僂牒杜芈督
 獮釘菁幪鷄醜莫礪鋒肭釁陞扭砮鏘珊榜狂陴棹鮫滯气
 併併鉅鯉拊庖鞞愆蠟鯨纒顛蓐澈氈矧胄鸚駟蝮喉渴闖
 食啖廓峯莢鯢鰒輻閱鷓郅阢依荏豨蕝谿荃汶謝研滯糈
 痲茨枸廢馐繫彥勸虞鷓蟻罍愨鮪纒智蝟咄篋差颯驛衫
 躡寺湮頤辘鸛瘰蠟磁聶痲鰕構

(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、韩国汉字等,出现在用字总表的3 118位以后,即覆盖率达到99.67%之后,较去年的2 871位相对靠后,汉字的规范化使用范围相对更广。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	网络(新闻)	广播电视
繁体字	908	376	615	251
异体字	310	210	209	45
旧印刷字形	65	40	56	8
不规范的简化字	12	6	10	6
旧计量用字	4	4	3	1
韩国汉字	2	0	2	0
日本汉字	78	45	66	10

对于上述字形,我们选取了2008年中出现而在2007年中未出现的部分字,考察了其在报纸语料中的使用情况,具体如下(注:括号内的文字为对应的规范用字):

(1) 使用旧字形、异体字或日本汉字,比如:

“綠茸茸的草甸(绿)、硕士研究生邢璐(邢)、国家稅费(税)、德清县(清)、文娛活动室(娱)、光緒皇帝(绪)、韩中佛教文化交流协会(教)”等使用了旧字形;

“湖南省外事办(事)、今年的情况(况)、庞戈(庞)、过去曾多年实行白澳政策(曾)、行政冗节(冗)、低廉(廉)、廉洁奉公(廉)、还剩七金(剩)”等使用了异体字;

“对韩系车的打压是直接的(压)”使用了日本汉字。

(2) 错用别字,比如:

“触决(解)、柳暗花明(柳)、锯木榧(屑)、人民币,钱币(币,注:“币”是“匝”的异体字)、鸡翅(翅)、女人裹小脚(裹)、很熟悉(很)、好看的T恤(恤)、方案均需国家药监局批准(均)、致富的道路(富)、曝光度(曝)、智障(智)、增添了活力(添)、胆固醇含量减少(胆)”等。

分析造成这些别字的原因,多为用形码输入的过程中,错误输入了相近的形码。对上述这些不规范字、别字的使用,我们逐条对照了纸版本的报纸内容,其使用完全同电子版。由此可见,汉字的使用还需进一步向规范化的方向迈进。

(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

频率比值(简称“频比”)^①能够在一定程度上反映不同媒介、不同领域语言的特点。汉字频比的计算方法同往年,分别计算三种媒体语料的前2500字的频率比值,得到各自的特色用字。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)汉字频比分析

媒 体	前2500字中频率比值在前的20个汉字
报纸	禹荔粤舰圳穗墅践龄莞废铺咨汕衬厨苑岗谐绣
广播电视	咱窠嘛您呢啊扁呀殖它吗彦萌么怎湾你跟候喂
网络(新闻)	憔悴祛痘媾鞞橘浪寞虐羌曝寨翁页烂娱蒂杆腾

从上表可以看出,广播电视语料中的语气词、叹词、称谓词等,反映了有声媒体口语化的特点;报纸语料中的“践、谐、舰”等字由于“实践科学发展观、建设和谐社会、护航舰队”等在报纸语料中出现频繁,“莞、禹、粤、汕、圳”等字则由于报纸语料中选择了广州、深圳等地域性报纸;“浪、腾、页”等表明了网络媒体语料采集来源及形式的特点,“羌、曝、寨、橘”等则由于网络上“羌族文化、曝光、山寨文化、蛆橘”等新闻链接较多。这些汉字在用字总表以及各媒体用字表中的频次、排序情况见附表1。

(七) 2008年、2007年、2006年用字总表的比较

1. 汉字使用情况

表1-8 2008年、2007年、2006年汉字使用情况

媒体	类型 年度	总 字 次			字 种 数		
		2006年	2007年	2008年	2006年	2007年	2008年
报纸		399 488 842	531 287 305	548 899 104	8 326	9 082	8 156
广播电视		53 029 167	86 109 014	98 787 148	6 194	6 539	6 569
网络(新闻)		526 476 397	389 656 861	344 031 539	8 142	8 393	7 984
全部语料		978 994 406	1 007 053 180	991 717 791	9 231	10 123	9 271

表1-8列出了2006年、2007年、2008年三年的汉字使用情况。总字次反映

① 见本书497页术语定义。

了每年调查的语料规模。仅从这三年的数据来看,字种数与语料规模之间基本是正相关的。但也有例外的情形,如报纸语料 2008 年总字次最多,但字种数却最少;网络(新闻)语料 2007 年比 2006 年少,但字种数却多。

通过对报纸 2007 年、2008 年两年的语料用字情况作对比调查,发现只出现在 2007 年的汉字中有较多的粤语方言字,这些字多出现于《广州日报》中的一个方言栏目,2008 年该栏目未出现,这是造成 2008 年字种数相对减少的一个原因。

2. 共用、独用情况

表 1-9 2008 年、2007 年、2006 年用字总表共用独用情况比较

类型 年度	字种数	共用 字种数	独用 字种数	部分共用 字种数	独用字举例(按频次排序前 15 个字)
2008 年	9 271	7 882	602	787	鏈 霖 獠 吳 堯 寧 貫 邗 裸 半 烜 鏈 弧 鮪 憲
2007 年	10 123		1 205	1 036	璿 込 萬 蕘 沈 葦 煖 咏 菝 猱 齡 嘑 楔 翮 屹
2006 年	9 231		656	693	柜 訛 楓 鸞 碼 蝥 筍 犴 犁 襠 焯 碘 淘 輓 璿

表中的部分共用字种数是指只在某两年中出现的字种数量。由上表可知,共用的字占全部字种数的大多数;独用的字大多是低频字,使用具有偶然因素。2008 年独用字中出现频次最高的 15 个汉字见表 1-10,表中列出了具体出现的媒体情况。

表 1-10 2008 年独用字出现情况

汉字	频次	文本数	出现的媒体		
			平面	网络	有声
鏈	292	63	√	√	√
霖	35	11		√	
獠	33	1		√	
吳	21	9		√	
堯	20	4	√	√	
寧	19	8	√	√	
貫	19	6	√	√	
邗	19	3	√	√	
裸	19	2	√		

(续表)

𦍋	17	9	√	√	
烜	17	5	√		
鎡	16	5	√	√	
瓠	14	1		√	
鯧	14	1			√
愆	14	4	√	√	

对于频次较高的独用字,可以从其使用环境分析它们的独用状况,从中看出社会生活对用字频次的影响。经调查,2008年的这些独用字主要用于以下几个方面:一是人名,“愆、瓠、鎡、烜、堯、昊、獯”都用于人名。二是地名,“𦍋”、“邗”是用于两个镇名,其中“𦍋”用于汶川县的一个镇名。三是机构名,“鎡”用于表示台湾鎡震股份有限公司,这家军火公司因其名字的怪异且与台湾政治、军事密切相关而备受大众关注。四是旧词新义,“𦍋”古为“梅”的异体字,本来已经罕用,因其在网络传播中被赋予“形神俱呆”的意思而重新获得生命力。五是引用性的文字,“裸”主要用于《光明日报》的《“裸”之考辨》一文中,着重论述这个汉字的来龙去脉。六是直接使用异体字,“𦍋”本为“𦍋”的异体字,在语料中用于“故事、湖南省外事办”;此外,“贯”主要用在“传贯、赙贯、贯器店”等词语中,其中“传贯”是韩国一种独有的租房模式;“鯧”字的出现是由于新闻报道“黑鯧鱼人工养殖规模逐渐增大”,并且具有较好的经济效益。

3. 汉字使用覆盖率

表 1-11 列出了三年不同覆盖率对应的字种数。从表中可以看出,覆盖率分别达到 80%、90%、99% 的字种数大体会随着语料规模的增加而增加。

表 1-11 2008 年、2007 年、2006 年汉字使用覆盖率比较

覆盖率	年度	语料类型		报纸	广播 电视	网络 (新闻)	全部 语料
		字种	数量				
达到 80% 的 字种	2008 年	数量	609	546	587	604	
		比例(%)	7.47	8.31	7.35	6.51	
	2007 年	数量	603	550	573	595	
		比例(%)	6.64	8.41	6.83	5.88	
	2006 年	数量	591	528	589	591	
		比例(%)	7.10	8.52	7.23	6.40	

(续表)

达到 90% 的字种	2008 年	数量	979	898	940	971
		比例(%)	12.00	13.67	11.77	10.47
	2007 年	数量	976	920	928	964
		比例(%)	10.75	14.07	11.06	9.52
	2006 年	数量	955	904	954	958
		比例(%)	11.47	14.59	11.72	10.38
达到 99% 的字种	2008 年	数量	2 431	2 302	2 270	2 384
		比例(%)	29.80	35.04	28.43	25.71
	2007 年	数量	2 439	2 389	2 281	2 394
		比例(%)	26.86	36.53	27.18	23.65
	2006 年	数量	2 401	2 379	2 340	2 377
		比例(%)	28.84	38.41	28.74	25.75

各个年度覆盖率达到 80%、90%、99% 的字种(为叙述方便,统称高频字)占年度总字种数的比例见图 1-1。从图中可以看出,高频字占字种数的比例通常是随着总字种数的增加而减小的。^①这说明当语料规模达到一定的值之后,高频字的分布比较稳定,高频字的字种数会随着语料规模的增加而少量增加,但它所占字种的比例会减小;语料中所增加的字种数多为偶然使用,对高频字的影响不大,但对高频字所占比例会有影响,规模越大,高频字所占字种的比例越小。

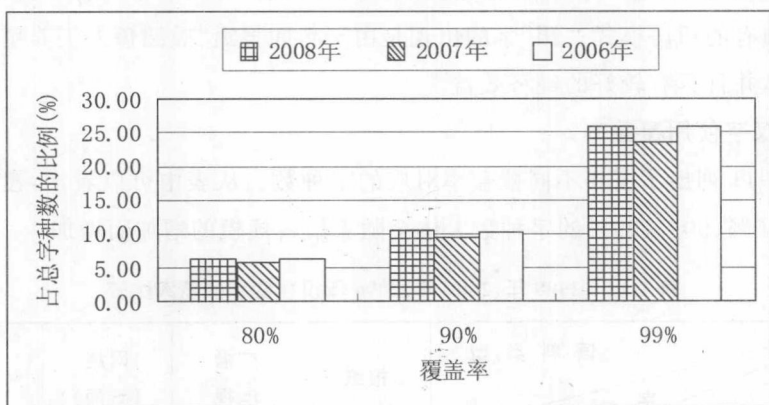


图 1-1 三年不同覆盖率字种数比例

4. 高频字

高频字在年度之间的差别可以很好地反映每年社会用字的变化,并进一步反映年度语言生活的特色。这里选择各年度用字总表的前 600(覆盖率在 80%

^① 2008 年、2007 年、2006 年的总字种数分别为 9271、10123、9231。