



《义务教育常用词表》的研制及义序法的运用

苏新春



《义务教育常用词表》的研制及 语义法的运用

厦门大学嘉庚学院

厦门大学国家语言资源监测与研究教育教材中心

苏新春

香港城市大学

2018-3-19



一 研制背景及内容

- (一) 研制背景
- 《义务教育语文课程标准》对小学和初中的语文教学中词汇的学习和使用有明确要求。现行各种教材也设有名称不同而功能相同的“识字组词”“字词积累”“读一读，写一写”“词语比较”等词汇学习栏目，但对词汇学习的质和量却一直没有明确的要求与范围。
- 不同教材之间的词汇面貌和词汇量差异非常大，在我们所统计的教材中共有词只有40%^①，生词的共用部分比例更低。
- 长期以来中小学语文教学界对学习性词表表示出了很高的期盼，从叶圣陶、吕叔湘、张志公等前辈学者到一线教师，都表达了对学习性词表研制的愿望。



-
- 国家语委于2011年批准设立“基础教育学习性词表的研制”（YB125-29）的课题。
 - 于2016年8月批准设立“国家语委语言文字科研项目优秀成果后期资助”课题——“《基础教育学习性词表》的分级、验证及推广”（HQ135-1）。
 - 由厦门大学国家语言资源监测与研究教育教材中心承担，苏新春为课题负责人。



中国语言绿皮书 A 系列

《义务教育常用词表（草案）》

教育部语言文字信息管理司 组编

苏新春 主编

《义务教育常用词表》课题组

2017 年 3 月

- 教育部语言文字信息管理司组编
- 苏新春主编
- 商务印书馆出版，2018年6月
- “中国语言绿皮书” A卷之一



- **（三）词表内容**

- 《义务教育常用词表（草案）》共包括词14323条。共分四级：
 - 一级词1807条；
 - 二级词4997条；
 - 三级词4779条；
 - 四级词2740条。
- 分别对应于小学1-2年级的第一学段，3-4年级的第二学段，5-6年级的第三学段，初中1-3年级的第四学段。



音序表

(所有词条按音序排列, 后面带有音、级、词性、义类)

1	阿姨	āyí	1 名	壹一 Bb	31	爱人	àirén	2 名	壹一 Dc
				名	壹一 Bb			名	壹一 Dc
2	啊	ā	1 语气	玖五 Bb	32	爱屋及乌			
3	哎	āi	2 语气	玖五 Bb		ài wū - jí wū	3		伍五 Ga
4	哀兵必胜				33	爱惜	àixī	1 动	伍五 Ka
	āi bīng - bì shèng		4	伍八 Ca	34	爱心	àixīn	1 名	叁三 Bb
5	哀悼	āidào	4 动	伍五 Xb	35	隘口	àikǒu	4 名	肆二 Dc
6	哀告	āigào	4 动	陆九 Jb	36	碍手碍脚			
7	哀鸣	āimíng	4 动	伍六 Ab		àishǒu - àijiǎo	3		捌四 Fd
8	哀求	āiqiú	3 动	陆九 Jb	37	安	ān	2 动	陆三 Ab
9	哀伤	āishāng	3 动	伍五 Ba	38	安定	āndìng	3 形	捌六 Aa
10	哀思	āisī	4 名	叁三 Bc	39	安顿	āndùn	4 动	陆一 Fb
11	哀叹	āitàn	4 动	伍五 Bg	40	安放	ānfàng	3 动	伍一 Ea

义序表

(所有词语按义类排列, 词语后面括号中的数字是词的分级)



壹 生物

一、人

A 泛称

a 人

人(1)、人口(1)、人类(2)、
人士(2)、人物(2)
人员(2)、劳动力(3)、人手
(3)
个人(2)
家伙(2)
家(1)

他们(1)、她们(1)
别人(1)、人家(1)、某(2)、
旁人(2)
哪个(1)、谁(1)

e 前人古人

鼻祖(4)
创始人(3)
古人(1)、先辈(2)、前人(3)、
先驱(4)

f 今人后人

世人(2)
后代(2)、接班人(2)、传人
(3)、后人(3)

B 性别

d 成年人

大人(1)

e 老人

老人(1)
爷爷(1)、公公(2)、老人家
(2)、老爷爷(2)
老奶奶(1)、老太太(1)、奶
奶(1)

D 亲属

a 亲戚眷属辈次

亲人(1)、亲戚(2)、亲属(2)
至亲(3)、本家(4)
远亲(2)
家人(1)、家属(2)、家眷(4)



-
- 音序表中“义类编号”显示的是词的语义分类。如“散”字：
 - 当义类是“性质与状态/性状/紧-松/松/松散”，义类编号是“捌三Db02”时，拼音是上声，词性是形容词；
 - 当义类是“运动与变化/事态变化/结束/终了/散场”，义类编号是“柒三Bb02”时，拼音是去声，词性是动词。
 - 参照的义类系统为《现代汉语分类词典》（商务印书馆，2013年）。该义类系统收录现代汉语通用词目83000余条，按五级语义层划分。有一级类9类，二级类62类，三级类508类，四级类2057类，五级类12659类。



二 词表的性质与功能

- (一) 词表性质

1. 是学习性词表而非通用性词表。

学习性词表反映了对学习内容与教学标准的要求，要在一定条件下通过一定教学手段以达到预定的教学目的。收词及内部的分级分等要符合词表使用者的认知需求和认知特点。根据学习等级的要求来划分词级是学习性词表最重要的内容之一。

通用性词表，反映的是社会普通成员在一般言语交际活动中使用的通用性词语。它一般是按使用状况来排列，如频率、分布率或综合二者而成的使用度。



2.是面向母语学习者的基础教育词表。

仅就“学习性”来看，《义务教育常用词表（草案）》与《对外汉语词汇大纲》《汉语国际教育用音节汉字词汇等级划分》是相同的。

但从学习者的身份、特点及学习目的来看，二者有着巨大差异。

“对外汉语”“汉语国际教育”面向的是第二语言学习者，他们已经具备了一定的母语能力，而基础教育面向的是正在进行母语学习的学习者。

前者已经完成了第一语言和“第一认知世界”的塑形，这时需要完成的是第二语言能力和“第二认知世界”的学习，功能目标明确；

后者要完成的是第一语言和“第一认知世界”的塑形，语言学习与语言能力、逻辑能力、对世界的认知是紧密联系在一起。



• (二) 词表功能

• 1. 书面语词汇系统的学习功能。

- 表现在词汇上，要学习的就应是书语词汇而非口语词汇，是承载着思想观念、历史现实、政治经济、文化社会等的通用性规范性词汇，而非“捉蛴蛴玩家家”、“拿筲箕淘米，拿火钳夹炭”的生活俚语俗词。
- 书面通用词汇的学习主要是由语文课程来完成的。历史、地理、数学、物理、化学等学科都有与语文词很不一样的学科高频词，对它们的学习主要是由所在学科来完成的。



- 2.对母语社会的认知功能。
- “文以载道”通常指的是文学作品，其实义务教育阶段的语言学习也有同样的功能，这里的“道”指的是语义及背后的文化世界。母语学习者学习语言的过程就是认知世界的过程。他们通过语言学习来认识整个世界。他们生存于其中的自然环境、社会环境和文化环境，都是通过语言来承载和传递的。“锄禾日当午，汗滴禾下土。谁知盘中餐，粒粒皆辛苦。”这首诗传递的绝不止于农民辛苦耕种，要爱惜粮食不浪费等表层语义，在诗句的背后其实还传递着更多的底层文化信息。



- 一个人的文化身份、文化认知、文化认同正是在长期的语言浸润中才获得的。义务教育阶段需要学习的正是能体现中小学生学习需求与认知特点的词，能反映中小学生学习于斯长于斯的生活、学习的环境，有利于进一步掌握语言文字知识的必需储备。这与仅仅把汉语作为交际工具的“对外汉语教学”的二语学习有很大的不同。
- 如对外汉语教材中的“麻婆豆腐”位于中国饮食文化类词的首位，这是经特意挑选而获得了极为显著的文化代表义；在义务教育语文教材中却无一例，因为中小学生学习从小由“豆”到“豆浆”“豆腐”，再到“麻婆豆腐”，是一种自然生活积累所得。



-
- 同样是“语言文化的学习”，二语学习者是“定向”“任务”式，母语学习者是“熟视无睹”“潜移默化”“润物无声”式。
 - 如中国的传统二十四节气，在中国传统文化的生产、生活、知识，甚至文学创作中都有重要位置，词表将它们收进来，就是在词的背后建立起了完整的节气文化语境，而这在其他性质的词表中则是不必要的。



-
- 3.体现词汇的习得规律。
 - 词汇的学习有“知”“晓”“用”三个层次。
 - “知”为知道，是听过见过这个词，但还不能准确理解它的意义，一般不会用到它。
 - “晓”是懂得、理解，它比“知”进了一层，不仅知道还能较为准确地说出它的意思，有了特定语境或许会使用到。
 - “用”是运用，即不仅知晓理解，还会在个人的话语与写作中熟练自如地使用。
 - “知”“晓”“用”三个层次就是“知道”“懂得”“使用”的差别。人们在认知一个语义类的词时，往往存在着这样的差异。



- 如“害怕/惧怕”类，本词表从一至四级的收词分别是：
- “怕、害怕”；“生怕、谈虎色变”；“恐惧、大惊失色、恐怖、唯恐、闻风丧胆、提心吊胆”；“战战兢兢、畏惧、望而却步、毛骨悚然、惧怕、噤若寒蝉、魂飞魄散、魂不附体、战栗、胆怯、望而生畏、不寒而栗”。
- 语义由浅至深，语体由口语至书语，表义由指称至描绘。
- 该语义类没收入词表的词还有不少，如“惧怯、惟恐、畏惮、畏怯、畏难、可怖、疑惧、骇惧、骇怕、怵头、忌惮、狼顾、惶惑、失容、戒惧、心悸、震悚、打怵、犯怵、发怵、犯慄、悚然、丧魂落魄、魄散魂飞、惊魂未定、心胆俱裂、视为畏途”等，这些词可归入“知”的范围甚至更远。



-
- 本词表大体按“晓”的程度来设定词汇的收录范围。
 - “晓”的程度要难于“知”。一个人对词汇“知”的范围很宽，将它们都纳入词表会加大词表的容量和学习难度。
 - “晓”的难度是低于“用”，能为词汇教学与词汇测试提供更大一些的范围，为学习积累、提高词汇“用”的能力做好储备。



三 词表的规模与收词标准

- (一) 词表规模
- 本词表在确定收词规模时主要考虑了以下几方面的因素。
- 1. 义务教育阶段语文学习任务所应掌握的语义表达范畴。这里使用了语义表达范畴而没用词汇量的说法，因为语义域是语言表达交际的实在内容，一个语义域可以用很多词来表达；词汇量是语义域的表达形式，表达同一语义域的词形稍有不同，就可以成为不同的词。



-
- 2. 现有语文教材及长期教学经验中总结出的词汇教学量。
 - 在四套样本教材中，各教材的词总数在2.3万~2.9万条之间。共有词是10460条，占每套教材词汇总数的36.83%~45.12%之间。调查的教材愈多，共有词愈少，主要原因就是各教材并没有一个共同的选词标准。这种状况在港台的语文教材及民国语文教材中也差不多。



- 3.教材词汇的常用频率。
- 覆盖四套样本教材总语料的
 - 80%用了2457词，频次是35次，临界的五词是“恐惧、山林、神话、用心、准确”；
 - 90%用了7078词，频次是10次，临界的五词是“细腻、增多、转来转去、下场、眼帘”；
 - 95%用了14405词，频次是4次，临界的五词是“盐碱、幸得、印刷、贮蓄、杂事”；
 - 96%用了20584词，频次只有3次，临界的五词是“田田、水泵、尚书郎、柔婉、吴承恩”。
- 根据对累加覆盖率95%的用词考察，“较为常用”的“通用”词也是大致在1.5万左右。



• (二) 收词标准

- 本词表所收应是普通话中的通用词，它们应具有通用性、常用性、基础性、语文性、规范性等特点。具体的收词原则如下：

1. 收录普通话的通用词。

- 普通话词具有通用、普遍、稳定的特点，不具备这些特点的词不予收录。如收了“爸爸”“父亲”，不收带方言色彩的“阿爸”“爹爹”“老子”；收了“手腕”，不收“腕子”“手腕子”；收了“拇指”，不收“大拇哥”“拇哥”“大指”；收了“富裕”“富庶”，不收“肥实”“裕如”“活络”“厚实”；



- 2.主要收语文词，一般不收专名。
- 语文词指的是能满足于一般交际需求的词，它们具有很好的通用性、常用性、普遍性。
- 如普通话词汇系统中的“虫”类有252个词（见《现代汉语分类词典》），本词表只收了“虫、昆虫、甲虫、害虫、知了、蟋蟀、萤火虫、蚯蚓、蜈蚣、蜜蜂、春蚕、蝴蝶、蜻蜓、蜘蛛、蚊子、苍蝇、蚂蚁、蚂蚱”等18个。
- 又如在一级词中收录了少数表示我们国家、政权、政党象征或代表的名词，如“中国”“五星红旗”“共产党”“解放军”“北京”“天安门”等，但重点还是在收录“国家”“国旗”“国歌”“首都”类通用词上。



-
- 3.注重收录词的原形，不收重叠、无表意作用的儿化词、变换语素等变形词。
 - 如收“摇头”“点头”“看”，不收“摇摇头”“点点头”“看看”；
 - 收“船”，不收“船儿”；
 - 收“胸有成竹”“八仙过海”“别出心裁”，不收“成竹在胸”“八仙漂海”“独出心裁”。



-
- 4.一般不收组合叠加词。
 - 组合叠加词如：“口渴” “心想” “长大” “白兔” “浇水” “你好” “瓶口” “葡萄架” “清香袅袅” “色彩明丽” “受用不尽” “思潮起伏” “物产丰富” “消磨时光” “雪花飞舞”，及“小熊” “水桶” “小刀”等。



-
- 5.适当收入当代产生、稳定性强、已进入普通话的词。
 - 如“硬件” “软件” “数码” “摄像” “签证” “配置” “联手” “客户” “汇率” “驾照” “互动” “股东” “高铁” “峰会” “多媒体” “盗版” “研发” “网址” “网吧” “发言人”。



- 6.收录能独立使用、作单音词用的汉字。
- 由于汉字与汉语词的独特关系，还会遇到单字词与复音词的问题。如“爸”与“爸爸”、“妈”与“妈妈”、“筷”与“筷子”、“鞋”与“鞋子”、“窗”与“窗户”、“厕”与“厕所”，二者之间的风格庄谐、语义广狭、语气轻重等皆有所不同，但基本意义相同。
- 在来自真实语料的描写性词表中，二者并收是正常的。但学习词表为保证在有限的规模中能使每个词都能起到专门的语义表达作用，更好地与字表保持互补作用，收录更接近当今语言使用习惯、符合书面语特点的双音词。



-
- 7.适当收录有较强表现力、较高稳定性、较广使用范围的成语。
 - 成语是汉语词汇中富于表现力的成分，是中小学生学习词汇的重点。对若干含有3500常用字以外的字、常用度高的成语如“韬光养晦”“吹毛救疵”“神采奕奕”“东施效颦”“虎视眈眈”等，也适当收录，以体现“以词促字”的学习原则。
 - 词表没有收录“语”中的“谚语”“歇后语”，因这类语的数量大，独立性强，地域性和文化性鲜明，意义内涵和语用功能丰富灵活，对语境依赖性高。



-
- 8.有多种词形的，只收推荐词形。
 - 收录《第一批异形词整理表(草案)》以及《264组异形词整理表(草案)》等国家试行规范标准及行业规范标准中已作出判断的推荐词。
 - 如收“吃里爬外”“一塌糊涂”“缥缈”“规诫”“黏稠”。
 - 不收“吃里扒外”“一蹋糊涂”“飘渺”“规戒”“粘稠”。
 - 上述规范没有规定的，则参考权威辞书选择推荐词形。如：
 - 只收“淳厚”“做证”“银铛”“反照”“啰唆”“执着”；
 - 不收“醇厚”“作证”“郎当”“返照”“啰嗦”“执著”。



四 词表的分级

- 本词表共分四级，努力在不同的词级中体现易难、浅深、先后的差别。划分词级时主要依据了以下几点。
 1. 根据词汇的认知规律。词汇认知在广度上有“知” “晓” “用”范围之别，以基本词为核心，呈现逐层扩大的趋势。
 2. 根据长期以来教学与使用教材的经验。（采用位序法）
 3. 根据频率高低的排列顺序。常用度高，词义明白浅近，贴近日常生活和学习实际的词容易理解，这样的词就排在初级。
 4. 参照《义务教育语文课程常用字表》，努力做到词汇学习与汉字学习相衔接。



- 下面是在多套教材中用位序法得到的前100个词：

- “我们、座、要、对、说、着、间、住、家、红、鸟、的、摘、成、梅花、鱼、好看、祖国、星星、头、落、望、山、朵、说话、放、跳、快、找、种、就、船、滴、春风、前、秋天、三、旁边、漂亮、当、蓝天、世界、树、圆、一定、紫、还、床、盖、树木、黑、哪里、瞧、画、这么、跟、爬、颜色、多少、满、许多、能、还是、有趣、只、吗、高兴、一、雪白、阳光、写、有、岁、喝、回来、流、数、了、四、人、收、外、东、下面、嘴、西、给、大家、走、有的、早晨、老、冬天、更、爷爷、玩、送、蓝、办法、啊”。

- 称其为位序法，是着眼调查时依据的是词的出现形式；
- 称其为经验法，是因为位序法体现出来的是经验性的做法；
- 称其为认知法，是显示这种形式、经验所反映出来的是中小学生的词语认知习惯。



-
- 《语文课程标准》规定，小学阶段要掌握一级常用字2500个，初中阶段学习二级常用字1000个。词表收录了能独立作词用的汉字，即为单音词。
 - 词表根据汉字与词汇的繁衍关系，在处理词表与字表关系时，遵循了“以字带词”“词不越字”“以词促字”的原则。
- ### 1.收录能独立成词的字。
- 从《义务教育语文课程常用字表》3500常用字中选用了1543个有完整表意、能独立使用的字进词表。还收入了若干有特殊功能的汉字。如“甲”“乙”与“子”“丑”等22个天干地支字，因为它们不但经常用作序数词，而且在中国传统文化中还有很强的记岁记年记时记史的人文价值。



2. “以字带词”的原则。

《语文课程标准》列有“300基本字”，指出“本表是识字、写字教学的基本字表。这些字构形简单，重现率高，其中的大多数能成为其他字的结构成分。……这些字应作为第一学段教科书识字、写字教学的重要内容”^⑮。这些字所表示的意义也都是基本义，且大都能独立使用。所以它们也应包括在一级词中。300基本字在一级词表中作为单音词的有246个，另外54个汉字则出现在复合词中。在初稿中一级词无论是单用还是复合词中，都没有出现“军”“业”“舌”“卫”四个字。为了与“300基本字”相对接，将含有四字的“解放军”“军人”“冠军”、“作业”“毕业”“学业”、“舌头”、“卫生”“卫星”等安排在了一级词中。



3. “词不越字”的原则。

对要求初中才学的二级常用字，尽量不出现在小学阶段的一、二、三级词中。如“蠢”“吮”“骚”“秉”“怠”“嘲”属二级字，“蠢事”“蠢笨”“吮吸”“秉性”“秉烛夜游”“怠工”“怠慢”“嘲笑”“嘲讽”“嘲弄”就没出现在小学阶段的词表。含有3500常用字表以外字的词一般也不予收录，如“不落窠臼”中的“窠”，“春意盎然”中的“盎”，“集腋成裘”中的“裘”，“涓涓细流”中的“涓”，“出类拔萃”中有“萃”，“遐迩闻名”中的“遐”。这并不是说这些词不好，而是为了尽量保持与字表的对接，形成有效的知识序列，更好地保持字词知识教学的层次关系。



4. “以词促字”的原则。

这一原则是对“词不越字”原则的灵活运用。

- “袜、馒、禾、饺、蕉、膏、狸、筷、旷、箏、诵、柿、姥、澡、蜻、厕”。如严格参照汉字分级的要求，这些字都应该在第四学段才出现，就会导致“袜子”“馒头”“禾苗”“饺子”“香蕉”“牙膏”“狐狸”“筷子”“旷课”“风筝”“背诵”“西红柿”“姥姥”“洗澡”“蜻蜓”“厕所”这些非常常用的词要到初中才出现，这显然不合适。
- “旷”是二级字，“旷野”在四级词，但“旷课”可提前出现；
- “诵”属二级字，“吟诵”在四级词，“诵读”可提前出现；
- “澡”是二级字，“洗澡”也可提前出现。



词表研制中尤为重视语义规律 语义观主要表现在以下几方面

1. 中小学生的母语学习是认知心智逐步成熟、语言使用逐步精细化的过程。
2. 中小学生的母语学习是对本民族文化的习得过程。
3. 学习性词表的研制中，语义法比频率法、分布法具有更强的针对性与科学性。
 - A 语义类覆盖率比词语覆盖率更有说服力；
 - B 位序法的本质是语义认知法；
 - C 词级划分的本质是在同一义类中追求表义的细腻与丰赡；



语义法

- 语义法主要指的是语义分布上的广-窄、难-易、深-浅、粗-细、熟悉-陌生的差异上。
- 如亲属关系，父母、兄弟姐妹、祖孙的关系就是基本的语义关系，而随着少子化，堂与表、姑表与姨表的关系就逐渐陌生了。
- 而语义的差异很大程度上表现在表意的“精细”上，即我们通常说的同义词的掌握上，同义词掌握得越多，语义表达愈精细，语义掌握的难度也就愈难。而这些现在都变得可以测量的了。
- 本词表的语义调查借助的语义分类系统是《现代汉语分类词典》。词表收词量不大，但具有较广的语义分布面。表现出的结果是愈是上层语义，覆盖愈广的特点。



-
- 语义观直接影响到了词表的性质、功能、规模及分级。
 - 语义问题超越了“纯工具”的范畴，因此，它也就不是单纯由词频能决定的了。
 - 语义问题本质上会受到人的认知规律的影响。好的基础教育常用词表必须立足于儿童的认知需求、认知能力、认知需求。