

# 小学汉字形声字表音特点 及其分布的研究

舒 华 武宁宁 郑先隽 周晓林

**提要** 本研究考察了小学生所学汉字的表音特点和分布规律,从小学课本中汉字信息的统计特点和学习方式探讨儿童掌握形声字表音规则的可能性和必要性。认为语言材料数据库的研究是了解人的语言获得和语言加工过程的一条很好的途径。

一个学生在小学期间将学习大约三千汉字,需学习的汉字几乎占一般成年人掌握汉字的二分之一,儿童通过学习也会了解汉字本身的规律和特点,为以后的学习和提高打下基础。所以,在小学阶段的汉字学习是非常重要的,教师所教授的不应该仅仅是单个孤立的汉字,还应该引导学生认识字与字之间联系、特点,使他们从中获得规律性的知识,有助于他们继续学习。

儿童在学前已发展了大量的口头词汇,在头脑中建立了大量语音与语义的联系,在识字初期,儿童面临的主要问题和困难是要建立大量字形与字音之间的联系,把大量已掌握的口头词汇转化为书面词汇,尽快学会阅读,克服阅读困难。由于汉字是非拼音文字,没有明显的形—音对应关系,许多字形与字音的关系是任意的,这就增加了学习的困难。然而,语言学家的的工作已表明,占现代汉字 80% 以上的是形声字,形声字带声旁,声旁可以在一定程度上提供字的语音信息,尽管声旁提供的语音信息不是完全可靠的。小学学习期间儿童是否有可能利用形声字的表音规律? 认识形声字的表音规律是否能帮助儿童更好地学习汉字? 这一直是教育学家和心理学家所关心的问题。

近年来越来越多的心理语言学家的的工作表明,人的语言加工特点会受到语言文字使用度的影响,也就是说,语言及其使用的许多规则性会反映在人的语言加工中,影响语言的学习过程。英语儿童词汇发展的研究表明,语言和文字系统的学习是不可能全部靠简单记忆或强化作用的,由于学习的内容远超出人能记忆的范围,学习中必定包含某种一般规则的抽取,而语言规则的学习不仅与儿童认知因素有关,还很大程度上受语言文字的规则性特点的影响(Derwing & Baker, 1979; Nagy & Anderson, 1984)。因此,要回答表音规律在汉字学习中的作用问题,来自两个方面的研究证据是重要的,一是关于汉字表音规则性特点的研究,二是关于儿童汉字认知和学习规律的研究。

关于形声字声旁表音规律,语言学家已进行了长期的探讨。早期研究认为形声字的表音度是较低的,周有光(1978)的研究表明完全规则的形声字大约只占汉字的 30%。近年来一些基于计算机统计的工作对形声字表音特点做了更细致的分析,Yin, W., (1991)的研究表明,约有 36% 的声旁能提供可靠的整字读音信息, 48% 的声旁提供整字读音的部分信息, 16% 的声旁完全不提供整字的读音信息。李燕、康加深(1993)对声旁与整字读音的声、韵、调关系的研究表

明,声旁与整字读音的声、韵、调全相同的情况占 38%,声旁与整字读音的声、韵相同,调不同的占 18%,声旁与整字读音的韵相同,声、调不同的占 16%。Perfetti & Zhang(1992)的研究表明,形声字的表音规则性还与字的频率有关,即规则性随着字频的降低而增强。

虽然语言学家、心理学家对汉字形声字表音程度的统计还存在不同的看法,认知实验研究已经表明,儿童汉字学习包含着一般规则的学习(李燕萍,1995;舒华,1997;Shu & Anderson,1997)。例如,研究表明,一、二年级学生还不能较自觉地使用汉字的表音线索,儿童在中年级后可逐步了解。掌握形声字的声旁表音规则,这种认识帮助他们记忆学过的汉字,也帮助他们学习和推理未学过的字。儿童掌握声旁表音规则存在很大的个体差异,语文能力高的儿童对声旁表音规律的认识较早、较好,而语文能力低的儿童认识较晚、较差,这表明声旁表音规律的认识与儿童语文能力的发展存在较高相关(舒华,曾红梅,1996;Yang & Peng,1997;Shu, Wu & Zhou,1997)。

要探讨儿童对汉字形声字表音规则的认识,应该对小学汉字表音的统计学特点有所了解,到目前为止,还缺乏该方面的细致的研究。本研究的目的主要就是考察小学生所学汉字的表音特点和分布规律,从小学课本中汉字信息的统计特点和学习方式看儿童掌握形声字表音规则的可能性和必要性,为心理学和教育学研究提供参考。

## 一 小学汉字基础信息库的建立

我们选用北京市小学统一使用的六年制语文课本建立了小学汉字基础信息库,该库收集了学生从语文课本中学习的生字,共 3262 个字。为了分析汉字的各种特征及儿童学习情况,对库中的汉字做了多种编码。编码的信息包括:

### 1. 声旁代码:

参照语言学家的分类(倪海曙,1982;周有光,1980),我们给每个声旁一个独立的代码,共给 1426 个声旁编码。

### 2. 声旁类型:

参照语言学家的工作及现代汉语字典,我们将所有汉字按声旁类型分为八类,其中凡是涉及可能的多音、繁简问题,在分类上有困难的字,我们都将其归入特定的类别:

一类:规则形声字——声旁读音与整字读音的声母相同、韵母相同、声调相同;

二类:规则形声字——声旁读音与整字读音的声母相同、韵母相同、声调不同;

三类:不规则形声字——声旁读音与整字读音的声母不同、韵母相同;

四类:不规则形声字——声旁读音与整字读音的声母相同、韵母不同;

五类:不规则形声字——声旁读音与整字读音的声母不同、韵母不同;

六类:非形声字;

七类:其它形声字——多音形声字:声旁或整字是多音字。

八类:其它形声字——涉及古字、简化字等。

### 3. 声旁的位置:

对库中所有的汉字,按其声旁在字中的位置共做了七种编码:在右、在左、在上、在下、在内、在外、其它。

### 4. 声旁是否独立:

声旁不独立的字指汉字的声旁不独立成字,而只能出现在合体字中,如“沈”;声旁独立的

字指其声旁本身也是一个字,有自己的读音,如“清”。

### 5. 声旁的位置是否固定:

声旁位置固定的字指其声旁只能出现在汉字的某一位置,如形声字“消”的声旁“肖”只能出现在字的右边;声旁位置不固定的字指其声旁可能在汉字的不同位置出现,如“妈”的声旁“马”,可能出现在字的右边(如“妈”),或在字的下边(如“骂”)。

### 6. 声旁是否可以作为形旁:

两种编码分别是:其声旁只可能做声旁的字,如“销”的声旁“肖”在统计范围内只能做声旁;其声旁既可做声旁又可能做形旁的,如“码”的声旁“马”可在“骑”字中做形旁。

### 7. 汉字作为生字出现的课本册数:

将所有的汉字按其作为生字第一次出现的课本册数进行了编码,从一年级到六年级共做了12个编码。

### 8. 汉字的频率:

选用国家标准一级字频度表(摘自《汉字信息字典》,科学出版社,1988)统计字频。

## 二 小学形声字表音特点及分布分析

对小学汉字基础库中形声字表音特点及分布进行统计分析,结果如下:

### 1. 声旁表音规则性及其与字频、年级的关系

首先,对小学各年级所学生字的数量及其占小学总汉字量的比率及平均频率进行统计。从表1中可以看出,一至四年级识字量较大,尤其是二、三年级的识字任务较重。随年级的升高,课本中所学生字的平均频率急剧下降。这主要与中年级后语文学学习重点转向阅读理解,阅读课文中引入的低频字迅速增加有关。

表1 小学各年级的识字量及平均频率

	一年级	二年级	三年级	四年级	五年级	六年级
字数	526	853	622	511	376	337
比率	.17	.26	.19	.16	.12	.10
平均频率*	1.00	.28	.18	.08	.04	.02

\*注:平均频率的单位为千分之一,因此表中的1.00表示百万次出现一千次,.02表示百万次出现二十次。本文的其他表中平均频率的含义相同。

按使用频率将汉字分成高频字、次高频字、中频字和低频字四个等级,其中高频字指每百万次出现一百次以上的字,次高频字指每百万次出现十次到一百次之间的字;中频字指每百万次出现一次到十次之间的字;低频字指每百万次出现一次以下的字。统计各年级所学的不同频率段的汉字占识字总量的比率,结果表明,低年级课本中以高频、次高频字为主,随着年级的升高,在识字总量中高频字数量减少,低频字数量增多。高年级所学的字中,次高频字和中频字占主要部分,低频字有较大幅度的增加(见表2)。

按基础库中的声旁类型,将小学汉字分为四种:规则形声字(一、二类)、不规则形声字(三、四、五类),其他形声字(七、八类)、非形声字(六类),并统计这四种字在各频率段、各年级的出现比率。在以上四种字中,非形声字不提供字的读音线索,规则形

表2 不同频率的字在各年级识字量中的比率

	一年级	二年级	三年级	四年级	五年级	六年级
高频字	.76	.47	.34	.20	.10	.05
次高频字	.20	.41	.44	.44	.37	.30
中频字	.03	.11	.18	.27	.40	.40
低频字	.01	.01	.04	.09	.13	.25

按基础库中的声旁类型,将小学汉字分为四种:规则形声字(一、二类)、不规则形声字(三、四、五类),其他形声字(七、八类)、非形声字(六类),并统计这四种字在各频率段、各年级的出现比率。在以上四种字中,非形声字不提供字的读音线索,规则形

声字中声旁提供可靠的读音线索,不规则形声字提供部分或不提供读音线索,其他形声字提供读音线索的情况比较复杂,本研究中未作进一步分析。

表3 不同类型的字在各频率段中的比率

	高频字	次高频字	中频字	低频字
非形声字	.41	.20	.08	.03
规则形声字	.21	.33	.42	.58
不规则形声字	.23	.35	.33	.27
其他形声字	.15	.12	.17	.12

表4 不同类型的字在各年级识字量中的比率

	一年级	二年级	三年级	四年级	五年级	六年级
非形声字	.45	.29	.23	.15	.12	.13
规则形声字	.18	.31	.33	.38	.38	.41
不规则形声字	.22	.27	.33	.32	.35	.30
其他形声字	.15	.13	.11	.15	.15	.16

从表3、表4可以看到,随着汉字频率的降低,汉字总量中非形声字的数量稳定减少,规则形声字的数量稳定增加,而不规则形声字和其他形声字在各种频率上的分布变化不大。从儿童学习汉字的年级分布看,一年级课本中学习的主要是非形声字,占一年级总识字量的45%,随着年级的升高,各类字的年级分布呈与频率分布同样的变化趋势。

统计分析表明,汉字形声字有随着频率的降低增强表音规律的趋势,可以预计,随着年级的增高,儿童学习越来越多的汉字后有可能了解汉字的表音规律;一年级儿童学习的主要是非形声字,还不足以认识汉字的表音规律,然而,一年级的学习是非常重要的。统计表明,一年级的非形声字中,80%的字将成为后来学习的汉字形声字的声旁,当中高年级儿童进入大量阅读阶段后,他们在阅读中遇到的生字主要是中低频字,利用声旁信息有利于学习、记忆和推理生字的读音,因此,帮助儿童了解和掌握声旁表音规则是不可忽视的。

## 2. 声旁与形声字的关系

声旁对形声字读音最直接的影响是声旁本身的读音。我们进一步对声旁及其构成的整字的读音做了分析。我们将两类规则形声字(一、二类)和三类不规则形声字(三、四、五类)作为标准形声字,并将标准形声字编码为声旁独立成字的和声旁不独立成字的,前者指字的声旁本身也是汉字,有自己的读音,后者指声旁本身不能独立成字,只能作为整字的部件,没有读音或其读音很少为人所知。从对标准形声字中声旁独立成字与声旁不独立成字的比率的分析结果中可以看出,绝大多数字的声旁本身也是独立的汉字,有其明确的读音(见表5)。

表5 标准形声字中声旁是否独立的分析

	声旁独立成字	声旁不独立成字
字数	1850	161
比率	.92	.08

我们还对小学阶段学习的汉字中能做声旁的字与其构成的整字的相对频率进行了分析,结果表明,总体说来,能做声旁的字的频率高于整字的频率,声旁字的平均频率是约每百万次七百三十次,而声旁字构成的整字的平均频率约为每百万次十五次(见表6)。

也就是说,当儿童遇到一个声旁独立的熟悉字时,其声旁的熟悉度一般大于整字的熟悉度;而当儿童遇到一个声旁独立的生字时,其声旁在很大的概率上已经是儿童熟悉的。从这个角度来看,声旁的读音有可能影响整字的读音。

从小学语文课本中引入生字的方式来看,低年级儿童学习大量的非形声字,分别占一年级和二年级识字总量的45%和29%,而在这些儿童早期学习的非形声字中,有70%以上的非形声字成为儿童后来学习的形声字的声旁(见下页表7)。数据分析中

表6 声旁独立成字的形声字中声旁与整字的频率比较

	声旁	声旁构成的整字
平均频率	.73	.15

可以预测,中年级以上的小学儿童的汉字读音会强烈地受到声旁的影响。

表7 一、二年级学习的非形声字中  
是否可做声旁的分析

	可做声旁	不可做声旁
字数	367	122
比率	.75	.25

### 3. 声旁特征的复杂性

我们进一步对声旁特征的复杂性做了分析。首先,把声旁位置分为七种,分别统计了声旁在各种位置上的字数。然后,根据声旁是否可以在字的不同位置上出现,统计了声旁位置固定的字数,表8、表9中给出了分析的结果。结果表明,在标准形声字中,

大部分字的声旁占据汉字的右半边,比率高达65%;在其余的字中,声旁在下半部、声旁在上半部、声旁在左边和声旁在其他位置的字各占四分之一左右。然而,绝大多数形声字的声旁在汉字中位置是不固定的(超过百分之八十),声旁位置固定的字,即声旁只能在字的左边、右边或其他位置之一的字是很少的。

我们还统计了标准形声字中,字的声旁同时可以在其他字中作为形旁的字数。从表10看出,

表8 标准形声字中声旁位置的分析

	声旁在右	声旁在左	声旁在下	声旁在上	其他(在内、在外等)
字数	1229	140	200	168	207
比率	.65	.07	.10	.08	.10

出,标准形声字中,大部分字带的声旁在汉字组成中只能做声旁,但仍有约13%的字中带的声旁在汉字组成中也可以作为字的形旁。

表9 标准形声字中声旁位置稳定性分析

	声旁位置不固定	声旁位置固定
字数	1665	348
比率	.83	.17

表10 标准形声字中声旁是否可作形旁的分析

	声旁不可作形旁	声旁可作形旁
字数	1757	256
比率	.87	.13

从以上分析中可以看出形声字中声旁的特点,如声旁位置的复杂性、声旁与部首可能的混淆等都会增加整字中声旁识别的困难。对小学儿童来说,正确识别汉字的表音部分——声旁并不是一件容易的事情。

## 三 小 结

对小学语文课本中3262个汉字的表音特点及其分布的统计分析表明:小学二、三年级是识字高峰期,这一阶段学生接触的生字量最大。随着年级的升高,小学生所学的中、低频字比例上升,高频字比例下降。低年级儿童学习的高频汉字中,非形声字、不规则字占很大的比率,它表明低年级儿童认识形声字表音规律的可能性很小。随着年级的升高,学生学习生字的频率下降,非形声字比例下降,规则形声字比例上升,其分布比率与前人研究的结果大体相似。统计分析揭示了两个相互关联的事实:(一)高年级儿童在独立学习中遇到的主要是低频字;(二)汉字的表音规律性随着字频的降低而稳定增强。这两个事实告诉我们:小学高年级学生在学习中有可能了解形声字的规则性,而了解形声字的规则性、掌握汉字表音线索对于他们有效地学习汉字、扩大字汇量有重要作用。

儿童在学习了解形声字表音规则性的可能过程如何呢?统计分析表明,一方面,绝大多数形声字的声旁本身也是汉字,一般来说,能做声旁的字的频率高于其构成的整字的频率,小

学低年级学习的70%以上的非形声字成为儿童后来学习的形声字的声旁。所以,声旁与整字的频率关系、各类汉字学习的先后顺序,都有利于儿童发现声旁的表音作用。中年级后,儿童的汉字读音将会较强地受到声旁的影响。另一方面,在标准形声字中,形声字的声旁可能出现在多种位置上,大部分字的声旁位置是不固定的,还有少部分字的声旁既可作声旁又可以作形旁。这些分析表明,声旁表音规则的掌握不是一件容易的事,首先,声旁位置的不固定、声旁与形旁可能的混淆等给形声字中声旁的识别带来困难;另外,形声字中大量不规则字的存在,尤其是高频字的不规则性,给儿童了解形声字的表音规则,掌握汉字表音线索,确定生字读音都带来很大的困难。从小学汉字库统计分析所得的这些预测与我们从儿童汉字读音实验中得到的数据结果(舒华,曾红梅,1996;舒华,武宁宁,1996;Shu, Wu & Zhou, 1997)非常相近。

本研究结果表明,语言材料数据库的研究是我们了解人的语言获得和语言加工过程的一条很好的途径。

## 参考文献

- [1]倪海曙《现代汉字形声字字汇(初稿)》,语文出版社,1982。
  - [2]周有光《现代汉字中声旁的表音功能问题》,《中国语文》1978年第3期。
  - [3]周有光《汉字声旁读音便查》,吉林人民出版社,1980。
  - [4]李燕、康加深《现代汉语形声字声符研究》,收入陈原主编《现代汉语用字信息分析》,上海教育出版社,1993。
  - [5]李公宜、刘如水主编《汉字信息字典》,科学出版社,1988。
  - [6]四省市六年制《小学语文课本》(一至十二册),浙江教育出版社,1984。
  - [7]北京市九年义务教育六年制《小学语文课本》(一至八册),北京教育出版社,1995。
  - [8]李燕平《正字法规则信息在汉字识别中的作用》,北京师范大学硕士论文。
  - [9]舒华、曾红梅《儿童对汉字结构中语音线索的意识及其发展》,《心理学报》1996年第2期。
  - [10]舒华《汉语儿童词汇和阅读获得的研究》,收入彭聃龄等主编《汉语认知研究》,山东教育出版社,1997。
  - [11]舒华、武宁宁《儿童汉字读音的规则性和一致性效应》(未发表手稿,1996)。
  - [12]Derwing, B. & Baker, W., Recent research on the acquisition of English morphology. In P. Fletcher & M. Garman (Eds.), *Language Acquisition*, New York: Cambridge University Press, 1979.
  - [13]Shu, H. & Anderson, R. C., The Chinese writing system and the development of reading ability. In Wang, Inhoff & Chen (eds), *Reading Chinese Script: A Cognitive Analysis*. Mahwah, NJ: Lawrence Erlbaum, 1998.
  - [14]Shu, H. & Anderson, R. C., Role of radical awareness in the character and word acquisition of Chinese children, *Reading Research Quarterly*, Vol. 32, 78-89, 1997.
  - [15]Yang, H & Peng, D. L., The learning and naming of Chinese characters of elementary school children, The paper presented in the 7<sup>th</sup> International Conference on the cognitive processing of Chinese and other Asian language, Hong Kong, 1995.
  - [16]Shu, H, Wu, N. & Zhou, X., Reading compound characters: A developmental study. The paper presented in the 3<sup>rd</sup> International Conference on cognitive processing of Chinese language, Beijing, 1997.
  - [17]Yin, W., On reading Chinese characters-A experimental and neuropsychological study. Unpublished doctoral dissertation, University of London, 1991.
- (舒华 武宁宁 郑先隽 周晓林 北京师范大学心理学系, 邮编:100875)