

学者新论：中文信息技术标准：汉字注音？拼音正词法？

教育部语言文字应用研究所研究员、博士生导师 冯志伟

人民网 2003年4月16日

代前言

冯志伟教授是教育部语言文字应用研究所研究员、博士生导师。2002年10月中旬，在河南省开封市召开了“中国语文现代化学会第五次学术会议”。这篇文章，是根据冯教授的报告摘编的。

冯志伟教授的文章，提出了语音、历史和识别的三个正词法原则结合的方法来施行和完善现代汉语拼音正词法，从而使全国文化教育工作和中文信息技术发展有更严格规范的和统一的语言文字根据。冯教授的主张，对国内和国际的规范标准的实行与建设都很有现实意义。

现代汉语拼音正词法是1988年中国政府颁布的，是具有法律意义的国家标准规范，而且，也是联合国等国际组织使用的国际标准规范。这个标准规范的主要内容，是中文书写中的“词”和“词界”的确立。

1989年，“汉字优越”论和“汉字文化”论开始泛滥，鼓吹者的目标之一，是否认中文有“词”和“词界”、进而否定现代汉语拼音正词法，恢复半个世纪以前逐渐淘汰的“汉字注音”法。在报刊媒体的鼓动宣传和某些部门人士的支持下，“汉字优越”论终于成了气候，现代汉语拼音正词法和相关的规范标准的建设受到极大干扰和破坏，而被淘汰的汉字注音法大有回笼之势。

中文信息技术不光是输入输出汉字，更还有句子和段落的处理、还有排序等数据结构管理和中文算法语言等等，这些，都需要明确的词界。为了发展中文文献管理和翻译自动化，中国已经投入了三十年左右的时间、上亿元的资金和数不胜数的人力，可是，许多基本问题依然处于一筹莫展或停滞不前的状况，究其技术原因，主要问题就是词界模糊、单个汉字连成一片。

以前，中文书写没有标点符号，需要很大的“句读”工夫才能避免歧义和正确阅读。“句读”，通俗说，就是分句和确定“句界”。引进拼音文字的标点符号方法，句界清晰，为避免歧义和正确阅读提供了极大的方便，而中文还是中文，并没有因为使用拼音文字的标点符号而不再是中文了。试想，如果取消标点符号，所有的中文文献都是一个个汉字连成一片，那么，现在的人工阅读和计算机文献处理，会是个什么样子？多数人会同意，那将是一场灾难性的技术和文化的倒退。

同样，如果没有现代汉语拼音正词法确定词界的方法，那么，许多中文文献工作、中文数据管理自动化和翻译自动化等工作，就会面临巨大的困难。跟使用标点符号一样，确定词界不是取消汉字、也不是用英文等西方文字代替中文书写，而是在使用标点符号的基础上，进一步避免歧义和保证正确阅读，为中文数据管理和自动

翻译等提供基本的文字方法依据，使计算机操作、人工操作、说话和书写等等，有统一的标准规范。

汉字注音是一个个汉字连成一片的方法；现代汉语拼音正词法是确定词界的方法。恢复汉字注音、还是坚持现代汉语拼音正词法？这不但是全国基础教育和文献管理的大事，而且是中文信息技术标准依据的大事，也是中国政府提交给联合国的国家标准和国际标准是否有效的大事。

吕叔湘、王均和周有光等老前辈科学家，一开始就看到了“汉字优越”论的危害，认为那是主张文化倒退。他们主张继续执行和完善现代汉语拼音正词法，一步步搞好中文现代化建设。然而，他们坚持科学真理的声音被一浪高过一浪的媒体炒作淹没了，直到前几年网络热掩盖了“汉字优越”论的热浪。网络热冷却后，“汉字优越”论又开始抬头，跟尊儒复古思潮结合，继续试图否定现代汉语拼音正词法和恢复汉字注音。令人注意的是，随着“汉字优越”论和“汉字文化”论的泛滥，已经有人试图把文字方式作为搞文化分裂和国家分裂的政治工具。

主持国际通用字符库标准的联码（Unicode）组织，每年都召开国际学术会议，讨论各种语言文字的数据管理文字方式和字符标准规范等问题。随着中文信息技术市场的发展，近些年来，联码会议也越来越多地提出了中文文字方式的“词界”问题。中文书写的“词界”和现代汉语拼音正词法，已经超越国家行政和文化地理的范围，成为国际科研学术界十分关注的一个重要问题。

在这个时候，看到冯志伟教授的《英德法语的正词法与汉语拼音正词法》，深受启发，特别向人民网做了推荐。承蒙人民网编辑部的要求，特做此介绍，且当作前言，奉献给读者。（米阿仑）

中文信息技术标准：汉字注音？拼音正词法？

《中华人民共和国国家通用语言文字法》第一章第十八条规定：“国家通用语言文字以《汉语拼音方案》作为拼写和注音的工具。《汉语拼音方案》是中国人名、地名和中文文献罗马字母拼写法的统一规范，并用于汉字不便使用或不能使用的领域。初等教育应当进行汉语拼音教学。”

《汉语拼音方案》不仅是注音的工具，而且也是拼音的工具，要拼音，就必然会涉及到词儿连写的问题。在《汉语拼音方案》中，采用 y、w 和隔音符号“'”来隔音，就是为了解决词儿连写的问题，而词儿连写的问题，就是汉语拼音正词法的问题。我们在这里讨论汉语拼音正词法，是完全符合《中华人民共和国国家通用语言文字法》第一章第十八条的规定的。最近一段时间，似乎特别强调《汉语拼音方案》给汉字注音的功能，而非常忽视甚至试图削弱《汉语拼音方案》给汉语拼音的功能。因此，我们认为，有必要在这里借鉴国外正词法的研究，来进一步加强《汉语拼音正词法》的研究和改进。

问题的提出

“正词法”在英语里叫 Orthography。ortho-是“正确”的意思，-graphy 是拼写的意思，整个 Orthography 就是“正确的拼写法”的意思。1958年2月11日，国家正式公布了《汉语拼音方案》，用拉丁字母来拼写汉语，这就出现了汉语拼音的正词法问题。

汉字一连串写下来，既不分词又不连写，只有“字”的书写单位，没有“词”的书写单位。汉语普通话中的词有单音节的，有多音节的，多数是双音节的。用汉语拼音来拼写普通话时，应该分词连写，也就是以词作为书写单位，使得语言中的表义单位能够在书面形式上反映出来。这是拼音书写形式与汉字书写形式的根本区别。

比如，“中华人民共和国”不是写成 zhong hua ren min gong he guo，也不是写成 zhonghuarenmingongheguo，而是写成 Zhonghua Renmin Gongheguo。这样的写法反映了词语的实际结构，阅读起来比全分或全连都来得清楚。可见，汉语拼音正词法就是以词为单位来正确地书写汉语普通话的拼写法。

任何拼音文字都应该有自己的正词法。英语、法语、德语等采用拉丁字母作为拼音文字的语言，在长期的历史过程中，发展出适合于本语言特点的正词法。汉语拼音虽然不是拼音文字，但是，在它的实际使用中，特别是在与语言信息处理有关的应用中，也有正词法的问题。为了深入地研究汉语拼音正词法问题，有必要对英语、法语和德语的正词法发展情况进行研究，以便作为我们的借鉴。

英语的正词法

英语分古代英语（公元700年-1100年）、中古英语（公元1100年-1500年）和近代英语（公元1500年至今）。公元9世纪丹麦入侵英国，至11世纪初，丹麦实际上统治了英国，丹麦语和英语同属日耳曼语族，加速了古英语的简化过程。

1066年法国入侵，法语成为官方语言，法语属罗曼语族，与英语不是同一个语族，英语吸收了大量的法语词汇，语法上也受到了法语的影响。在法国入侵之后，英国实际上存在着3种语言：法语是官方语言，拉丁语是宗教语言，英语是社会下层使用的语言。

后来由于政治上的历史变动，英语逐渐成为官方语言，学校开始教英语，英语成为了英国的全民共同语。在英语上升为官方语言的过程中，吸收了大量的法语借词，公元1250年-1400年的150年内，约有10000个法语词汇进入了英语，75%至今还在使用。

由于基督教会的影响，拉丁语词汇也大量进入英语，使得英语的同义词大量增加。例如，同是表示“火”的三个词，fire 来自英语，flame 来自法语，conflagration 来自拉丁语；同是表示“问”的三个词，ask 来自古英语，inquire 来自法语，interrogation 来自拉丁语。

文艺复兴时期对古希腊、罗马文化的研究，大量的拉丁语、希腊语词汇涌入英语。例如，arithmetic（算术），grammar（语法），logic（逻辑）来自希腊语，arbitrator（仲裁人），executer（执行者），item（条款）来自拉丁语。

这些情况，使得在现代英语中，借词占了 80% 的比重，这些来自不同语言的大量借词，减少了同形词和同音词产生的机会。大量吸收外来词 -- 这是英语正词法得以形成和巩固的一个重要因素。

英语正词法中比较严重的问题是如何解决英语的书面拼写形式和读音之间的矛盾问题。

由于文艺复兴时期印刷术的推广和教育的普及，英语的书面拼写形式逐渐统一，形成了规范的形式，而这些规范的形式又通过印刷的方式进一步固定下来。

与此同时，英语的语音发生了很大的变化。例如，gnash（咬牙）和 gnat（小昆虫）中，g 不发音，knight（骑士）和 know（知道）中，k 不发音，而在中古英语中，这几个字母全都要发音。又如，辅音字母前的 r（如 arm）和词尾的 r（如 father）不发音，词尾的 e（如 live）不发音，而在中古英语中，它们都是要发音的。

这种情况，使得规范的书面形式与实际发音之间出现了很大的矛盾。在英语正词法中，基本上保持历史上原来的词形，从而减少了同形词和同音词产生的可能性。

18 世纪开始，英国强调语言的规范化，要求语言准确有力。1755 年，约翰逊 (Samuel Johnson) 编写了第一部英语词典，把英语词的拼写形式固定了下来。这样，英语的正词法就能够以词典作为规范的根据。

尊重历史，保留词形的原有形式，这是英语正词法得以形成的另一个重要因素。

书面拼写形式与实际发音不一致，当然也会给英语的学习和使用带来一些困难，为了统一读音，英国从维多利亚时代就开始推行标准语，著名语音学家琼斯 (Daniel Johns) 编写的《英语发音词典》 (English Pronunciation Dictionary) 成了人们必须遵守的发音规范。

我们可以看到，英语的正词法主要是遵从历史原则，同时也适当考虑语音原则。

法语的正词法

法语属罗曼语族，它的祖先是拉丁语。拉丁语的使用早在罗马帝国时代就一分为二了：书面的拉丁语古文有严格的语法规则，专为贵族和僧侣使用，口头拉丁语俗体是人民大众使用的活语言，又称民间拉丁语。民间拉丁语随着罗马帝国版图的扩大，由军队和商人传到了法国南部的高卢人那里，并逐渐取代了高卢语，同时又受了高卢语的影响。公元 7-8 世纪，原来的民间拉丁语变成了一种新的语言 -- 这就是古代法语。

12 世纪末，以巴黎为中心的方言逐渐成为了法语的全民语言，17 世纪以来，由于发展商品经济的需要，法国一直十分注意语言文字的标准化和规范化，主张清除法语中的不纯洁成分。

在长期的历史发展过程中，法语的语音有了很大的变化，而书面的拼音形式却已经基本固定下来，语音的演变并没有反映在书面的形式上，因此，法语与英语一样，也存在着书面的拼写形式与实际读音之间的巨大差异。这一方面固然给学习法语带来了一定的困难，但另一方面却为区分同形词和同音词提供有力的手段。

为了解决书面拼写法和实际读音的矛盾，法语的正词法提出了许多行之有效的规则。

例如，法语词末的辅音 **-ds, -s, -x** 等，大多数是古音在书面形式上的遗迹，在现代的口语中已不发音。利用这些形式不同而实际上已经不发音的词末辅音，就可以区分同形词和同音词。**poids**（重量），**pois**（豌豆），**poix**（松香）3 个词，发音都是 [pwa]，是同音词，由于词末辅音 **-ds, -s, -x** 不发音，利用这些形式各不相同的词末辅音，就可以把这 3 个同音词在书面形式上区别开来。

可见，法语正词法主要是遵从历史原则，并把这个原则作为区别同音词的重要手段。

德语的正词法

德国人在使用拉丁字母之前，是使用鲁纳字母（**Rune**）的。这种字母的形式很像拉丁字母和希腊字母。德国人在接受了基督教之后，就放弃了鲁纳字母而改用拉丁字母。公元 2 世纪，拉丁字母的地位在德国进一步巩固。到了公元 7 世纪，拉丁字母成了西欧各国大部分民族文字的基础。但是，德国使用的拉丁字母是歌德体的字母，与西欧各国并不完全一样。在使用拉丁字母来拼写德语的过程中，德语的正词法逐渐形成。

1596 年，语言学家魏歇勒（**Wecherer**）建议，名词的第一个字母大写，这种办法有利于辨识名词，大大地方便了阅读，但也由此而引起了关于字母的大写和小写的一系列正词法问题的争论。

1876 年，德国召开了正词法会议，会上建议取消字母组合 **th, Thal** 写为 **Tal**（山谷），并研究了 [s] 和长元音的书写规则。

1901 年，德国召开了第二次正词法会议，通过了第一次正词法会议所提出的各项建议，并采用了著名语言学家格里木（**Grimm**）的建议，放弃了歌德字体，提倡使用西欧各国普遍使用的拉丁字母。

著名词典编纂专家杜登（**Konrad Duden, 1829-1911**）编写了德语正词法词典，为德语正词法的研究立下了功劳。

德语正词法主要根据下列原则：

1. 语音原则：这个原则要求词中字母的发音应与该字母的表音一致。例如，p, t, k, f 等字母在词中的任何位置或与任何字母相结合，均依次读为 [p], [t], [k], [f]。这个正词法原则还要求单词的发音与其拼写法尽量保持一致。例如，kalt (冷) 读为 [kalt]，oft (经常) 读为 [oft] 等等。在德语中，70%-80% 的词和字母是符合语音原则的。由于坚持了语音原则，使得德语的字母和单词易学易记，给学习和使用带来了极大的方便。

2. 形态原则：这个原则要求词根和词干的写法，在由它们派生出来的词中保持不变。也就是说，在合成词中，在名词的变格形式中，在动词的变位形式中，在形容词的比较级和最高级形式中，尽量保持词根和词干的原有形式，必要时采用变音 (Umlaut)。

3. 识别原则：这个原则要求把声音相同而意义不同的几个词在写法上加以区别，以便在书面形式上区分同音词。例如，Lid (眼皮) 与 Lied (歌曲) 都读为 [lid]，根据识别原则，把 Lied 中的 [i] 写为 'ie'；malen (画图) 和 mahlen (磨碎) 都读为 [maln]，根据识别原则，把 mahlen 中的 [a] 写为 'ah'。

4. 历史原则：这个原则要求保持历史上的某些传统写法。例如，ziehen (拉) 一词中的字母 'h' 是没有什么作用的，因为词根里 'i' 的长音已经写为 'ie'，完全没有必要在加上一个 'h' 来表示长音，但由于这个 'h' 在历史上曾经表示过一种特殊的而现在已经不复存在的辅音，根据历史原则，可以把这个 'h' 保留在 ziehen 的词形中。

在这四条原则中，“语音原则”是德语正词法的最基本的原则，这个原则使得德语的书面形式比英语和法语易读，易写，易记。

然而，德语有 19 个元音，23 个辅音，共 42 个音位 (phoneme)，而拉丁字母只有 26 个，这就不可能处处都遵循语音原则。为了弥补字母之不足，同一个字母可以表示不同的音位，同一个音位可以表示不同的字母，也可以采用字母组合来表示一些单字母不能表示的音位。为了补充语音原则的不足，德语正词法还规定了如下的字母拼写规则：

1. 同一个音位可以用不同的字母或字母组合来表示。
2. 同一个字母可以表示两个音位的组合。
3. 用不同的字母组合方式表示长元音，并使之成为区别同音词的手段。表示长元音的手段主要有：①在元音后加哑音 h 来表示长元音；②用重叠元音字母的方式来表示长元音。

采用不同的方法，一些同音词就可以区别开来了。例如，

Aale (鳊鱼) -- Ahle (锥子)

Moor (沼泽地) -- Mohr (黑种人)

由此可以看到，德语正词法十分注意语音原则，但也考虑到了形态原则，识别原则和历史原则，以区别同音词，减少拼音文字使用中的同形现象。

英语、法语和德语的正词法都考虑到了历史原则，德语正词法还考虑到了识别原则。这些语言的正词法原则可以给汉语拼音的正词法提供宝贵的经验。

中国文字的拼音化运动

在明朝末年，欧洲一些国家的传教士来中国传教，开始用拉丁字母给汉字注音，学习汉语和汉字，并出版了一些汉语译音读物，专供欧洲传教士之用。如1605年意大利耶稣会教士利玛窦 (Matteo Ricci, 1552 - 1610) 在北京出版的《西字奇迹》，1626年法国耶稣会士金尼阁 (Nicolas Trigault) 在杭州出版的《西儒耳目资》。著名语言学家罗常培先生说，《西儒耳目资》是“明末耶稣会在中国音韵学上的第一个贡献。”

鸦片战争之后，各国传教士蜂涌来华，他们用拉丁字母给圣经注音，拼写中国的方言。1867年，英国人威妥玛 (T. F. Wade) 出版了《语言自述集》，用26个拉丁字母来拼写北京官话，形成了有名的威妥玛方案。

这些都是用拉丁字母来拼写汉字读音的最早尝试。这样的尝试激发了知识分子对中国文字拼音化的兴趣，在清朝末年，开始了切音字运动。

在切音字运动中提出的拼音字母是多种多样、琳琅满目的，可以归纳为三大系：

① 假名系：模仿日文假名，采用汉字部首作为拼音符号。例如，1892年卢懋章的《一目了然初阶》一书中提出的“中国切音新字”，1901年王照的“官话合声字母”等都属于假名系。

② 速记系：采用速记符号作为拼音符号。例如，1896年到1897年两年间出版的蔡锡勇的《传音快字》、沈学的《盛世元音》、王炳耀的《拼音字谱》等书中提出的方案都属于速记系。

③ 拉丁系：采用拉丁字母作为拼音符号。例如，1906年朱文熊的《江苏新字母》、1908年刘孟扬的《中国音标字母》和江亢虎的《通字》、1909年黄虚白的《拉丁文臆解》等书中提出的方案都属于拉丁系。

1911年辛亥革命成功，1913年在北京成立了“读音统一会”，通过了注音字母方案，这些字母选自古汉字，具有民族特色，1918年由教育部公布实行，并在全国推广，由于语文工作者的热心倡导，政府的积极支持，注音字母顺利地普及到各级学校，成为字典、词典的注音工具，在帮助学习汉字，普及拼音知识等方面起了很

好的作用。这就是民国初年的注音字母运动。

在清末的切音字运动中，已经开始考虑到拼音正词法问题。卢戢章在他的《一目了然初阶》一书中，已经用短横连接音节的方法实行了初步的分词连写。蔡锡勇的《传音快字》一书中，提出了“连书”的概念。沈学的《盛世元音》一书中，还把“连书”与词类问题结合起来讨论，主张按词类“缮写连书”。王照的《官话合声字母》、朱文熊的《江苏新字母》、刘孟扬的《中国音标字书》等书中，都用不同的方式表示了以词为单位的书写方法。

但是，学者们对于以词为书写单位，还停留在初步的感性认识阶段，还没有总结出一套较为系统的分词连写规则，更谈不上建立正词法的理论。而且，基本上没有考虑历史原则。

在民国初年的注音字母运动中，由于这个运动的倡导者只是仅仅用注音字母来给汉字注音，因而对于分词连写的问题几乎没有涉及。所以，我们可以说，注音字母运动对于拼音正词法是没有什么贡献的。

1917年陈独秀在《新青年》上提出了“文学革命”的口号。1918年钱玄同在《新青年》上发表了《中国今后的文字问题》一文，促进了人们对拼音化问题关注。1923年，《国语月刊》出版了特刊《汉字改革号》，发表了钱玄同的《汉字革命》、赵元任的《国语罗马字的研究》、黎锦熙的《汉字革命军前进的一条大路》等论文，开始了国语罗马字运动。1928年9月26日，由政府正式公布了国语罗马字方案，作为注音字母的第二式。这个方案用拉丁字母来拼写汉语，用字母来表示声调，考虑的比较周密，已经接近拼音文字。但是，没有很好地得到推广。

在国语罗马字运动中，黎锦熙明确地提出了“词类连书”的问题，并且拟订了“复音词类构成表”，提出了正词法的初步规则。1928年黎锦熙的《国语模范课本》和1929年赵元任的《最后五分钟》，通过拼音读物，对国语罗马字作了较为系统的试验。肖迪忱于1934年、孙先六于1936年曾经拟订了国语罗马字的分词连写条例，但都不够完整，不够系统，不够成熟，而且也完全不考虑历史原则，很少考虑识别原则。

30年代初期兴起了拉丁化新文字运动。这个运动一开始，就以集体名义制定了《中国汉字拉丁化的原则和规则》，于1931年9月在苏联海参威召开的中国新文字代表大会上通过。1934年到1937年间，上海、北京、天津等城市成立了新文字研究团体，出版了拉丁化新文字的书籍和杂志。1938年10月成立了陕甘宁边区新文字协会。1941年吴玉章发表了《中国拉丁化新文字的写法规则》一文，对拉丁化新文字的规则作了系统的总结。

在拉丁化新文字运动中，正词法问题受到了普遍的重视，许多学者对此作了深入的研究。林汉达发表过不少关于分词连写的文章，出版过第一本拼音文字的词表《国语拼音词汇》，倪海曙的《中国拉丁化新文字的写法》一文，提出了词的拼写法规则68条，对拉丁化新文字的正词法作了详尽而系统的总结。可惜，这些总结中，并没有注意到历史原则和识别原则，导致同音词的分化很差，成为而后汉语拼音正

词法的一个隐患。

汉语拼音正词法是否应当考虑历史原则和识别原则？

1958年2月1日，中国第一届全国人民代表大会第五次会议通过了《汉语拼音方案》，内容包括：

- ①字母表：规定字母的形式、名称和顺序。
- ②声母表和韵母表：规定汉语中声母和韵母用字母来表示的方法。
- ③声调符号：规定调号形式和标调方法。
- ④隔音方法：规定隔音符号和隔音字母 y, w, yu 用法。

显而易见，①②③三项是关于拼写音节的规则，而④项已经超出了单个音节的范围，为拼写多音节的单词创造了条件。

这个方案在全国推广，在语文教学，词典编纂，文献检索，信息处理等许多领域中得到了广泛的应用。

1970年，联合国的地名标准化委员会在雅典举行的专家组会议上，决定采用《汉语拼音方案》作为拼写中国汉语地名的国际标准。1982年国际标准化组织(ISO)的第46技术委员会(TC 46)投票通过，决定采用《汉语拼音方案》作为用拉丁字母拼写汉语的国际标准，该国际标准的编号为 ISO-7098. 这样一来，《汉语拼音方案》不仅是中国的国家标准，而且还是国际标准，是世界各国应该遵守的国际规范。

陆志韦出版了《汉语的构词法》一书，为汉语拼音正词法的研究提供了相当详尽的资料。周有光的《汉字改革概论》一书，对汉语拼音正词法的理论和方法作了全面的论述。在《汉语拼音方案》公布之后出版的《现代汉语词典》、《汉英词典》、《汉语拼音词汇》以及大量的拼音读物，也都采用了分词连写的办法，这些都为汉语拼音正词法的制定积累了多方面的经验。

1982年正式成立了汉语拼音正词法委员会。该委员会提出，汉语拼音正词法应该以现代汉语语法中定义的词作为正词法的拼写单位，同时也要照顾到阅读和理解的方便，并且要照顾到某些国际上拼写的习惯。该委员会同时也开始研制汉语拼音正词法的基本规则，参考过去个人研制的和集体草拟的正词法规则，《汉语拼音方案》公布以来出版的各种拼音读物，各种以《汉语拼音方案》为基础并以词为拼写单位的词书，各种信息处理用的汉语拼音分词连写的实践经验，广泛听取了各方面人士的意见，并请教育界、出版界、信息处理界和语文界的专家学者进行了讨论，又与有关单位协作，进行了10万多词的拼写试验。在这些工作的基础上，于1984年10月公布了《汉语拼音正词法基本规则（试用稿）》。这个试用稿经过几年的实践，又作了适当的修改，于1988年7月制定了《汉语拼音正词法基本规则》，由国家教育委员会和国家语言文字工作委员会联合公布施行。

但是，我国的汉语拼音正词法，仍然基本上没有考虑历史原则和识别原则，只是单纯地考虑语音原则，致使同音词难以区分，同形词或词形易混词相当多，造成阅读上的不便，在汉字的计算机拼音输入法中，同音词的选择增加了用户的麻烦，降低了输入的效率，这成为拼音输入法的一个弊病。我国的拼音正词法中，是不是也应当适当的考虑历史原则和识别原则呢？这个问题希望大家讨论。

在信息网络时代，中文的信息处理成为一个重要的研究领域，也是语文现代化的一个方面。在中文信息处理中，汉语书面文本的自动切分是一个非常困难的问题。从原则上说，汉语拼音中的词应该与用汉字书写的汉语书面文本中切分出来的词保持一致。因此，汉语拼音正词法的深入研究，对于中文信息处理具有重要的作用。我们应该借鉴国外正词法研究的宝贵经验，推进我们汉语拼音正词法的研究工作。

注：本文根据“中国语文现代化学会第五次学术会议”（2002年10月11-14日，河南开封）的报告摘编。报告作者冯志伟教授是教育部语言文字应用研究所研究员、博士生导师。

来源：人民网 2003年4月16日

（责任编辑：董盟君）