国际标准ISO 7098: 2015的四个特色

冯志伟[1,2]
(1.教育部语言文字应用研究所, 北京 100010; 2.杭州师范大学, 杭州 311121)

摘要: 2015年12月15日, 国际标准ISO 7098: 2015正式在日内瓦出版。本文对该国际标准的四个特色进行分析。

关键词: ISO 7098: 2015; 国际标准
中图分类号: G250.7
DOI: 10.3772/j.issn.1673-2286.2016.12.007

1 引言

1958年2月11日, 全国人民代表大会一致通过《汉语拼音方案》作为拼写汉语普通话的国家标准[1]。汉语拼音成为中国初等教育的教学内容, 每位学生都应学习和掌握《汉语拼音方案》。通过汉语拼音给汉字注音, 提高汉字学习效率, 帮助学生进一步学习文化和科学技术。汉语拼音在电报拼音化、视觉通信、文献编目、排序检索、人力资源管理中得到很好的应用, 在中国受到普遍的欢迎[2]。

1979年, 中国代表周有光在华沙召开的ISO/TC 46(国际标准化组织第46技术委员会)第十八届会议上, 建议将《汉语拼音方案》作为国际标准。

1982年, 在南京召开的ISO/TC 46第十九届会议上, 正式通过ISO 7098《文献工作——中文罗马字母拼写法》(ISO 7098 *Information and Documentation: Chinese Romanization*)国际标准。澳大利亚、加拿大、法国、日本、韩国、德国等投赞成票, 英国弃权, 由于技术原因, 美国投反对票。因大多数国家支持标准通过, 从此汉语拼音从中国的国家规范成为国际标准。

1991年, 在巴黎召开ISO/TC 46第二十四届会议上, 对原ISO 7098进行技术修改后, 颁布ISO 7098《信息与文献——中文罗马字母拼写法(1991)》, 简称"ISO 7098(1991)"。

20世纪90年代初制定ISO 7098(1991)时, 正处于世界进入信息时代的关键时期。为适应信息时代需求, 中国开始研制计算机汉字输入与输出。使用ISO 7098(1991)的汉语拼音, 可通过拼音-汉字转换的方法输入、输出汉字。由于汉语拼音是中国初等教育不可或缺的内容, 促使ISO 7098(1991)成为汉字输入、输出的一种便捷手段。在移动通信中使用汉语拼音在移动电话上输入汉字, 推动移动电话在中国的普及; 汉语拼音在汉语国际教育中也发挥很好的作用, 成为国外学生学习汉语和汉字的有用工具。

2 ISO 7098: 2015的修订过程

普通话是中国各民族的通用语言, 也是联合国工作语言之一。ISO 7098(1991)对全球信息与文献工作具有重要意义, 为满足当前国内外对汉语拼音实际应用的迫切需要, 有必要修改ISO 7098(1991)。

为此, 2011年3月教育部成立ISO 7098(1991)修订工作组, 由语言文字应用研究所研究员冯志伟担任组长, 傅爱平、李志江、黄伟、颜伟4位专家参加, 启动ISO 7098(1991)的修订工作。

2011年5月6日, ISO/TC 46第三十八届会议在悉尼召开, 中国代表在会议上提出修改ISO 7098(1991), 以便反映中文罗马化的新发展和实际应用需要的建议。

会后，中国国家标准化管理委员会（Standardization Administration of the People's Republic of China）正式向国际标准化组织（International Organization for Standardization，ISO）提出修订ISO 7098（1991）的新工作项目（New Working Item Proposal）提案，该提案的国际编号：N 2358。

2012年5月6日—11日，ISO/TC 46第三十九届会议在柏林举行，此次会议接受N 2358提案，并将该提案直接作为ISO 7098的工作草案（working draft），成立ISO 7098（1991）国际修订工作组，ISO 7098（1991）修订正式列入ISO的工作日程。

2013年6月3日—7日，在巴黎召开ISO/TC 46第四十届会议，中国代表在会议上正式向ISO/TC 46秘书处提交ISO 7098的委员会草案（committee draft）。

2014年5月5日—9日，ISO/TC 46第四十一届会议在华盛顿召开。中国代表在5月7日上午举行的第3工作组（Working Group 3，WG3）会议上，就ISO 7098（1991）修订问题重申立场，会后向ISO/TC 46秘书处提交ISO 7098国际标准草案（Draft of International Standard，DIS）。

2015年6月1日—5日，ISO/TC46第四十二届会议在北京召开。根据大会安排，中国代表在6月2日的专题报告会上发表《ISO 7098国际标准及其在人机交互中的应用》，且用生动的实例说明在数字化环境下，汉语拼音在人机交互中发挥的巨大作用，并受到各国代表的热烈欢迎；在6月3日的WG3会议上，就各国对ISO 7098DIS稿提出的意见，中国代表详细说明了处理情况，并向参会人员出示DIS修改稿。

会后，中国代表将DIS修改稿提交至ISO/TC 46秘书处，根据ISO/TC 46第四十二届会议的决议，ISO/TC 46秘书处于2015年7月27日将DIS修改稿分发给ISO/TC 46各成员国进行委员会内部投票（Committee Internel Balotting，CIB），CIB投票于2015年9月18日截止。ISO/TC 46秘书处N 2526号文件公布投票结果，ISO/TC 46中没有弃权的19个国家（保加利亚、加拿大、中国、克罗地亚、丹麦、爱沙尼亚、法国、德国、伊朗、意大利、日本、韩国、拉脱维亚、挪威、俄罗斯、泰国、乌克兰、英国、美国）都投了赞成票，获得全票通过。值得注意的是，在1982年对ISO 7098投反对票的美国和投弃权票的英国，均投了赞成票[3]。这说明ISO 7098在世界范围得到越来越多的国家支持。于是ISO 7098的修订工作进入出版阶段，形成新的修订稿，称为ISO 7098：2015。

2015年11月12日，中国代表向ISO/TC 46秘书处提交ISO 7098：2015的最终版本，并转至ISO总部准备出版。2015年12月15日，ISO正式出版ISO 7098：2015，有助于大幅提高文献自动化工作水平，使汉语拼音在国际文献工作中发挥更大的作用，并进一步推动汉语拼音走向世界。汉语中大多数常用词都是多音节词，因此，在国际文献和信息工作中，把单音节拼音连写为多音节的汉语单词是理所当然的事情，有必要把按词连写的规则引入国际标准[4]。

3 ISO 7098：2015的特点

ISO 7098：2015有四个引人注目的特点：一是将汉语拼音按词连写的规则引入国际标准；二是提出把汉字文本自动译音为拼音的方法；三是进一步完善汉语拼音的音节形式总表；四是给声调和标点符号补充16进制unicode代码，扩充罗马字母的字符集。以下分别进行具体说明。

3.1 将汉语拼音按词连写的规则引入国际标准

在中世纪之前，希腊人和罗马人已经知道“单词”的含义，尽管在文本中相邻单词间没有空白，其仍可识别出对应单词。

7世纪，爱尔兰人开始使用“空白”来分隔文本中的单词，并且将该方法传到法国。9世纪开始，使用空白分隔单词的方法在欧洲流行开来。

空白的使用意味着承认“单词”这个概念，在单词与单词间插入空白成为在书面使用字母语言的一个标准，世界出版界和图书馆都遵循该标准。

在汉语拼音中，也有必要使用空白来分割单词而非分割音节。单词的分割是世界文明的一个优良传统。在制定《中文罗马字母拼写法》时，遵循这样的优良传统是有利的。

在汉语拼音中，一个拼音音节可表示若干个汉字，因而在表示汉字方面，拼音音节存在歧义。如《通用规范汉字表》中拼音音节/bei/可以表示为31个汉字：北、杯、卑、背、椑、悲、碑、鹎、贝、孛、邶、狈、备、钡、倍、悖、被、琲、棓、辈、惫、焙、蓓、碚、鞁、褙、糒、鞴、鐾、呗、臂；拼音音节/jing/可以表示为49个汉字：京、茎、泾、经、猄、荆、菁、旌、惊、晶、腈、鹊、睛、粳、兢、精、鲸、麠、鼱、井、阱、洟、刭、胫、颈、景、儆、憬、璥、璟、

警、劲、径、净、迳、胫、倞、痉、竞、竟、竫、婧、靓、敬、靖、静、境、獍 镜。

在汉语拼音中，单音节歧义指数很高。若不包括声调，基本汉语音节有405个，这些汉语音节可表示全部汉字的读音。而《通用规范汉字表》中有8 105个通用汉字，在这种情况下，一个汉语音节平均可以表示20多个汉字，因而不可避免会出现歧义[5]。但若将几个单音节连接起来构成多音节单词，拼音音节的歧义指数就会大幅降低，因此为给拼音音节排除歧义，必须把不同单音节连接起来构成多音节汉语单词。

为解决汉语拼音音节存在歧义问题，使用拼音音节的歧义指数概念来描述拼音音节的歧义。歧义指数（I）是汉语拼音音节歧义程度的数学描述，与该拼音音节可以表示的语言单位数（N）的关系为$I = N–1$。

“语言单位”既可是单音节汉字，也可是单音节或多音节单词。

上文案例中，拼音音节/bei/可表示31个汉字，即有31个语言单位，其歧义指数为30；拼音音节/jing/可表示49个汉字，即有49个语言单位，其歧义指数为48。但若将单音节/bei/和/jing/结合形成双音节单词/beijing/，其歧义指数将明显减少，因为/beijing/可以表示3个单词：北京、背景、背静，即有3个双音节语言单位，其歧义指数减少（N=2）。若将/beijing/第一个字母进一步大写为/Beijing/，则其歧义指数为0，说明/Beijing/是一个没有歧义的语言单位，即中国首都北京。因此，若将不同单音节的拼音音节连接成多音节的汉语单词，拼音音节的歧义指数将明显降低。这是把不同的单音节连接成多音节汉语单词的优越之处。

GB/T 16159—2012《汉语拼音正词法基本规则》包括音节分割或连接成单词的规则，常用词（名词、动词、形容词、代词等）拼写规则，固定短语拼写规则，人名和地名拼写规则，声调表示规则，在行末尾的连字符使用规则等[6]。

目前，在汉语语言学中对汉语单词还没有公认的明确定义，这样很难确定汉语普通单词的边界（切分界线），把单个音节连接起来构成多音节单词时也将很困难。不过，汉语专有名词中单词的界限相对清晰，由于汉语中多音节的命名实体界限根据有关规范和标准比较容易确定，因此，把不同单音节连接构成多音节专有名词难度不大。在国际文献和信息工作中，把不同汉语拼音单音节连接起来构成多音节专有名词，从而避免拼音歧义，不仅是必要的也是可能的。基于此，在国际标准ISO 7098：2015中增加命名实体按词进行音节连写的规定，即在汉语拼音中对于人名、地名、语言名、民族名、宗教名这5种命名实体，均按词进行连写，将“按词连写”这个重要方法引进国际标准，与ISO 7098（1991）相比是重大的进展。例如，对于命名实体“地名”的书写，按国际标准ISO 7098：2015规定“汉语地名中的专名和通名（包括行政区划名或地理特征名）分写，由多个汉字组成的专名、行政区划名或地理特征名应分别按单词连写，每一分写部分的第一个字母大写”。

根据ISO 7098：2015规定，北京市“朝阳路”路名中的专名部分“朝阳”和通名部分“路”，应当分写且中间留空白。由于“朝阳”由两个汉字组成，拼写时应连写为一个单词，且每个分写部分的第一个字母均大写，因此“朝阳路”的汉语拼音规范书写形式应是“Chaoyang Lu”。而“Chao Yang Lu”（Chao和Yang没有连写为一个单词）、“ChaoYang Lu”（Yang的首字母不应大写）、“chaoyang Lu”（chaoyang的首字母没有大写）、“Chaoyang lu”（通名部分lu的第一个字母没有大写）、“chaoyang lu”（专名chaoyang和通名lu首字母均未大写）、“chaoyanglu”（专名chaoyang与通名lu没有分写且通名和专名的首字母均未大写）等拼写形式均不符合ISO 7098：2015规定。严格执行ISO 7098：2015标准，排除不符合规定的拼写形式，使得地名拼写形式统一，改变地名拼写的混乱局面，有助于人们无歧义地进行沟通。

20世纪60年代，联合国地名专家组为便于国际交往，使各国地名的专名部分只有一种拼写形式，避免在国际交往中地名因语言文字的复杂造成混乱。1967年第二届联合国地名标准化会议做出决议，要求世界各国、各地区在国际交往中都使用罗马字母拼写地名，做到每个地名的专名部分只有一种罗马字母拼写形式。选择罗马字母是因为世界上大多数国家均习惯使用，这就是“单一罗马化”（single Romanization）原则。如果严格执行ISO 7098：2015，将“朝阳路”拼写为“Chaoyang Lu”，便十分有利于贯彻推行“单一罗马化”原则。

地名的单一罗马化，对于使用罗马字母的国家而言，国家的地名标准化即国际标准化；而对使用非罗马字母文字的国家（如中国、日本、俄罗斯、泰国、韩国、希腊等）而言，须制定国家地名罗马化方案，经联合国地名标准化会议通过后，作为地名罗马字母拼写的国际

标准。1977年9月，在雅典召开的联合国第三届地名标准化会议上，中国代表提出《采用汉语拼音作为中国地名罗马字母拼法的国际标准》提案获得会议通过。第三届联合国地名标准化会议作出决定，“注意到《汉语拼音方案》在语言学上是完善的，用于中国地名的罗马字母拼法是最合适的”“建议采用汉语拼音方案作为中国地名罗马字母拼法的国际标准”。从此，根据《汉语拼音方案》拼写我国地名成为中国地名单一罗马字母拼写的国际标准，在以罗马字母为文字（如英文、德文、法文等）的各国出版物上都应根据《汉语拼音方案》拼写中文地名的专名部分。

在中文罗马字母拼写发展过程中，曾使用过与《汉语拼音方案》不同的罗马字母拼写形式拼写中文地名。其中，以英国人威妥玛和詹里斯在1867年设计的威妥玛式拼音（Wade Giles）、我国学者赵元任在1928年设计的国语罗马字（Guoyeu Romatzyh，GR）、美国人肯尼迪在1943年设计的耶鲁拼音（Mandarin Yale）最为有名，根据“单一罗马化”原则，在对中文地名进行罗马字母拼写时，不应采用多种罗马化形式拼写法，只能选择单一的罗马化拼音形式（即《汉语拼音方案》规定的拼音形式）。因此，在实行“单一罗马化”原则时，不能使用威妥玛式拼音、国语罗马字拼音、耶鲁拼音，而应使用《汉语拼音方案》的拼音方法。“北京”曾经有“Peking”“Pekin”等拼写形式，根据“单一罗马化”原则，应根据ISO 7098：2015拼写为“Beijing”，不能使用“Peking”“Pekin”等拼写形式。

在国际标准ISO 7098：2015中，还对“字符译音”（transcription）做出说明。其指出“字符译音是指用字母的语音系统或转换语言的符号来表示某种语言中的字符，而不论该语言原本的书写方式”“字符译音系统必须以转换语言及其字母表的正字法为依据，因此字符译音系统的使用者必须对转换语言了解，并能准确地读出其字符”“字符译音不是严格地可逆转的”“字符译音可用来转换所有的书写系统”“它是唯一能够用来转换如中文、日文这样的不全使用字母的拼音文字系统及意音图形文字书写系统的方法”。在把“朝阳路”转写成汉语拼音“Chaoyang Lu”时，其中的专名部分“Chaoyang”遵循“单一罗马化”原则，通名部分“Lu”也准确地反映汉语普通话的读音。这样的转写应属于“译音”的范畴。由此可见，ISO 7098：2015为把汉字地名正确译音，为拼音路名“单一罗马化”提供明确的规范。

3.2 提出命名实体自动译音方法

ISO 7098：2015提出，在计算机辅助文献工作中有两种对命名实体进行自动译音的方法。一种是按音节全自动译音，另一种是基于规则按单词半自动译音。

3.2.1 按音节全自动译音

全自动译音程序能自动生成彼此间由空白分开的单个音节，该方法适用于任何应用系统和环境，其音节切分结果效果显著，这种全自动译音程序特别适用于将拉丁字母译音与原汉字混合存储的系统。使用该方法“北京市”可全自动地译音为/bei/、/jing/和/shi/3个音节。这种全自动方法很容易通过计算机程序实现，但译音出来的音节歧义指数较高。

3.2.2 基于规则按单词半自动译音

在与语言有关的科学研究和工业生产中，“词”是基本和必要的概念，因此有必要对“词”以统一界说，但很难简单地使用基于空白和标点符号等规则来决定单词间的界限。这样的规则没有考虑到复合词、缩写词、惯用语等的切分问题，且单词切分对于单词与单词间没有用空白分开的语言（如汉语、日语）更加复杂[7]。

在自然语言处理中，单词切分即将文本切分为有负载意义的语言单位的过程。例如，英语“the white house”，可以切分为3个有意义的单位“the”“white”“house”，译为一间白色的房子；而“the White House”，则只与一个语言单位相对应，即美国总统的官邸。这种有意义的单位称为单词的切分单位（Word Segmentation Units，WSU）。对于单词间有空白的语言（如英语），在将文本切分WSU时，只需使用空白作为基础确定WSU切分的边界即可，简单易行；但对于单词间没有空白的语言（如汉语和日语），或对于只在局部单词间有空白的语言（如泰语和韩国语），在将书面文本切分为WSU时，要求使用不同的方法[8]。

很多应用领域需将文本切分为单词，在翻译中，统计单词数量是计算翻译工作成本的主要方法。在翻译记忆系统和机器辅助翻译（Computer-Assisted Translation）的工具中，单词切分是其主要功能；在术语抽取工具中，单词切分也起着重要的作用；在术语管理工具中，有时也要提供单词切分的功能；在内容搜索时，

也要对文本进行切分，以便在内容管理系统和数据库使用搜索词进行匹配；此外，搜索功能要求关于单词边界的知识，文本-语音转换系统在单词的基础上生成语音，因此要求在单词查询时进行单词切分等。各种自然语言处理系统必须把文本切分为单词才能实现其功能。

国际标准ISO 24614-1：2010给出自然语言处理中单词切分的基本概念和一般原则，提出以可信赖且能复用的方式进行书面文本自动切分的导则，且这种导则是独立于语言的[9]。

国际标准ISO 24614-2：2011提出汉语、日语和韩语中切分WSU的具体规则。其中，部分规则是这3种语言共同的，尽管每种语言都有独自判别WSU的特殊规则[10]。

因此，在中文罗马字母拼写中应将由汉字表示的命名实体译音为拼音，以表示单词。在汉语中单词可由一个或多个音节组成，单词间的界限并不清楚，在目前技术条件下，全自动单词切分难以达到很好的质量，可采用基于规则按单词半自动译音的方法。

命名实体基于规则按单词半自动译音可使用如下资源。

（1）一套译音规则。在本标准中提出命名实体译音的一般规则。这些规则可用作命名实体半自动译音的资源。

（2）一个相关的译音词典。《汉语拼音词汇（专名部分）》包含大多数命名实体的拼音译音，是可用作命名实体半自动译音的另一种资源。使用这样的方法“北京市”的译音过程：/bei jing shi/、/beijing shi/、/Beijing shi/、/Beijing Shi/。

根据规则，首先，地名“北京市”被切分为/bei/、/jing/和/shi/3个音节；然后，把/bei/、/jing/结合成/beijing/，使其与行政区划名/shi/分开；最后，把每部分首字母大写，译音为/Beijing Shi/。如果在按词译音过程中出现歧义或问题，编辑人员可根据译音词典通过人机交互找出合适的命名实体译音。因此，这种方法是半自动的，但译音质量很高，音节的歧义指数较小甚至可降低至0。

3.3 对汉语普通话的语音系统进行全面说明

ISO 7098：2015对汉语普通话的语音系统进行全面说明，使国际人士对汉语普通话的语音获得全面理解。

汉语普通话的声母包括双唇音（bilabial）：b，p，m；唇齿音（labio-dental）：f；舌面前音（dorso-prepalatal）：d，t，n，l；舌根音（dorso-velar）：g，k，h；舌尖前音（apico-alveolar）：z，c，s；舌尖后音（apico-postalveolar）：zh，ch，sh，r；舌面音（dorso-palatal）：j，q，x；零声母（zero initial）：在韵母的左侧没有元音。汉语普通话的韵母包括4种。

（1）开口呼（Articulation A）：以a、o、e为介音或主要元音的韵母。例如，a、o、e、ei、ao、ou、an、ang、en、eng、ong、er，以及zi、ci、si和zhi、chi、shi、ri中的主要元音i。

（2）合口呼（Articulation B）：以u为介音或主要元音的韵母。例如，u、ua、uo、uai、uei、uan、uang、un、ueng。

（3）齐齿呼（Articulation C）：以i为介音或主要元音的韵母。例如，i、ia、ie、iao、iu、ian、iang、in、ing、iong。

（4）撮口呼（Articulation D）：以ü为介音或主要元音的韵母。例如，ü、üe、üan、ün。在不会产生歧义的情况下，汉语拼音使用u代替ü，以简化音节拼写。

汉语普通话的音节形式表（见表1），覆盖汉语普通话中除音节ê和儿化音节外的所有音节。

此外，在ISO 7098：2015中，我们还对于这个音节形式表做出如下的说明。

（1）⁂ 表示零声母（在韵母的左侧没有声母）。

（2）*在音节开头的u写为w。但是，当w后没有其他附加元音时，作为一个完整音节的u不能写作w，而应写为wu。

（3）十 在zi，ci，si等音节中的i与在其他大多数音节中的i读音不同。这样的i在国际音标中写为ɿ。

（4）ㄓ 在zhi、chi、shi、ri等音节中的i与在其他大多数音节中的i读音不同。这样的i在国际音标中写为ʅ。

（5）+在音节开头的i写为y。但当这个y后面没有其他附加元音时，不能写作y、yn、yng，而应写作yi、yin、ying。

（6）※在不会产生歧义的条件下，汉语拼音使用u代替*ü*，仅是为便于拼写，这些u仍应读为*ü*。

（7）[1] wei: ui实际是*uei*的简写。因此，在汉语拼音声韵配合表中，有shui而没有shuei，有dui而没有duei。

（8）[2] wen: un实际是uen的简写。

（9）[3] you: iu实际是iou的简写。由于在音节开头的i写为y，所以应拼写为you而非yu（采用yu这样的拼写方法会导致混淆）。

（10）在该声韵配合表中，略去了儿化音节和音节ê。

表 1 汉语普通话音节形式表

	b	p	m	f	d	t	n	l	g	k	h	z	c	s	zh	ch	sh	r	j	q	x	(Null)
a	ba	pa	ma	fa	da	ta	na	la	ga	ka	ha	za	ca	sa	zha	cha	sha					a
o	bo	po	mo	fo																		o
e			me		de	te	ne	le	ge	ke	he	ze	ce	se	zhe	che	she	re				e
ai	bai	pai	mai		dai	tai	nai	lai	gai	kai	hai	zai	cai	sai	zhai	chai	shai					ai
ei	bei	pei	mei	fei	dei	tei	nei	lei	gei	kei	hei	zei			zhei		shei					ei
ao	bao	pao	mao		dao	tao	nao	lao	gao	kao	hao	zao	cao	sao	zhao	chao	shao	rao				ao
ou		pou	mou	fou	dou	tou	nou	lou	gou	kou	hou	zou	cou	sou	zhou	chou	shou	rou				ou
an	ban	pan	man	fan	dan	tan	nan	lan	gan	kan	han	zan	can	san	zhan	chan	shan	ran				an
ang	bang	pang	mang	fang	dang	tang	nang	lang	gang	kang	hang	zang	cang	sang	zhang	chang	shang	rang				ang
en	ben	pen	men	fen	den		nen		gen	ken	hen	zen	cen	sen	zhen	chen	shen	ren				en
eng	beng	peng	meng	feng	deng	teng	neng	leng	geng	keng	heng	zeng	ceng	seng	zheng	cheng	sheng	reng				eng
ong					dong	tong	nong	long	gong	kong	hong	zong	cong	song	zhong	chong		rong				
er																						er
u	bu	pu	mu	fu	du	tu	nu	lu	gu	ku	hu	zu	cu	su	zhu	chu	shu	ru				wu *
ua									gua	kua	hua				zhua	chua	shua	rua				wa *
uo					duo	tuo	nuo	luo	guo	kuo	huo	zuo	cuo	suo	zhuo	chuo	shuo	ruo				wo *
uai									guai	kuai	huai				zhuai	chuai	shuai					wai *
ui					dui	tui			gui	kui	hui	zui	cui	sui	zhui	chui	shui	rui				wei * [1]
uan					duan	tuan	nuan	luan	guan	kuan	huan	zuan	cuan	suan	zhuan	chuan	shuan	ruan				wan *
uang									guang	kuang	huang				zhuang	chuang	shuang					wang *
un					dun	tun	nun	lun	gun	kun	hun	zun	cun	sun	zhun	chun	shun	run				wen * [2]
ueng																						weng *
i	bi	pi	mi		di	ti	ni	li				zi **	ci **	si **	zhi ++	chi ++	shi ++	ri ++	ji	qi	xi	yi +
ia					dia			lia											jia	qia	xia	ya +
ie	bie	pie	mie		die	tie	nie	lie											jie	qie	xie	ye +
iao	biao	piao	miao		diao	tiao	niao	liao											jiao	qiao	xiao	yao +
iu			miu		diu		niu	liu											jiu	qiu	xiu	you + [3]
ian	bian	pian	mian		dian	tian	nian	lian											jian	qian	xian	yan +
iang							niang	liang											jiang	qiang	xiang	yang +
in	bin	pin	min				nin	lin											jin	qin	xin	yin +
ing	bing	ping	ming		ding	ting	ning	ling											jing	qing	xing	ying +
iong																			jiong	qiong	xiong	yong +
ü							nü	lü											ju #	qu #	xu #	yu #
üe							nüe	lüe											jue #	que #	xue #	yue #
üan																			juan #	quan #	xuan #	yuan #
ün																			jun #	qun #	xun #	yun #

3.4 补充16进制unicode代码

给汉语特有的声调和标点符号补充16进制unicode代码，丰富拉丁字母内容。汉语是一种有声调的语言，这意味着，声调会影响意义。具有不同声调的同一音节，其意义可能有很大的不同。每个音节可具有四个声调中的一个，也可以没有声调。在这个CD稿中，对普通话的四个声调符号加圆括号进一步说明其性质。

（1）一声（高平调）：—

（2）二声（升调）：/

（3）三声（降/升调）：∨

（4）四声（降调）：\

ISO 7098：2015还增加汉语普通话声调图示（见图1）。

对图1的分示图分别进行展示，以更便于国外用户理解汉语普通话声调的性质（见图2）。

根据ISO/TC 46第四十一届全会决议精神，要求在ISO 7098：2015中增加扩充拉丁字符使用的材料，因此对汉语拼音的声调符号和标点符号补充16进制的unicode代码（hexadecimal code，hex）。

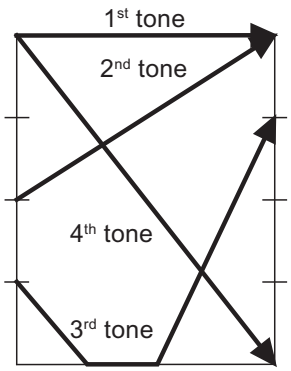

图 1 汉语普通话声调图示（综合图）

（1）一声（高平调）：—（hex：0304）

（2）二声（升调）：/（hex：0301）

（3）三声（降/升调）：∨（hex：030C）

（4）四声（降调）：\（hex：0300）

在实际文本中，声调符号附在音节的主要元音上。例如，/béi/、/què/，在音节/béi/中，声调符号附着在主要元音e上，标注为/é/；在音节/què/中，声调符号附着在主要元音e上，标注为/è/。如果区分元音大小写，则汉语普通话带声调符号的元音如表2和表3所示。

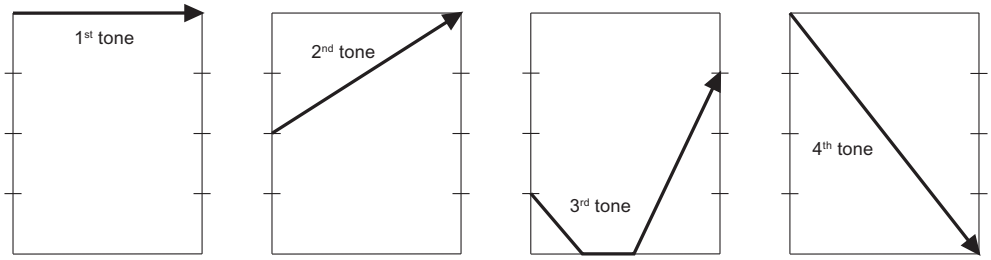

图 2 汉语普通话声调图示（分示图）

表 2 汉语带调元音小写字母16进制代码

Chinese vowel	1st tone		2nd tone		3rd tone		4th tone	
A	ā	hex：0101	á	hex：00E1	ǎ	hex：01CE	à	hex：00E0
E	ē	hex：0113	é	hex：00E9	ě	hex：011B	è	hex：00E8
I	ī	hex：012B	í	hex：00ED	ǐ	hex：01D0	ì	hex：00EC
O	ō	hex：014D	ó	hex：00F3	ǒ	hex：01D2	ò	hex：00F2
U	ū	hex：016B	ú	hex：00FA	ǔ	hex：01D4	ù	hex：00F9
Ü	ǖ	hex：01D6	ǘ	hex：01D8	ǚ	hex：01DA	ǜ	hex：01DC

表 3 汉语带调元音大写字母16进制代码

Chinese vowel	1st tone		2nd tone		3rd tone		4th tone	
A	Ā	hex：0100	Á	hex：00C1	Ǎ	hex：01CD	À	hex：00C0
E	Ē	hex：0112	É	hex：00C9	Ě	hex：011A	È	hex：00C8
I	Ī	hex：012A	Í	hex：00CD	Ǐ	hex：01CF	Ì	hex：00CC
O	Ō	hex：014C	Ó	hex：00D3	Ǒ	hex：01D1	Ò	hex：00D2
U	Ū	hex：016A	Ú	hex：00DA	Ǔ	hex：01D3	Ǔ	hex：00D9
Ü	Ǖ	hex：01D5	Ǘ	hex：01D7	Ǚ	hex：01D9	Ǜ	hex：01DB

汉语特有的标点符号及其对应的拉丁标点符号16进制的unicode代码如表4所示。

上述工作进一步丰富罗马字母字符集的内容，扩充罗马字母字符集，是汉语拼音对于罗马字母的贡献。

表 4 标点符号16进制代码对照表

Chinese mark		Latin mark		Note
。	hex：3002	.	hex：002E	full stop
、	hex：3001	,	hex：002C	special comma used to set off a short pause in the series
•	hex：2022	Space	hex：0020	disconnect mark
……	hex：2026 2026	…	hex：2026	horizontal ellipsis

4 结束语

ISO 7098（1991）将《汉语拼音方案》提高至国际标准的地位，是汉语拼音走向世界的第一步，如今ISO 7098：2015在罗马字母拼写时，进一步提出对命名实体按词连写的规则和自动译音方法，并对汉语普通话的语音系统进行全面描述，给汉语特有的声调符号和特有的标点符号增加了16进制代码，扩充罗马字母的字符集，这些富有成效的工作，迈开汉语拼音走向世界的新步伐。

参考文献

[1] 国家语委标准化工作委员会办公室.国家语言文字规范和标准选编[M].北京:中国标准出版社,1997:441.

[2] 国务院关于推广普通话的指示[EB/OL].(2011-01-17)[2016-12-13].http://www.seac.gov.cn/art/2011/1/17/art_58_106828.html.

[3] ALA-LC Romanization,Chinese,Rules of Application[EB/OL].[2016-12-13].http://www.loc.gov/catdir/cpso/romanization/chinese.df.

[4] Library of Congress,Pinyin Conversion Project,New Chinese Romanization Guidelines[EB/OL].[2016-12-13].http://www.loc.gov/catdir/pinyin/romcover.html.

[5] 中国国家标准化管理委员会.汉语拼音正词法基本规则:GB/T 16159—2012[S].北京:中国标准出版社,2012.

[6] Information and documentation:Romanization of Chinese:ISO 7098:2015[S].2015.

[7] Documentation—Romanization of Japanese(kana script):ISO 3602:1989[S].1989.

[8] Information and documentation—Transliteration of Korean script into Latin characters:ISO/TR 11941:1996[S].1996.

[9] Language resources management—Word segmentation of written text:Part 1:Basic concepts and general principles:ISO 24614-1:2010[S].2010.

[10] Language resources management—Word segmentation of written text:Part 2:Word segmentation for Chinese,Japanese and Korean:ISO 24614-2:2011[S].2011.

作者简介

冯志伟，1939年生，研究员，教授，博士生导师，研究方向：计算语言学、自然语言处理，E-mail：zwfengde2010@hotmail.com。

Four Distinguished Features of International Standard ISO 7098:2015

FENG ZhiWei[1,2]

(1.Institute of Applied Linguistics, Ministry of Education, Beijing 100010, China; 2.Hangzhou Normal University, Hangzhou 311121, China)

Abstract: A new International Standard ISO 7098:2015 was published at 15-December-2015 in Geneva. This paper analyzes four distinguished features of this new international standard.

Keywords: ISO 7098:2015; International Standard

（收稿日期：2016-10-12）