李宇明． 中国语言规划论．长春：东北师范大学出版社，2005：115-123．
原载《术语学与术语标准化》，2001（3）：29-32．

# 信息时代需要更高水平的语言文字规范

李 宇 明

**关键词** 信息处理；计算机；语言；文字；规范

**摘 要** 论述了语言文字是最基本的信息载体，分析了提高语言文字规范化水平对推进国民经济信息化的意义和作用，并对如何营造良好的社会语言环境提出了若干建设性的意见。

## Need High-Level Norm of Language in the Information Era

Li Yuming

**Key words** information processing, computer, language, character, norm

**Abstract** Author discusses that language is the basic information carrier, analyzes the significance and role of improving national economy information by high-level norm of language, and puts forward some constructive suggestions.

## 一、语言文字是信息处理的主要对象

21 世纪是广泛采用数字技术处理信息的时代。为迎接数字化信息时代的到来，《中华人民共和国国民经济和社会发展第十个五年计划纲要》提出，要“加速发展信息产业，大力推进信息化”，并且要“以信息化带动工业化”。这从一个侧面反映了国家的信息战略。但推进信息化有一个重要的前提，就是要提高语言文字的规范化水平。

**1. 语言文字是最基本的信息载体**

信息时代一个很重要的任务，就是对各种信息进行各种各样的加工处理。信息必须有载体负载。信息的载体很多，有语言文字，也有各种声音（除了语音）、图像、符号等等，在信息的诸多载体中，毫无疑问语言文字是最为主要的信息载体。这是因为：

（1）语言文字可以负载的信息领域几乎是无限的。非语言文字载体有其特殊的信息负载功能，有时可以辅助语言文字负载信息，但是同语言文字相比，其信息负载功能是非常受限的。

（2）人类使用语言文字负载信息非常方便，除使用“生理设备”和书写工具外，几乎不需其他特殊设备。

（3）非语言文字的信息最终需要翻译为语言文字信息，才能较好地为人所理解。

（4）人类历史上浩如烟海的信息主要是由语言文字负载的，今后依然是如此。

所以，语言文字是信息处理的主要对象，是声音信息处理、图像信息处理等所不能代替的。我国近几十年来所形成的许多著名的信息技术产业，如联想集团、方正公司、四通公司等，都与语言文字的处理，特别是与汉语汉字的处理有非常密切的关系。这种现象说明，当今语言文字信息处理的每一步进展，都会对我们的生活方式和工作方式带来巨大改变，如计算机的推广应用，出版印刷技术的重大革新等。同时也说明，语言文字信息处理的每一步进展，都会产生重大的经济效益。

**2. 语言文字信息处理的具体内容**

语言文字信息处理需要解决对语言文字本身的处理，以及对语言文字所负载的信息内容的处理等两个方面的问题。

除了语言文字研究等一些特殊领域之外，信息处理的目的当然是对语言文字所负载的信息内容的处理，但是对语言文字所负载的信息内容的处理有赖于对语言文字本身的处理，因为要让计算机高质量地理解语言文字所负载的信息内容，必须首先让计算机能够“看懂”文字、“听懂”语音，能够“写出”文字、发出语音，能够拥有一定的词汇量并且学会从语流和文字串中识别词语，能够掌握一定的语法规则、语用规则、语义常识乃至百科知识。

几十年的科技实践表明，对语言文字本身处理的难度是远远超出人们意料的。我国对汉语汉字的处理取得了举世瞩目的成就：汉字的输入输出、内码的存储和交换等获得了突破性进展；语音识别和语音合成预计在五年左右的时间内可以有较大的实用性进展。这是值得在文化史上和信息技术史上大书特书的事情，是无数科学工作者和语言文字工作者几十年间艰苦奋斗才取得的。但是，语音、文字都只是语言的物质外壳，“字词处理”，包括“语音处理”是对语言的物质外壳的处理，还谈不上真正的“语言处理”。当前，中文信息处理学界在不断完善“字词处理”的同时，正在向“词处理”、“句处理”和“篇章处理”的方向发展，即进行真正的“语言处理”。

“语言处理”的难度要比“字词处理”，包括“语音处理”的难度更大。语音和文字尽管复杂，但它们必然有明确的物质形式，而且其成员也是有限的，一种语言中有多少个音节、多少个字符，大体上是可以枚举穷列的。但是，语言却是开放的体系，“语言处理”必然要对语义、语法和语用进行处理。语义本身没有外显的物质形式，而且学术界对语义的认识还相当肤浅，十分有限；语法是词语构成（词法）、句子构成（句法）和篇章构成（超句法）的各种规则，其复杂程度要高于语音和文字；语用牵涉到交际环境和语言使用者的各种社会属性、交际心理和知识经验等等，处理的难度更大。因此，词处理、句处理、篇章处理的技术征程还相当崎岖漫长，可以预见，当前及今后相当长的一个阶段，中文信息处理的攻坚之战仍将是对语言文字本身的处理。

## 二、语言文字规范对信息处理的意义

语言文字的信息处理本来就有不小难度，如果语言文字的规范化程度不高，更会使信息处理的难度加大。为保证语言文字信息处理的顺利进行，必须尽量使语言文字规范化。语言文字规范，包括语言文字标准的制定与推行，是促进并实现语言文字规范化的重要举措。

**1. 我国语言文字规范化所取得的成就和存在的问题**

新中国成立几十年来，我国已经制定了不少语言文字规范（包括标准），这些规范为我国的语言文字规范化和社会语言生活现代化起到了重要的作用：

（1）确定了现代汉民族共同语（普通话）的规范，并在全国范围内进行推广，在一定程度上消除了因方言、语言分歧所造成的交际障碍，同时为汉语走向世界创造了语言条件。

（2）从字种、字形、字音、字量等方面对汉字进行了大规模的整理和简化，制定了现行汉字的规范和汉字在不同领域的一些应用规范，完善了新式标点符号，提倡横向书写，使现代汉语书面语在形式上有了基本规范。

（3）制定了汉语拼音方案及其有关规定，使现代汉语和现行汉字有了现代化的拼写工具与注音工具，使汉语的罗马字母转写有了国际认可的标准，并为汉语汉字的计算机输入提供了方便。

但是，也应当看到，我国已有的语言文字规范还有不能令人完全满意之处。例如，有些规范还不够完善；规范的系统性不很强，有许多领域、特别是许多语言文字的应用领域还没有规范，规范之间还有不协调的地方。造成这些问题的原因除社会政治因素，如“文化大革

命”时期的混乱之外，主要是：第一，语言文字及其应用非常复杂，科学研究的底子薄弱，不能为规范的制定提供有力的科研支持；第二，现有规范是在不同的时期制定的，不同时期规范制定者的语言文字观念和规范观念有差异，而且统一协调、更新维护的工作没有及时跟上；第三，随着时代的进步，社会语言生活在快速发展变化；第四，现代化的科研手段没有得到规范制定者的及时利用。

为了保证社会语言文字的规范化，加快国家的信息化进程，应当加大语言文字规范建设的力度，加快语言文字规范制定的步伐。

**2. 语言文字规范的两大类别**

语言文字规范主要分为两大类：一类是面向人的语言文字规范，一类是面向机器，包括计算机、多媒体、因特网和其他信息产品的语言文字规范。面向人的规范与面向机器的规范有许多不同，其中最主要的不同有三点：

（1）面向人的语言文字规范一般说来需要柔一些，需要有一定的弹性，面向机器的语言文字规范一般说来需要刚一些。

（2）面向人的语言文字规范要尽量保持稳定，面向机器的语言文字规范应根据技术的发展及时更新维护。

（3）哪些现象和领域需要规范，哪些现象和领域不需要规范，人和机器的要求也不大相同。

这些不同起因于人运用语言文字的特点与机器处理语言文字的特点不同。人的最大特点是多样性。我国民族多方言多，语言成分复杂；国民受教育的程度有较大差异，社会上的不同行业对语言文字规范的需求不同，而且人的语言文字习惯一经形成就不大容易改变。因此，面向人的语言文字规范需要有较大的弹性，需要尽量保持稳定，所要规范的是那些在人与人的交际中容易引起混乱的地方。机器的最大特点是要求一致性，而且更新换代较快。

**3. 语言文字规范化对信息化的意义和作用**

我国现有的语言文字规范，多是面向人的，面向机器的非常之少。当前，语言文字规范建设除了完善面向人的规范之外，还应特别重视面向机器的规范建设。面向机器的语言文字规范主要有两种：

a. 为便于机器处理而制定的语言文字规范，如规定基本的词汇集、语法集，规定计算机用的汉字部件及汉字部件名称，如规定机器用的轻声、儿化字表和词表等等。这些规范就人们的一般语言生活来说，有些是不必要的，有些带有很强的人为性。

b. 面向机器的语言文字规范是机器处理语言文字时的规范，如对汉语汉字的各种形、音的键盘编码规则，机器用的词语切分标准，汉语词性的机用标注方式，以及汉字的存储码和交换码等。

a 种规范是为了方便计算机的语言文字处理，b 种规范是计算机处理语言文字的各种带有较强技术性的“协议”，以便于成果共享，便于社会应用和行政、技术等方面的管理。这两种规范的制定，都需要有一定的语言文字学知识和计算机语言文字处理的知识，需要语言文字学家和计算机专家密切合作，需要国家的语言文字主管部门和信息产业主管部门、国家标准的主管部门相互协作沟通，只有单方面的努力是远远不够的。

从理论上来说，面向人的语言文字规范和面向机器的语言文字标准可以不同，但是，计算机、多媒体和因特网正在快速推广应用，“海量”真实文本正在成为计算机语言文字处理的对象，社会语言生活对计算机语言文字处理的影响越来越大，计算机语言文字处理的发展对社会语言生活的影响也越来越大，因此，面向人的语言文字规范对机器的语言文字处理会产生越来越多、越来越大或直接或间接的影响，面向机器的一些语言文字规范也会对社会语言文字的应用发生越来越多、越来越大或直接或间接的影响。这就要求在制定这两种语言文字规范时应统筹兼顾，尽量缩小差距，减少分歧。这种发展趋势对制定语言文字规范的要求

越来越高，制定规范的难度也越来越大。

## 三、为信息化营造良好的社会语言环境

**1. 应加速社会语言生活的规范化**

前面提到，计算机、多媒体和因特网正在快速推广应用，“海量”真实文本正在成为计算机语言文字处理的对象。此种状况下，每个语言文字使用者所输出的语言文字，都可能成为计算机要处理的真实文本。例如，要实现寻呼电话的自动接转，寻呼台的“机器小姐”就必须听懂每个寻呼人的话语；因特网是现代化的信息媒体，逐渐成为最重要的信息“集散地”，这就要求计算机能够正确处理网络上的所有文本，包括 BBS 中出现的各种文本；掌上电脑常常采用手写的方式输入信息，因此必须识别各种各样的手写字体。计算机介入日常生活越深，每个语言文字使用者所输出的语言文字成为计算机处理文本的可能性也就越大。

当然，计算机是为人服务的，应能为人们的语言生活提供各种便利。但是，真实文本的处理难度很大，为促进语言文字信息处理的顺利发展，也为社会成员能够有效利用信息产品，应当加速社会语言生活的规范化，以增强计算机要处理的各种文本的规范性，加快我国信息化的进程。

**2. 强化人们的信息化意识、规范意识和法律意识**

社会语言生活规范化，首先要求人们在语言文字领域要有信息化意识、规范意识和法律意识，特别是要认真贯彻《中华人民共和国国家通用语言文字法》，认真执行国家颁布的通用语言文字规范。在这方面，公务员、新闻媒体、印刷出版行业、教育战线应起模范带头作用，通讯、交通、旅游、商务等社会公共服务行业要自律。青年人是使用现代信息产品最积极的社会阶层，是现代信息产品的最大受益者，因此应当成为语言文字规范化的尖兵。其次，要逐渐完善执法保障，有一套切实可行的推广普通话、推行规范汉字的规章制度，特别是应当建立有效的劳动市场准入制度，把掌握国家通用语言文字的水平同职业要求有机联系起来。第三要通过学校教育和社会教育，迅速提高全社会的运用普通话和规范汉字的能力。

如果说语言文字的规范化过去主要是文化问题、交际问题和国家形象与国民内聚力问题的话，那么今天它又成为重要的社会经济问题，成为综合国力的一个重要构成要素，成为高科技发展的一个“瓶颈”问题。这个问题解决得好坏，直接关系到我国的信息化进程和信息安全，并影响到国家的综合国力。因此，语言文字规范化问题，应当引起政府、社会各行各业和每一位公民的高度重视。

**参考文献**

[1]陈敏，王翠叶．　中文信息处理的现状与展望．　语言文字应用，　1995　（4）．

[2]董振东．　机器翻译与汉语研究．　语文建设，　1992　（4）．

[3]范继淹，徐志敏．　人工智能与语言学．　中国语文，　1980　（4）．

[4]傅永和．　中文信息处理．　广州：　广东教育出版社，　1999．

[5]黄昌宁．　关于处理大规模真实文本的谈话．　语言文字应用，　1993　（2）．

[6]黄昌宁，童　翔．　汉语真实文本的语义自动标注．　语言文字应用，　1993　（4）．

[7]侯敏．　计算语言学与汉语自动分析．北京：　北京广播学院出版社，　1999．

[8]李宇明　（2001a）．　通用语言文字规范和标准的建设．　语言文字应用，　2001　（2）．

[9]李宇明　（2001b）．　规范语言文字，推进信息化进程．　中国教育报，　2001－05－07．

[10]刘坚　主编．　二十世纪的中国语言学．　北京：　北京大学出版社，　1998．

[11]刘开瑛．　中文文本自动分词和标注．　北京：　商务印书馆，　2000．

[12]刘涌泉，乔毅．　应用语言学．　上海：　上海外语教学出版社，　1991．

[13]陆俭明． 跨入新世纪后我国汉语应用研究的三个主要方面． 中国语文， 2000 （6）．
[14]陆致极． 计算语言学导论． 上海： 上海教育出版社， 1990．
[15]全国人大教科文卫委员会教育室， 教育部语言文应用管理司． 中华人民共和国国家通用语言文字法学习读本． 北京：语文出版社， 2001．
[16]孙茂松 邹嘉彦． 汉语自动分词中的若干理论问题． 语言文字应用， 1995 （4）．
[17]王宁． 计算机古籍字库的建立与汉字的理论研究． 语言文字应用， 1994 （1）．
[18]许嘉璐． 《语言文字学及其应用研究》． 广州： 广东教育出版社， 1999．
[19]许嘉璐 （2000a）． 现状和设想——试论中文信息处理与现代汉语研究． 中国语文， 2000 （6）．
[20]许嘉璐 （2000b）． 《未成集－－论新时期语言文字工作》． 北京： 语文出版社， 2000．
[21]姚亚平． 《中国计算语言学》． 南昌： 江西科学技术出版社， 1997．
[22]袁琦，陈力为． 九十年代中文信息处理技术的基本任务． 语文建设， 1992 （1）．
[23]袁毓林． 计算机语言学的理论方法和研究取向． 中国社会科学， 2001 （4） ．
[24]于根元． 《二十世纪的中国语言应用研究》． 太原： 书海出版社， 1996．
[25]于根元主编． 《世纪之交的的应用语言学》． 北京： 北京广播学院出版社， 2000．
[26]于玮． 现代信息社会呼唤语言文字规范化． 光明日报， 2000－11－05．
[27]俞士汶． 关于计算语言学的若干研究． 语言文字应用， 1993 （5）．
[28]张庆旭． 汉语真实文本自动语义标注取得突破性进展． 语文建设， 1993 （9）．