

# 人工智能中的语言问题

李 蓝

**摘 要：**本文从四个方面讨论人工智能中的语言问题。首先探讨人工智能时代的语言定义。第二部分介绍语言的生理机制。第三部分介绍目前语言工程学及人工智能领域的语言工程方向。最后本文比较了人类语言编码机制与机读二维码在信息交换方面的差别，认为人类语言因为必须遵循时间顺序原则，就信息交换的效率来说，远不及机读扫码的信息交换方式。因此，以后的人工智能机器人必须考虑人-机、机-机的双向能力设计。

**关键词：**人工智能 语言工程 仿生技术

## 一、人工智能时代的语言定义

关于什么是语言，中国人两千多年前就有思考。孔子曾说“不学诗，无以言”（《论语·尧曰篇第二十》）。这里的“言”，就是说话的意思。孔子死后，他的学生们把他的“言”整理成一本书，名叫“论语”。其性质和“某某语录”一样。

《诗·大雅·公刘》“于时语语”，毛传：“直言曰言，论难曰语。”陈奂传疏：“论难者，理有难明，必辩论之不已也。”这个讨论被许慎直接引入《说文解字》：直言曰言，论难曰语。

于此可见，在中国古代，一般的说话叫“言”，讨论问题时说的话叫“语”。“语”和“言”的意义有所不同。

下面是两个现在的通行定义，代表现代中国主流学术界对语言的传统认识。

《现代汉语词典》（第6版）（中国社会科学院语言研究所词典研究室2012）：

人类所特有的用来表达意思、交流思想的工具，是一种特殊的社会现象，由语音、词汇和语法构成一定的系统。语言一般包括它的书面形式，但在与“文字”并举时只指口语。

《辞海·语言文字分册》（上海辞书出版社2010）：

语言，人类最重要的交际工具。它同思维有密切的联系，是人类形成和表达思想的手段，也是人类社会最基本的信息载体。人类借助语言保存和传递文明的成果。语言是人区别于其他动物的要本质特征之一。

这两个定义都强调语言的人类属性，强调语言对人类社会的重要性。

但从维纳开始，科学家开始不再承认语言只属于人类，否认语言只属于人类的专有属性。陈明远（1984）在《语言学与现代科学》里说：

现代科学中的“语言”概念，比过去大大扩充了。什么是“语言”呢？语言是人类社会特有的一种信息系统，是人们用来进行交际和思维活动的有效工具。有声语言（口语）及其文字记录（书面语），不过是广义语言中的特殊形式，即“自然语言”的形式。随着现代科学和计算机技术发展的需要，各式各样的人工语言已被设计出来了。这些人工语言也同祥具有词汇形式、语法规则、语句结构和语义内容，等等。语言学课题已经成了现代科学的基本课题之一。

到我们今天，人类社会已完全进入了计算机时代，语音机器人已与人类进行了完整的“语言交际”，语言的确已不再专属于人类，现代计算机被人类赋予了语言能力。

## 二、人类语言的生理机制

本节主要以《新编心理语言》（桂诗春，2000）和《生理学概要》（朱大章、陆利民，2002）《语音学教程》（林焘、王理嘉，1992）三书为基础，简要介绍人类语言的生理基础。

人类要能说出话，首先要有管理整个语言行为和语言能力的神经中枢。人类控制语言能力的大脑神经机制是借助一些病例逐渐发现的。

1861年，法国外科医生，神经解剖学家保尔·布罗卡（P. Broca）在巴黎召开的人类学会议上介绍一个病例。一个病人的发音器官完全是正常的，但说话非常困难，而这个病人能听懂别人的话，能用面部表情或借助手势与别人交流。当这个病人去世后，解剖检查发现，他的大脑第3额左回发生严重病变，出现一个鸡蛋大小的肿瘤。布罗卡医生遂设想，这个区域可能是大脑语言中枢的所在地。现在，大脑这个区域的语言功能已基本确认，医学界把这个区域称为布罗卡区。

1874年，德国神学家卡尔·韦尼克（Wernike）报告了另一病例。这个病人能说话，也能听见别人说话，但他无法理解自己说的话，也听不懂别人的话。病人去世后解剖检查发现，这个病人大脑第3额上回有病变。韦尼克于是推测，这个区域可能与理解语言的意义有关。这个发现现在也已基本得到证实，这个大脑区域遂被命名为韦尼克区。

现在进一步的研究发现，韦尼克区是指大脑左半球后部颞顶叶较广泛的区域。

布罗卡区和韦尼克区组成了语言中枢的主要部分。

人类语言的生理机制除了神经机制外，还有心理机制。心理机制包括记忆、输入、理解、输出等几个方面。

1. 记忆。在语言信息处理模型里，外部信息需要经过一系列的心理结构来理解、编辑、存储和提取，其核心部分是记忆。言语的习得和产出都离不开记忆。没有了记忆我们的一切言语和思维活动，感知和推测能力都成了无源之水。

2. 语言输入的心理过程。语言输入是个认知和言语交际的过程，也是极为复杂的生理、心理过程。学习者通过听觉、阅读把感知到的语言符号先在韦尼克区内进行解释，然后传递到布罗卡区以备表达之用。在这个过程中，学习者要完成符号辨认和理解两个阶段。作者的思想通过语言符号传达给学习者。学习者则对一连串的语言符号进行信息处理，使之意义化，并加以理解、接受然后储存到记忆中，形成各种各样的图式。

输入（听、读）的关键在于理解，仅仅辨认语言符号而不能理解是无意义的。学习者为了达到最终的理解，应使语言符号意义化，将其与自己已具备的知识、经验联系起来加以思考、推测、判断、证明或修正。学习者已具备的同材料相关的知识结构（图式）是理解的基础。学习者的知识面越大，图式的数量越多，质量越高，理解也越好。

3. 语言产出的心理过程。人们在说话、写作时，先从韦尼克区提取词汇，然后将它们传递到布罗卡区，在这里进行文字、语义和句法的编码，再决定言语的具体形式或语音，并向书写中枢或言语运动中枢发出指令，完成言语的产出。

人类产生语言有大脑中形成后，还需要发音器官的协作动作，最后才能发出语音。人类的发音器官由三个部分组成。

一是动力系统。动力部分由肺、横膈膜、气管组成。当需要发音时，肺部排出气流，气流受横膈膜调节，经气管冲击声带，声带振动后发出声音。

二是声带。声带位于喉头的中间，是两片富有弹性的带状薄膜。两片声带之间的空隙叫声门，肌肉的收缩，杓状软骨活动起来可使声带放松或收紧，使声门打开或关闭，从肺中出来的气流通过声门使声

音振动发出声音，控制声带松紧的变化可以发出高低不同的声音来。

三是共鸣区。包括口腔、鼻腔和咽腔三个部分。口腔（包括唇、齿和舌头）后面是咽腔，咽头上通口腔、鼻腔，下接喉头。口腔和鼻腔靠软腭和小舌分开。软腭和小舌上升时鼻腔关闭，口腔畅通，这是发出的声在口腔中共鸣，叫口音。软腭和小舌下垂，口腔成阻，气流只能从鼻腔中发出，这是发出的音主要在鼻腔中共鸣，叫做鼻音。如果口腔没有阻碍，气流从口腔和鼻腔同时呼出，发出的音在口腔和鼻腔同时产生共鸣，叫鼻化音（也叫半鼻音或口鼻音）。

### 三、人工智能的两种工程模式

人类获得了语言能力，也发明了记录语言的文字符号，但在录音技术发明之前，人类的声音信号一直无法真实记录和保存，无法进行有效的科学研究。

1877年7月18日美国科学家爱迪生发明了录音机。这个录音机可以将声波转换成金属针的震动，然后将波形刻录在圆筒形蜡管的锡箔上。当针再一次沿着刻录的轨迹行进时，便可以重新发出录下的声音。这是人类历史上最重要的发明之一，人类从此后可以突破时空限制，记录声音，保存声音，研究声音。

1941年，在研究探测德军潜艇的声呐技术的过程中，美国贝尔实验室发明了语音视频分析仪。这种仪器利用声音的音色、音高、音强、音长等物理属性，把声音的能量信息转换成绘制在专用图纸上的可视化图片信息，这是人类第一次看见声音。从而揭开了人类研究语音的新篇章。这个发明的重要性可以和录音机相提并论。

到现在，科学家已经可以不要专门仪器，只要安装了语音分析软件，利用普通的计算机就可以生成语图了。进入计算机时代以来，声音处理技术的发展完全可以用日新月异来形容。到20世纪70年代，已可以初步人工合成语音（语音指的是人类发出的声音），可以不要发音

器官，也不要人，就可以直接“做”声音。

最近，高度智能化，可以和人类自由对话的语音机器人已出现，最著名的有日本的安德鲁和安德罗，已取得沙特公民身份的索菲亚，还有最近刷屏的谷歌助手。

这是两种不同工程方向的语音机器人。安氏兄妹和索菲亚是一类，属仿真机器人，这种机器人的终极目标是做出一个和真人一样的机器人。从语音仿真程度看，索菲亚的声音“机器味儿”还比较重，真人的语音仿真水平不如日本的安德鲁和安德罗。面部表情是索菲亚的仿真表情做得更好，基本可以和人类互动。

谷歌智能语音助手完全不考虑人类的外形仿真问题，只做语音。对话的智能水平已达到普通人的标准，语音的仿真水平可以说已到了完美的地步，完全可以以假乱真，接电话的“活人”完全不知道对方是一个语音机器人。

但这三种人工智能机器人用的都是同一种智能语音技术：程控法。这种方法完全不考虑人类发音器官如何运作等问题，完全从程序控制、语音识别、语音合成出发，最后生成对话的语句，用机器内部的播音喇叭播出声音。重要的技术难点有两个：一是复杂语义的理解及话语应对策略的设计，二是语音合成的仿真程度。

编程法合成语言的历史始于20世纪60年代。最初的做法是先写好一些词语，做成语图，放在数据库里，成为机器人理解人类语言的“词典”，计算机“听”到人类的语音后，把声音录下来，做成语图，送到数据库中去匹配，匹配上之后把相应的语图转换成声音或文字符号，播放出来或显示出来，从而达到人机交互。

外还有一种完全仿人类语言形成过程的智能语音机器人。这种机器人设计的目标是完全模仿人类的发音器官，做出像人一样说话的机器人来。日本的科学家对此兴趣一直不减。

语音机能仿真的技术目前看已完全无法和程控法竞争，在智能语音机器人这个领域，完全可以预见，以后这种方法不可能有竞争力。

但这种方法通过对人类语音形成的过程的完整模仿，可以使科学家充分了解人类语言的发音原理，在语言医学领域会有很大的发展空间。

#### 四、人机通讯与机机通讯

2018年5月9日，Google首席执行官皮查伊展示了一段谷歌语音助手为一位女士预订理发的录音，这段对话展示的语音技术可以说技惊四座。下面是我们转写的对话内容。

(铃声响起)

Hair Salon: Hello how can I help you?

Google Assistant: Hi,I'm calling to book a women's haircut for a client.I'm looking for something on May 3rd.

Hair Salon: Sure,give me one second.

Google Assistant: Mm-hmm.

Hair Salon:Sure, what time are you looking for around?

Google Assistant:At 12 pm.

Hair Salon:We do not have a 12 pm available. The closet we have to that is a 1:15.

Google Assistant:Do you have anything between 10 am and 12 pm?

Hair Salon:Depending on what servise she would like. What service is she looking for?

Google Assistant:Just a women's haircut, for now.

Hair Salon:Okay,we have a 10 o'clock.

Google Assistant:10 am is fine.

Hair Salon:Okay,what's her first name?

Google Assistant:The first name is Lisa.

Hair Salon:Okay,perfect. So I will see Lisa at 10 o'clock on May 3rd.

Google Assistant:Okay great,thanks.

Hair Salon:Great.Have great day.Bye.

从技术的角度说，这个谷歌语音助手输出端语音合成的拟真度堪称完美。句法成分之间的韵律构成和词语之间停顿时长都得到精确控制，贯穿整个语段的语调和美式英语的契合度非常高。从输入端看，谷歌语音助手没有听错任何词语或句子，并用准确的回答展示了其语义分析的水平和能力。

这是一段特定对象和特定内容的对话，不是如人类聊天一样东拉西扯的开放式话题。因此，还不能就此对谷歌的语音技术做出完整评价。但就这段对话展示的语音技术来说，已接近达到完美水平，在对话题复杂度和对话内容宽泛度进行扩展后，一个全能的语音助手已呼之欲出。

最后我们要探讨一个事实上已经来临的问题：如果是两个谷歌语音助手进行对话，那应该如何设计信息交换符号？换句话说，如果不是人机交互而是机机交互，如果以后形成机器人社区，机器人之间交流信息，还会继续使用人类语言吗？

回答这个问题可以从两个方面来观察。

人类语言的编码和解码都是线性方式：说话人沿着时间轴线编码，形成符号串，再沿着时间轴线一字一句输出（说话）。听话人逆向而行：沿着时间轴线接收声音信号，反向解码，提取话语信息，理解说话人的意思。

在信息传递过程中，人类语言编码和解码都要花费大量时间，最后才获得有效信息。

如果采用二维码的来编码解码，那会发生什么情况？

二维码（2-dimensional bar code），是一种用某种特定的几何图形

按一定规律在平面（二维方向上）分布的黑白相间的图形来记录数据符号信息的编码。

以日本设计的 QR CODE 为例，二维码在编码解码方面有以下优点：

1. 信息存储容量大；
2. 可有效表现各种符号，包括汉字；
3. 可从任意方向读取信息；
4. 传递的是信息，没有歧义；
5. 有比较强的抗污损能力；
6. 可视化符号，可快速编码，解码。

尤其是其中的可视化符号，直接传递信息，无固定读取方向这三项，这可以使二维码可瞬间编码解码。就信息交换来说，二维码具有人类语言无法比拟的效率和速度。

因此，下一步设计智能机器人时，我们应该为机器人设计两套信息交换符号系统。一套用于人机交互，一套用于机机交互。

### 参考文献

1. 陈明远：《语言学与现代科学》，成都：四川人民出版社，1984年。
2. 桂诗春：《新编心理语言》，上海：上海外语教育出版社，2000年。
3. 夏征农、陈至立：《辞海》，上海：上海辞书出版社，2010年。
4. 林焘、王理嘉：《语音学教程》，北京：北京大学出版社，1992年。
5. N. 维纳：《控制论——关于在动物和机器中控制和通讯的科学》，罗劲伯、侯德彭、龚育之、陈步译，北京：科学出版社，1961年。
6. 中国社会科学院语言研究所词典研究室：《现代汉语词典》（第6版），北京：商务印书馆，2012年。
7. 朱大章、陆利民：《生理学概要》，上海：复旦大学出版社，2002年。