[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 17/28 (2006.01)

[12] 发明专利申请公开说明书

[21] 申请号 03825926.5

[43] 公开日 2006 年 3 月 1 日

[11] 公开号 CN 1742273A

[22] 申请日 2003.4.23 [21] 申请号 03825926.5

[30] 优先权

[32] 2002.12.10 [33] US [31] 10/315,732

[86] 国际申请 PCT/US2003/012514 2003.4.23

[87] 国际公布 WO2004/053725 英 2004.6.24

[85] 进入国家阶段日期 2005.8.5

[71] 申请人 国际商业机器公司

地址 美国纽约

[72] 发明人 高雨青 顾 良 刘富华

杰弗里·索里森

[74] 专利代理机构 中国国际贸易促进委员会专利商标事务所

代理人 康建忠

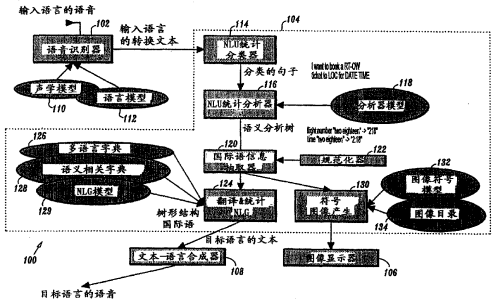
权利要求书 3 页 说明书 7 页 附图 3 页

[54] 发明名称

多模态语音－语音语言翻译和显示

[57] 摘要

提供一种把源语言的自然语句翻译成符号表现和/或目标语言的多模态语音－语音语言翻译系统和方法。系统(100)包括把源语言的自然语句(402)输入系统(100)的输入装置(102)；接收呈机器可读形式的自然语句(402)，并把自然语句(402)翻译成符号表现(404)和/或目标语言(406)的翻译器(104)；和显示自然语句的符号表现(404)的图像显示器(106)。另上，图像显示器(106)指示目标语言(406)的文本，符号表现(404)和源语言(402)的文本之间的相关性(408)。



1、一种语言翻译系统，包括:

把源语言的自然语句输入系统的输入装置;

接收呈机器可读形式的自然语句，并把自然语句翻译成符号表现的翻译器; 和

显示自然语句的符号表现的图像显示器。

2、按照权利要求 1 所述的系统，还包括以可听见的方式产生目标语言的自然语句的文本-语音合成器。

3、按照权利要求 1 所述的系统，其中输入装置是把口语单词转换成机器可识别文本的自动语音识别器。

4、按照权利要求 1 所述的系统，其中翻译器还包括:

从自然语句分析结构信息，并输出自然语句的语义分析树表示的自然语言理解分析器。

5、按照权利要求 1 所述的系统，其中翻译器还包括:

对自然语句的元素分类，并依据类别标记元素的自然语言理解统计分类器; 和

从分类句子分析结构信息，并输出分类句子的语义分析树表示的自然语言理解分析器。

6、按照权利要求 5 所述的系统，其中翻译器还包括抽取自然语句的独立于语言的表现的国际语信息抽取器。

7、按照权利要求 6 所述的系统，其中翻译器还包括通过使独立于语言的表现的元素与可视描述相联系，产生自然语句的符号表现的符号图像生成器。

8、按照权利要求 6 所述的系统，其中翻译器还包括把独立于语言的表现转换成目标语言的自然语言生成器。

9、按照权利要求 1 所述的系统，其中翻译器把自然语句翻译成目标语言的文本，图像显示器显示目标语言的文本以及符号表现。

10、按照权利要求 3 所述的系统，其中翻译器把自然语句翻译成目

标语言的文本，图像显示器显示目标语言的文本，符号表现和源语言的文本。

11、按照权利要求 10 所述的系统，其中图像显示器指示目标语言的文本，符号表现和源语言的文本之间的相关性。

12、一种翻译语言的方法，所述方法包括下述步骤：

接收源语言的自然语句；

把自然语句翻译成符号表现；和

显示自然语句的符号表现。

13、按照权利要求 12 所述的方法，其中接收步骤包括下述步骤：

接收口语自然语句作为声学信号；和

把口语自然语句转换成机器可识别的文本。

14、按照权利要求 13 所述的方法，还包括下述步骤：

从自然语句分析结构信息，并输出自然语句的语义分析树表示。

15、按照权利要求 14 所述的方法，还包括从语义分析树，抽取自然语句的独立于语言的表现的步骤。

16、按照权利要求 13 所述的方法，还包括下述步骤：

对自然语句的元素分类，并依据类别标记元素；和

从分类句子分析结构信息，并输出分类句子的语义分析树表示。

17、按照权利要求 16 所述的方法，还包括从语义分析树抽取自然语句的独立于语言的表现的步骤。

18、按照权利要求 17 所述的方法，还包括通过使独立于语言的表现的元素与可视描述相联系，产生自然语句的符号表现的步骤。

19、按照权利要求 18 所述的方法，还包括下述步骤：把独立于语言的表现转换成目标语言的文本，并显示目标语言的文本以及符号表现。

20、按照权利要求 19 所述的方法，还包括以可听见的方式产生目标语言的文本的步骤。

21、按照权利要求 20 所述的方法，还包括突出所显示符号表现的和目标语言的可听文本对应的元素的步骤。

22、按照权利要求 19 所述的方法，还包括下述步骤：使目标语言的

文本，符号表现和源语言的文本相关联，和显示目标语言的文本，符号表现和源语言的文本之间的相关性。

23、一种机器可读的确实包含可由机器执行，从而实现翻译语言的方法步骤的指令程序的程序存储装置，所述方法步骤包括:

接收源语言的自然语句;

把自然语句翻译成符号表现；和

显示自然语句的符号表现。

多模态语音-语音语言翻译和显示

美国政府对此发明具有已付费的许可，和在有限的情况下根据如与海军空间和海军作战系统中心签定的合同 No.N66001-99-2-8916 的条款的合理的条款要求专利所有人许可其它人的权利。

技术领域

本发明涉及语言翻译系统，更具体地说，涉及一种多模态语音-语音语言翻译系统和方法，其中源语言被输入系统中，被翻译成目标语言，并由各种模态（modality），例如显示器、语音合成器等输出。

背景技术

可见图像用于人类交流非常古老和基本。从洞穴壁画到今天的儿童图画，图画、符号和图象表示在人类表达中一直起着重要的作用。图像和空间形体不仅被用于表现场景和物理对象，而且还被用于表现过程和更抽象的概念。随着时间的过去，象形文字体系，即可视语言已进化成更多地依赖于约定俗成，而不是依赖于它们的表现力的相似性的字母表和符号体系。

可视语言在有限的领域中被广泛使用。例如，在世界的多数地方，交通符号和公共场所中的国际性设施图标，例如电话，公共厕所，饭店，紧急出口等得到广泛接受和理解。

在过去的二三十年内，人们一直对用于人/机交互作用的可视语言，例如图形界面，图形编程语言等兴趣强烈。例如，微软的 Windows™界面使用带有文件夹，文件柜，垃圾桶，绘图工具和已成为个人计算机的标准的其它惯用对象的桌面隐喻，因为它们使计算机更易于使用和更易于学习。但是，随着由于旅行的便利，通信媒体，例如因特网的速度的提高，和市场的全球化的结果，全球社会变得越来越小，可视语言将在

语言不同的人们之间的交流中起日益重要的作用。另外，可视语言能够方便根本不能说话的人(例如聋人)，或者文盲间的交流。

由于可视语言的下述特征：(1)国际性-可视语言不依赖于特定的口头语言或书面语言；(2)由可视表示法的使用而产生的易学性；(3)便于绘图能力受损者使用的计算机辅助创作和显示；(4)自适应性(例如对于视力受损者的更大显示，对于色盲者的重新着色，对于初学者的消息的更明确表达)，和(5)高级的可视化技术，例如动画的应用(参见 Tanimoto, Steven L.,"Representation and Learnability in Visual Languages for Web-based Interpersonal Communication",IEEE Proceedings of VL 1997,1997 年 9 月 23-26)，可视语言用于人类间的交流的潜力巨大。

发明内容

提供一种把源语言的自然语句翻译成符号表现和/或目标语言的多模态语音-语音语言翻译系统和方法。本发明使用自然语言理解技术对口语句子中的概念和语义分类，把句子翻译成目标语言，并使用可视显示(例如图画，图像，图标或任何视频片段)向双方，例如讲话者和听者表示句子中的主要概念和语义，以帮助用户相互理解，并帮助源语言用户核实翻译的正确性。

旅行者熟悉可视描述(例如在关于行李和出租车的机场标记中使用的那些可视描述)的有效性。本发明通过连同口语输出一起，把这些及其它这样的图像包含到要显示的符号表现中，把同样的特征引入到交互式谈话模型中。符号表现甚至可包含动画，以静态显示所不能的方式指示主语/宾语和动作关系。

根据本发明的一个方面，语言翻译系统包括把源语言的自然语句输入系统的输入装置；接收呈机器可读形式的自然语句，并把自然语句翻译成符号表现的翻译器；和显示自然语句的符号表现的图像显示器。系统还包括以可听见的方式产生目标语言的自然语句的文本-语音合成器。

翻译器包括对自然语句的元素分类，并依据类别标记元素的自然语言理解统计分类器；和从分类句子分析结构信息，并输出分类句子的语

义分析树表示的自然语言理解分析器。翻译器还包括抽取自然语句的独立于语言的表现的国际语信息抽取器，和通过使独立于语言的表现的元素与可视描述相联系，产生自然语句的符号表现的符号图像生成器。

根据本发明的另一方面，翻译器把自然语句翻译成目标语言的文本，图像显示器显示目标语言的文本，符号表现和源语言的文本，其中图像显示器指示目标语言的文本，符号表现和源语言的文本之间的相关性。

根据本发明的另一方面，提供一种翻译语言的方法。所述方法包括下述步骤：接收源语言的自然语句；把自然语句翻译成符号表现，并显示自然语句的符号表现。

接收步骤包括下述步骤：以声学信号的形式接收口语自然语句；和把口语自然语句转换成机器可识别的文本。

在本发明的另一方面，所述方法还包括下述步骤：对自然语句的元素分类，并依据类别标记元素；从分类句子分析结构信息，并输出分类句子的语义分析树表示；和从语义分析树抽取自然语句的独立于语言的表现。

此外，所述方法还包括通过使独立于语言的表现的元素与可视描述相联系，产生自然语句的符号表现的步骤。

在另一方面，所述方法还包括下述步骤：使目标语言的文本，符号表现和源语言的文本相关联，并显示目标语言的文本，符号表现和源语言的文本之间的相关性。

根据本发明的另一方面，提供一种机器可读的确实包含可由机器执行的，从而实现翻译语言的方法步骤的指令程序的程序存储装置，所述方法步骤包括接收源语言的自然语句；把自然语句翻译成符号表现；和显示自然语句的符号表现。

附图说明

结合附图，根据下面的详细说明，本发明的上述及其它方面、特征和优点将变得更明显，其中：

图 1 是根据本发明的一个实施例的多模态语音-语音语言翻译系统的

方框图；

图 2 是图解说明根据本发明的一个实施例的把源语言的自然语句翻译成符号表现的方法的流程图；

图 3 是图解说明源语言的自然语句的符号表现的多模态语音-语音语言翻译的例证显示；

图 4 是用源语言和目标语言如何与符号表现相关联的指示符，图解说明源语言的自然语句，该句子的符号表现和翻译成目标语言的句子的多模态语音-语音语言翻译系统的例证显示。

具体实施方式

下面将参考附图说明本发明的优选实施例。在下面的说明中，未详细描述公知的功能或构造，以避免使本发明掩蔽在不必要的细节中。

提供一种把源语言的自然语句翻译成符号表现和/或目标语言的多模态语音-语音语言翻译系统和方法。本发明通过增加设备显示的输入句子的图形或符号表现的附加翻译，延伸了语音识别技术，自然语言理解技术，语义翻译技术，自然语言产生技术和语音合成技术。通过包括可视描述(例如，图画，图像，图标或视频片段)，翻译系统向(源语言的)说话者指示语音被适当地识别和理解。另外，可视表现向双方指出由于翻译多义性的缘故，可能不正确的语义表现的多个方面。

任意语言的可视描述本身是一个挑战-尤其是对于抽象对话来说。但是，由于用于创建“国际语”（interlingua）表示法，即与语言无关的表示法的自然语言理解处理的缘故，在翻译过程中，可获得匹配适当图像的额外机会。在这种意义上，可视语言可被看作语言产生系统以其为目标的另一目标语言。

要明白可用各种形式的硬件、软件、固件、专用处理器，或者它们的组合物来实现本发明。在一个实施例中，本发明可用软件实现成确实包含在程序存储装置上的应用程序。所述应用程序可被上传到包含任何适当体系结构的机器上并由所述机器执行。最好，在具有硬件，例如一个或多个中央处理器(CPU)，随机存取存储器(RAM)，只读存储器(ROM)

和输入/输出(I/O)接口，比如键盘、光标控制装置(例如鼠标)和显示装置的计算机平台上实现所述机器。所述计算机平台还包括操作系统和微指令代码。这里描述的各种进程和函数或者是微指令代码的一部分，或者是通过操作系统执行的应用程序(或者它们的组合)的一部分。另外，各种其它外围设备可与计算机平台连接，例如附加的数据存储装置和打印装置。

还要明白，由于可用软件实现在附图中描述的一些系统组件和方法步骤，因此根据对本发明进行程序设计的方式，系统组件(或者进程步骤)之间的实际连接可能不同。根据这里提供的本发明的教导，本领域的普通技术人员能够设想本发明的这些及类似的实现或构造。

图 1 是根据本发明的一个实施例的多模态语音-语音语言翻译系统 100 的方框图，图 2 是图解说明把源语言的自然语句翻译成符号表现（representation）的方法的流程图。下面将参考图 1 和 2 详细说明所述系统和方法。

参见图 1 和 2,语言翻译系统 100 包括把自然语句输入系统 100(步骤 202)的输入设备 102，接收呈机器可读形式的自然语句并把该自然语句翻译成符号表现的翻译器 104，和显示自然语句的符号表现的图像显示器 106。可选的是，系统 100 将包括以可听见的方式产生目标语言的自然语句的文本-语音合成器 108。

最好，输入装置 102 是与用于把口语单词转换成计算机或机器能够识别的文本单词(步骤 204)的自动语音识别器(ASR)耦接的麦克风。ASR 接收声音信号，比较该信号和输入源语言的声学模型 110 和语言模型 112，从而把口语单词转换成文本。

可选的是，输入装置是直接输入文本单词的键盘或者把手写文本转换成计算机能够识别的文本单词(步骤 204)的数字写字板或者扫描仪。

一旦自然语句呈计算机/机器可识别的形式，翻译器 104 就处理所述文本。翻译器 104 包括自然语言理解(NLU)统计分类器 114，NLU 统计分析器（parser）116，国际语信息抽取器 120，翻译和统计自然语言生成器 124 和符号图像生成器 130。

NLU 统计分类器 114 从 ASR 102 接收计算机可识别文本，查找句子中一般类别的位置，并标记某些元素(步骤 206)。例如，ASR 102 可输入句子“I want to book a one way ticket to Houston,Texas for tomorrow morning”。NLU 分类器 114 将把 Houston,Texas 分类为地点“LOC”，并把它替换到输入句子中。此外，one way 将被解释成票的类型，例如往返或单程(RT-OW)，tomorrow 将被“DATE”替换，morning 将被“TIME”替换，从而得到句子“I want to book a RT-OW ticket to LOC for DATE TIME”。

分类后的句子随后被发送给 NLU 统计分析器 116，抽取结构信息，例如主语/动词(步骤 208)。分析器 116 与分析器模型 118 交互作用，从而确定输入句子的语法结构，和输出语义分析树。可关于特定的领域，例如运输、医疗等构成分析器模型 118。

随后国际语信息抽取器 120 处理语义分析树，确定输入源句子的独立于语言的含义，也称为树形结构国际语(步骤 210)。国际语信息抽取器 120 与规范器（canonicalizer）122 耦接，规范器 122 把由文本表示的数目转换成由周围上下文确定的恰当格式化的数字。例如，如果输入文本“flight number two eighteen”，那么将输出数字“218”。此外，如果输入“time two eighteen”，那么将输出时间格式的“2: 18”。

一旦确定了树形结构国际语，那么初始的输入源自然语句可被翻译成任何目标语言，例如一种不同的口语，或者翻译成符号表示法。对于口语来说，国际语被发送给翻译&统计自然语言生成器 124，从而把国际语转换成目标语言(步骤 212)。生成器 124 访问多语言字典 126，以便把国际语翻译成目标语言的文本。随后用语义相关字典 128 处理目标语言的文本，明确表达要输出的文本的恰当含义。最后，用自然语言发生模型 129 处理该文本，用按照目标语言的可理解句子构成所述文本。目标语句随后被发送给文本-语音合成器 108，以便以可听见的方式产生目标语言的自然语句。

国际语还被发送给符号图像生成器 130，以便产生将在图像显示器 106 上显示的可视描述的符号表现(步骤 214)。符号图像生成器 130 可访

问图像符号模型，例如 Blissymbolics 或 Minspeak，以产生符号表现。这里，生成器 130 将抽取恰当的符号来创建代表初始源句子的不同元素的“单词”，并把“单词”聚集在一起，从而传达初始源句子的预定含义。另一方面，生成器 130 将访问图像目录 134，选择复合图像来表示国际语的元素。一旦构成了符号表现，它将被显示在图像显示装置 106 上。图 3 图解说明源语言的初始输入的自然语句的符号表现(步骤 216)。

除了本发明的翻译系统的功能优点之外，共用图形显示的存在极大地增加了说话者和听者的用户体验。不共用任何语言的人们之间的交流既困难又紧张。可视描述促进了共同体验的感觉，并向共同领域提供恰当的图像，从而通过手势或者通过连续的一系列交互作用简化交流。

在本发明的翻译系统的另一实施例中，显示的符号表现将指示口语对话的哪一部分对应于显示的图像。图 4 中图解说明了本实施例的例证屏幕。

图 4 图解说明说话者说出的源语言的自然语句 402，源句子的符号表现 404，和源句子 406 到目标语言(这里为中文)的翻译。线条 408 指示图像对应的每种语言中的言语部分，因为流畅的语言翻译通常要求词序的改变。通过连接单词和短语的可视描述，并指示在每种语言中，它们出现在口头短语中的何处，听者能够更好地利用说话者提供的韵律提示，当前的语音识别系统通常未记录的提示。

可选的是，当文本-语音合成器可听地产生其对应的单词或概念时，呈现在图像显示器上的每个图像被突出。

在另一实施例中，系统将检测讲话者的情感，并把“情感”，例如“:-)”包含到目标语言的文本中。通过关于音调和音色分析接收的声音信号，可检测讲话者的情感。另一方面，本领域中已知，通过经神经网络分析捕捉的讲话者的图像，照相机将捕捉讲话者的情感。随后使讲话者的情感与机器可识别的文本相联系，供稍后翻译之用。

虽然参考本发明的一些优选实施例表示和说明了本发明，不过在不脱离由附加权利要求限定的本发明的精神和范围的情况下，可在细节和形式方面做出各种变化。

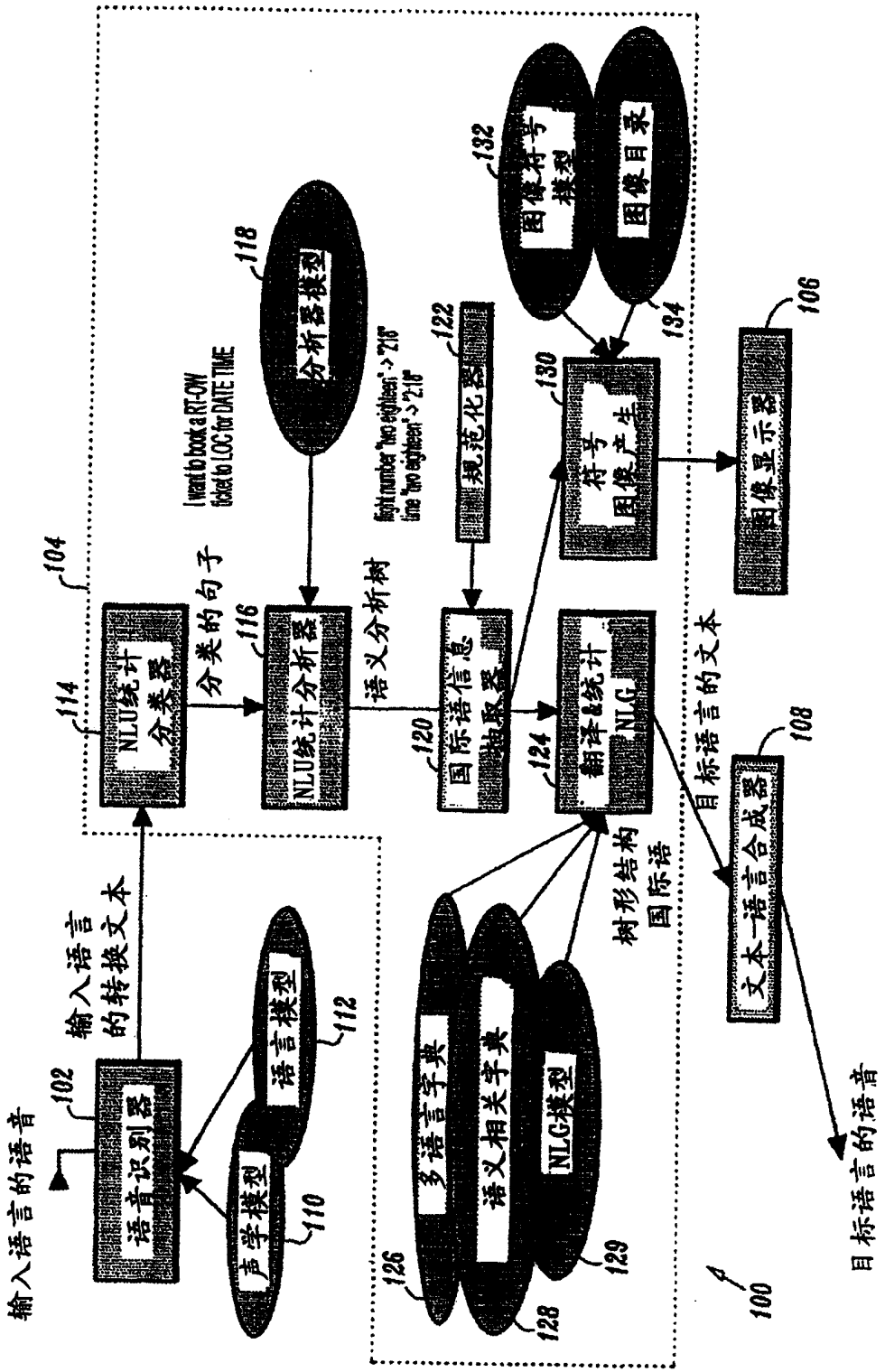

输入语言的语音
102
语音识别器
输入语言的转换文本
104
NLU统计分类器
114
分类的句子
116
NLU统计分析器
语义分析树
118
分析器模型
I want to book a RT-OW ticket to LOC for DATE TIME
flight number 'two eighteen' -> '218'
time 'two eighteen' -> '2:18'
122
规范化器
120
国际语信息抽取器
124
翻译&统计NLG
树形结构国际语
130
符号图像产生
132
图像符号模型
134
图像目录
106
图像显示器
目标语言的文本
108
文本-语言合成器
目标语言的语音
100
声学模型
110
语言模型
112
多语言字典
126
语义相关字典
128
NLG模型
129

图 1

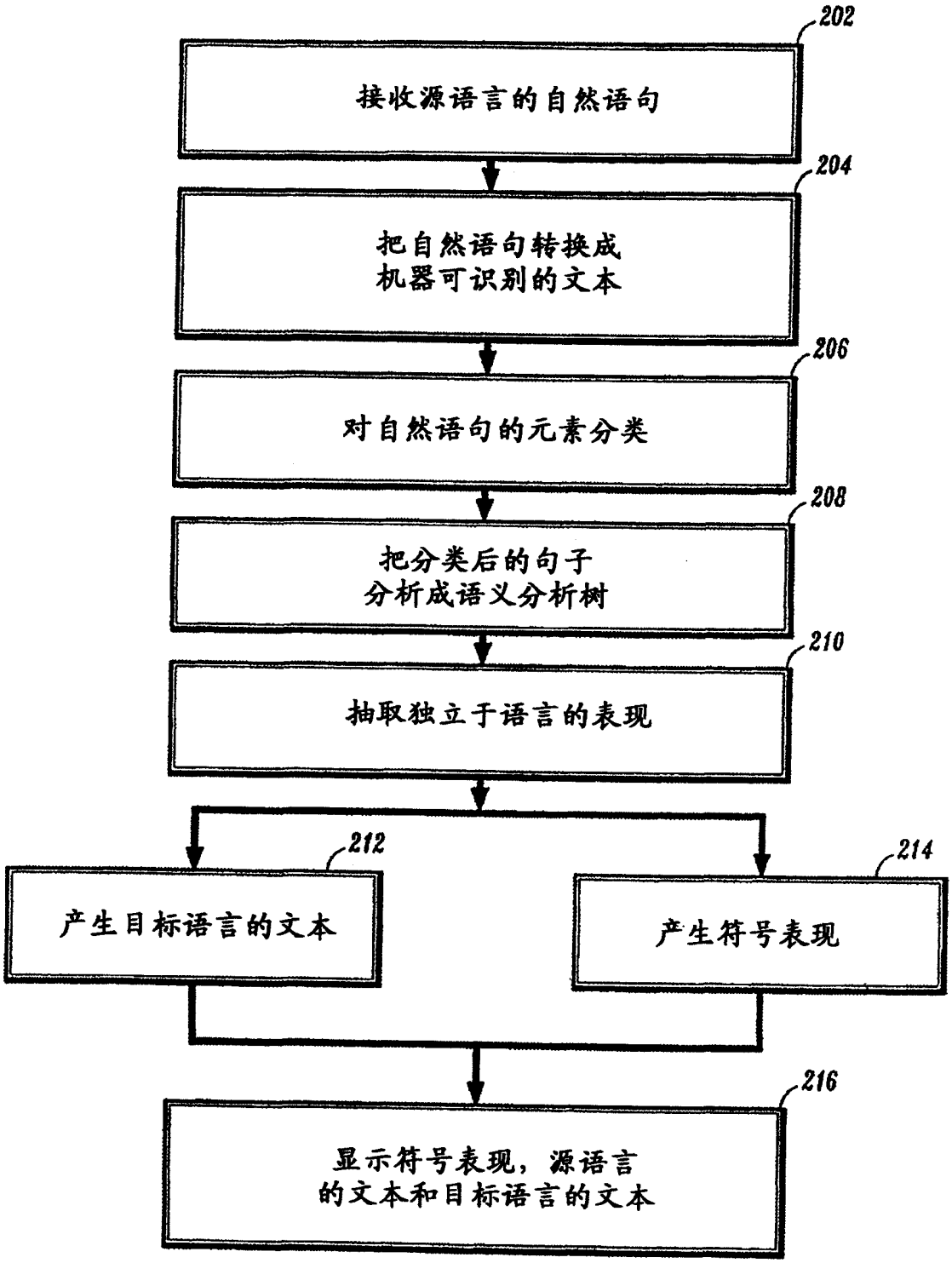

202
接收源语言的自然语句
204
把自然语句转换成
机器可识别的文本
206
对自然语句的元素分类
208
把分类后的句子
分析成语义分析树
210
抽取独立于语言的表现
212
产生目标语言的文本
214
产生符号表现
216
显示符号表现，源语言
的文本和目标语言的文本

图2

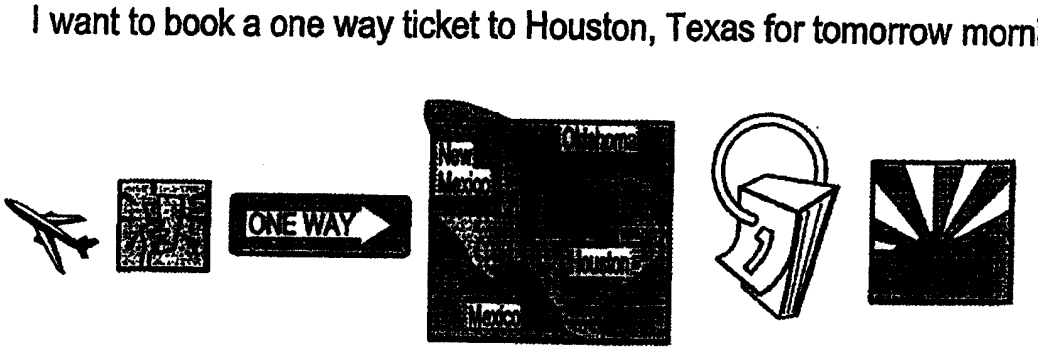

I want to book a one way ticket to Houston, Texas for tomorrow morning
ONE WAY

图 3

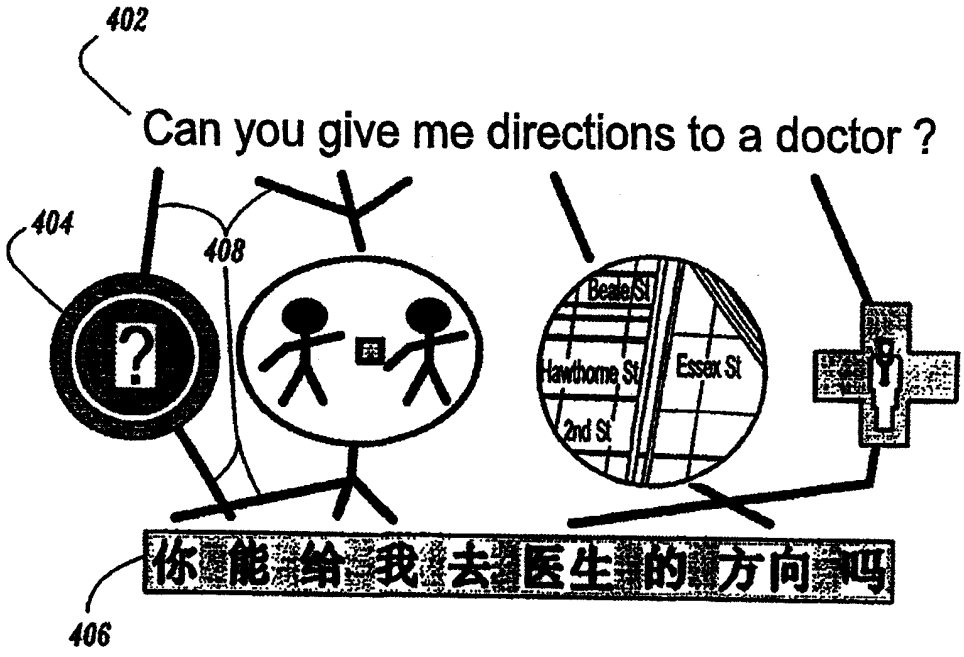

402
Can you give me directions to a doctor ?
404
408
Beale St
Hawthorne St
Essex St
2nd St
你能给我去医生的方向吗
406

图 4