

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

G10L 15/00

G06F 17/20



[12] 发明专利申请公开说明书

[21] 申请号 200410003433.9

[43] 公开日 2005 年 1 月 5 日

[11] 公开号 CN 1560834A

[22] 申请日 2004.3.9

[21] 申请号 200410003433.9

[71] 申请人 北京沃克斯科技有限责任公司

地址 100086 北京市海淀区北三环西路 43 号
青云当代大厦十七层

[72] 发明人 郑丽明

权利要求书 3 页 说明书 7 页

[54] 发明名称 一种口语识别理解的语法描述与识别分析的方法

[57] 摘要

本发明公开一种口语识别理解的语法描述与识别分析的方法，涉及到人工智能的语音识别技术和口语理解技术，特别是涉及到根据识别理解的任务编写计算机程序的语义句法规则和根据这种语义句法规则来理解输入计算机的语音或文字的意图。本发明把语义句法描述和识别理解操作有机地结合在一起，构成形式化的任务描述规范和可操作的识别分析规范，识别器同时完成识别和理解的双重任务，更加符合口语识别理解的实际应用要求，尤其当处理具有复杂语义结构的语音输入和文字输入的时候，从识别器就直接得到语义表达准确的理解结果，便于人机口语对话系统的设计、实现与移植。

1、一种口语识别理解的语法描述与识别分析的方法，其特征在于，把语义句法描述和识别理解操作结合在一起，识别器同时完成识别和理解的双重任务，更加符合口语识别理解的实际应用要求，尤其当处理具有复杂语义结构的语音输入和文字输入的时候，从识别器就直接得到语义表达准确的理解结果，便于人机口语对话系统的设计、实现与移植。

2、如权利要求1所述的口语识别理解的语法描述与识别分析的方法，其特征在于，构成形式化的任务描述规范和可操作的识别分析规范。

3、如权利要求2所述的口语识别理解的语法描述与识别分析的方法，其特征在于，所述的形式化的任务描述规范包括以下三部分：

- 1) 文本规则；
- 2) 句法规则；
- 3) 语义规则。

4、如权利要求2所述口语识别理解的语法描述与识别分析的方法，其特征在于，所述的可操作的识别分析规范包括以下三个步骤：

- 1) 语法编辑；
- 2) 语法编译；
- 3) 识别分析。

5、如权利要求3所述的形式化的任务描述规范，其特征在于，所述的文本规则，规定如下：

- 1) 文本注释：用“/*”和“*/”作为文本注释的起始和结束。在对文档的后续处理中，免除对“/*”和“*/”之间所有内容的处理；
- 2) 字符串：以双引号界定；
- 3) 空格：在语义句法文件中，除了文本注释、字符串、语义注解以外的地方，具有对文字的定界作用。

6、如权利要求3所述的形式化的任务描述规范，其特征在于，所述的句法规则，规定如下：

- 1) 规则定义：“文字符号 = 表达式；”。等号左边的文字符号，是规则名称。
等号与分号之间的表达式，是规则的语义句法的具体表示；
- 2) 规则展开：从等号开始，到分号结束，按照从左到右的顺序对表达式进行展开；
- 3) 规则引用：在规则名称前冠以'\$'。例如，Rule 是一条规则，\$Rule 表示对规则 Rule 的引用；
- 4) 句法符号和文本定界符号：(假定 A 和 B 表示两个简单表达式)：
 - a) 圆括号 ‘(’ 和 ‘)’，如 (A B)，表示对规则的封装，圆括号内部的表达式被作为一个整体单元对待；
 - b) 竖杠 ‘|’，如 A|B，表示匹配 A 或匹配 B；
 - c) 空格 ‘ ’，如 A B，表示先匹配 A，随后匹配 B，优先级比 ‘|’ 高：
A B | C D 等同于 (A B) | (C D)；
 - d) 减号 ‘-’，如 A-B，表示匹配 A，但不匹配 B；
 - e) 加号 ‘+’，如 A+，表示对 A 匹配一次以上，优先级比 ‘|’ 高：
A+ | B+ 等同于 (A+) | (B+)；
 - f) 星号 ‘*’，如 A*，表示对 A 匹配零次以上，优先级比 ‘|’ 高： A* | B* 等同于 (A*) | (B*)；
 - g) 前斜线 ‘/’ 和 ‘/’，如 /0.6/A | /0.5/B，表示按 0.6 概率匹配 A 或 0.4 概率匹配 B；
 - h) 方括号 ‘[’ 和 ‘]’，如 [A]，表示可以匹配 A，也可以不匹配 A。

7、如权利要求 3 所述的形式化的任务描述规范，其特征在于，所述的语义规则，包括语义标签、语义变量、语义注解，语义嵌套，规定如下：

- 1) 语义标签：由计算机可显示的符号组成，‘%’ 除外；
- 2) 语义变量，以 ‘%’ 开始，后接大小写字母符号（包括下连线 ‘_’ 和中连线 ‘-’），以空格和除下连线 ‘_’ 和中连线 ‘-’ 以外的其它标点符号结束；
- 3) 语义注解：由语义标签和语义变量串接组成，用花括号 ‘{’ 和 ‘}’ 作为语义注解的起始和结束。语义注解不参与识别器的匹配运算；
- 4) 语义嵌套：当前规则的语义注解的值，取自所引用规则的语义注解的值。

8、如权利要求 4 所述的可操作的识别分析规范，其特征在于，所述的语法编辑，

遵循权利要求 3 所述的规范，采用纯文本的形式，针对系统设计的任务，编写语义句法规则文件。

9、如权利要求 4 所述的可操作的识别分析规范，其特征在于，所述的语法编译，把采用纯文本编写的语义句法规则文件转换成为适合计算机进行动态优化匹配的语义句法网络。该语义句法网络，由连接网络起始点和终结点的一条以上相互交叉或互不相交的路径构成。从起始点到终结点的每一条网络路径，对应语义句法规则文件中的一条识别匹配的规则。每一条网络路径包含一个以上的节点和两条以上的连接节点的弧。每个节点上存储语块及其声学模型用于进行识别匹配；对于要求进行语义注解的节点，另外存储语义标签和语义变量。连接两个节点的弧还要存储权值，用于对节点识别匹配的值进行加权。

10、如权利要求 4 所述的可操作的识别分析规范，其特征在于，所述的识别分析，是指在权利要求 9 所述的语义句法网络中，对输入计算机的语音或文字进行动态规划匹配，找到最优匹配的路径。位于最优匹配路径上的节点如果要求语义注解，先将该节点最佳匹配的语块赋给语义变量，然后按照该最优路径上节点的先后顺序，把各个节点上的语义标签和语义变量的值，串接成为一个输出字符串，作为识别分析的输出。

一种口语识别理解的语法描述与识别分析的方法

技术领域

本发明涉及人工智能的语音识别技术和口语理解技术，特别是涉及根据识别理解的任务编写计算机程序的语义句法规则和根据这种语义句法规则来理解输入计算机的语音或文字的意图。

背景技术

口语识别理解的语法描述和分析方法是指这样的计算机技术：它是人机口语对话系统的一个组成部分。它把人机口语对话系统承担的任务，用人可以阅读的语义句法文件进行描述；经过编译，把语义句法文件转换成计算机可以直接调用的语义句法网络；在该语义句法网络中，对人的语音输入或文字输入进行识别分析，获得人发出语音信息或文字信息的意图。

口语识别理解系统一般使用分立的识别器语法和理解器语法。系统工作时，先用识别器语法将人的语音输入转换成文字字符串或个别词汇带有语义标签的文字字符串，然后用理解器语法对识别器的输出结果进行分析，得到人的语音输入所包含的意图。对于纯文字形式的输入，则直接用理解器语法进行分析，得到人的文字输入所包含的意图。

随着语音浏览器(Voice Browser)应用的兴起，要求同时具备识别理解功能的语音识别器语法受到重视。概括起来，现有的语音识别器主流语法可以归为三种：

- 1) EBNF(Extended Backus-Naur Form) 语法：一种传统风格的语法，包括循环（零次以上，一次以上），可选，“与”，“或”等标准的一元算子和二元算子；
- 2) JSGF(Java Speech Grammar Format)：在 EBNF 语法的基础上，新增加三种成分：(1) 适合 Java 风格的规则，包括：语法 URLs 指定规则，语法引入规则，规则的公共性/私有性指定规则等；(2) “语义标签”，主要打算

- 处理多语种的“多词同义”问题，如：“hi”，“hello”，“guten Tag”，“bon jour”；(3) 加权，实现简单的 n-gram 或“全元统计”模型。从 JSGF 控制的语音识别器获取识别结果的核心接口包括：(1) 前一选/前 N 选列表，其中每一个选择包含：(a) 语块列表；(b) 语义标签列表；(c) 与输入匹配的语法的名称；(d) 与输入匹配的公共规则（最高层符号）的名称；
- 3) V-XML 语音识别语法：是万维网联盟（W3C）为了适应“语音浏览器”发展的标准化要求而正在讨论的一种基于 XML 的 JSGF。

这三种语法都是以传统的“形式语言(Formal Language)”理论为基础的。从识别器的角度看，V-XML 语音识别语法和 JSGF 是等价的，统称为 JSGF 范式。JSGF 范式与传统的 EBNF 的本质差别在于引入了产生语义功能的标签配置，在一定程度上解决了分布式语音识别应用中多语种的“多词同义”问题。但是，即使在一个简单 的应用系统中，标签配置也存在明显的缺陷：(1) 标签的列表体现了话语中的语块顺序，但它不能对稍微复杂的“多句同义”，例如，语句“June the 1st”和“the 1st of June”产生同样的标签列表；(2) 当一个语句中多个语块与同一个的规则相匹配时，例如，“from <loc> to <loc>”，标签的列表变得异常复杂，需要进行后续处理。因此，大多数使用标签配置的 JSGF 范式的应用系统，最后都需要对语音识别器输出的标签列表进行类似传统自然语言处理中句法分析的特殊后处理。

发明内容

本发明的目的是提供一种口语识别理解的语法描述与识别分析的方法，可以良好地解决如何把识别器语法和分析器语法合二为一，如何在识别器处理具有复杂语义结构的语音输入或文字输入时，直接得到语义表达准确的识别结果等问题，便于人机口语对话系统的设计、实现与移植。

为达到上述目的，本发明的技术解决方案是提供一种口语识别理解的语法描述与识别分析的方法，该方法描述了一种形式化的任务描述规范，同时描述了一种可操作的识别分析规范。

所述的口语识别理解的语法描述与识别分析的方法，其所述的形式化的任务描述规范，描述了文本规则、句法规则、语义规则的基本规定。

所述的口语识别理解的语法描述与识别分析的方法，其所述的可操作的识别分

析规范，描述了系统任务的语法编辑、语法编译、识别分析的基本操作。

所述的口语识别理解的语法描述与识别分析的方法，其所述的文本规则，规定如下：

- 1) 文本注释：用“`/*`”和“`*/`”作为文本注释的起始和结束。在对文档的后续处理中，免除对“`/*`”和“`*/`”之间所有内容的处理；
- 2) 字符串：以双引号界定；
- 3) 空格：在语义句法文件中，除了文本注释、字符串、语义注解以外的地方，具有对文字的定界作用。

所述的口语识别理解的语法描述与识别分析的方法，其所述的句法规则，规定如下：

- 1) 规则定义：“文字符号 = 表达式；”。等号左边的文字符号，是规则名称。等号与分号之间的表达式，是规则的语义句法的具体表示；
- 2) 规则展开：从等号开始，到分号结束，按照从左到右的顺序对表达式进行展开；
- 3) 规则引用：在规则名称前冠以`$`。例如，`Rule`是一条规则，`$Rule`表示对规则`Rule`的引用；
- 4) 句法符号和文本定界符号：(假定 A 和 B 表示两个简单表达式)：
 - a) 圆括号 ‘(’ 和 ‘)’，如`(A B)`，表示对规则的封装，圆括号内部的表达式被作为一个整体单元对待；
 - b) 竖杠 ‘|’，如`A|B`，表示匹配 A 或匹配 B；
 - c) 空格 ‘ ’，如`A B`，表示先匹配 A，随后匹配 B，优先级比 ‘|’ 高：`A B | C D` 等同于`(A B) | (C D)`；
 - d) 减号 ‘-’，如`A-B`，表示匹配 A，但不匹配 B；
 - e) 加号 ‘+’，如`A+`，表示对 A 匹配一次以上，优先级比 ‘|’ 高：`A+ | B+` 等同于`(A+) | (B+)`；
 - f) 星号 ‘*’，如`A*`，表示对 A 匹配零次以上，优先级比 ‘|’ 高：`A* | B*` 等同于`(A*) | (B*)`；
 - g) 前斜线 ‘/’ 和 ‘/’，如`/0.6/A | /0.5/B`，表示按 0.6 概率匹配 A 或 0.4 概率匹配 B；
 - h) 方括号 ‘[’ 和 ‘]’，如`[A]`，表示可以匹配 A，也可以不匹配 A。

所述的口语识别理解的语法描述与识别分析的方法，其所述的语义规则，包括语义标签、语义变量、语义注解，语义嵌套，规定如下：

- 1) 语义标签：由计算机可显示的符号组成，‘%’除外；
- 2) 语义变量，以‘%’开始，后接大小写字母符号（包括下连线‘_’和中连线‘-’），以空格和除下连线‘_’和中连线‘-’以外的其它标点符号结束；
- 3) 语义注解：由语义标签和语义变量串接组成，用花括号‘{’和‘}’作为语义注解的起始和结束。语义注解不参与识别器的匹配运算；
- 4) 语义嵌套：当前规则的语义注解的值，取自所引用规则的语义注解的值。

所述的口语识别理解的语法描述与识别分析的方法，其所述的语法编辑，遵循所述的形式化的任务描述规范，采用纯文本的形式，针对系统设计的任务，编写语义句法规则文件。

所述的口语识别理解的语法描述与识别分析的方法，其所述的语法编译，把采用纯文本编写的语义句法规则文件转换成为适合计算机进行动态优化匹配的语义句法网络。该语义句法网络，由连接网络起始点和终结点的一条以上相互交叉或互不相交的路径构成。从起始点到终结点的每一条网络路径，对应语义句法规则文件中的一条识别匹配的规则。每一条网络路径包含一个以上的节点和两条以上的连接节点的弧。每个节点上存储语块及其声学模型用于进行识别匹配；对于要求进行语义注解的节点，另外存储语义标签和语义变量。连接两个节点的弧还要存储权值，用于对节点识别匹配的值进行加权。

所述的口语识别理解的语法描述与识别分析的方法，其所述的识别分析，在所述的语义句法网络中，对输入计算机的语音或文字进行动态规划匹配，找到最优匹配的路径。位于最优匹配路径上的节点如果要求语义注解，先将该节点最佳匹配的语块赋给语义变量，然后按照该最优路径上节点的先后顺序，把各个节点上的语义标签和语义变量的值，串接成为一个输出字符串，作为识别分析的输出。

所述的口语识别理解的语法描述与识别分析的方法，其任务描述规范和识别分析规范与任务无关，所以容易移植。

本发明的口语识别理解的语法描述与识别分析的方法，把语义句法描述和识别理解操作结合成一个整体，在识别器处理具有复杂语义结构的语音输入和文字输入的同时，直接得到语义表达准确的识别结果，便于口语识别理解系统的设计、实现和移植。

具体实施方式

本发明方法提出了口语识别理解的语法描述与识别分析的方法。为了保证准确地实施，本方法设计了形式化的任务描述规范和可操作的识别分析规范。

在语法描述规范中，由于既要进行句法运算，又要进行语义注解，总共设置了文本规则、句法规则、语义规则 3 套基本规定。考虑到语义注解的特殊要求，语义标签采用除了'%'以外的其它所有计算机可以显示的符号。

在识别分析规范中，把识别理解分为语法编辑、语法编译、识别分析 3 个基本步骤。考虑到识别与理解使用同一个语义句法规则文件，语法编辑在对计算机可能接受的语音输入或文字输入的语块的逻辑关系进行描述的同时，对需要语义注解的语块进行语义标签配置和语义变量配置。对于复杂语义结构的语音输入和文字输入，采用语义嵌套的方式。

语法编译对语法编辑得到的语义句法文件进行编译，产生语义句法网络，作为计算机进行识别匹配的数据结构。考虑到语义句法网络中处于相同位置上的多个语块在系统输入时的出现概率的差异，通过设立不同的权值进行调节。

识别分析在编译产生的语义句法网络中进行。对于计算机的语音输入或文字输入，在语义句法网络的起始点和终结点之间的所有网络路径中，用动态规划的方法进行匹配，找出最优匹配的路径。对于该最优匹配路径上具有语义注解的节点，先将该节点与输入匹配的语块赋给该节点的语义变量。最后按照时间先后的顺序，将该最优匹配路径上节点的语义标签和语义变量的值，串接成字符串，作为识别分析的结果。

语法描述规范和识别分析规范作为一个整体，构成口语识别理解的一体化系统，识别器同时完成识别和理解的双重任务，更加符合口语识别理解的实际应用要求。尤其当处理具有复杂语义结构的语音输入和文字输入的时候，从识别器就直接得到语义表达准确的理解结果，便于人机口语对话系统的设计、实现与移植。

实施例

用从简单语义结构到复杂语义结构的系统输入为例，说明本发明方法的使用（假定 A 和 B 表示两个简单表达式）：

1) 语义标签注解 A{symbol}，表示如果匹配 A，则输出“symbol”，如
语法规则：\$Rule=(美国|英国|法国|德国|日本|中国) {country}；

匹配规则：\$Rule

系统输入：“欢迎到中国观光旅游”

识别输出：“country”

2) 语义变量注解 A{%symbol}，表示如果匹配 A，则先把与 A 相匹配的值赋给变
量%symbol，最后输出%symbol 的值，如

语法规则：\$Rule=(美国|英国|法国|德国|日本|中国) {%country}；

匹配规则：\$Rule

系统输入：“欢迎到中国观光旅游”

识别输出：“中国”

3) 语义标签和语义变量混合注解 A{rulename=%symbol}，表示如果匹配 A，则
先将与 A 相匹配的结果赋给变量%symbol，最后输出“rulename=“和%symbol 的值，
如

语法规则：\$Rule=(美国|英国|法国|德国|日本|中国) {country=%name}；

匹配规则：\$Rule

系统输入：“欢迎到中国观光旅游”

识别输出：“country=中国”

4) 语义标签和语义变量混合注解 A{rulename(%symbol)}，表示如果匹配 A，则
先将与 A 相匹配的结果赋给变量%symbol，最后输出“rulename(“和%symbol 的值
以及”)”，如

语法规则：\$Rule=(美国|英国|法国|德国|日本|中国) {country(%name)}；

匹配规则：\$Rule

系统输入：“欢迎到中国观光旅游”

识别输出：“country(中国)”

5) 语义嵌套注解（当前规则的语义注解取自所引用规则的语义注解），如

语法规则

```

$InitQuerry=(想知道|请问) ;
$HaveOrNot=(有没有|剩没剩) ;
$loc0 = (北京|上海|广州|大连) {city=%CityName} ;
$loc = (北京|上海|广州|大连) {%CityName} ;
$Utterance_1=($HaveOrNot [从] $loc0 {%From} (到|去) $loc0
{%To}) {rule_1:depart=[%From];arrive=[%To]} ;
$Utterance_2=($InitQuerry [从] $loc {%From} (到|去) $loc
{%To} $HaveOrNot) {rule_2(\"%From\", \"%To\")} ;

```

匹配规则

(\$Utterance_1 | \$Utterance_2)

系统输入：有没有北京去大连的票

识别输出（\$Utterance_1 是与语音输入最佳匹配的路径）：

“rule_1: depart=[city=北京], arrive=[city=大连]”

系统输入：请问从北京去大连的票有没有

识别输出（\$Utterance_2 是与语音输入最佳匹配的路径）：

“rule_2(\"北京\", \"大连\")”