

基于维语中心语驱动句法分析器的结构及其实现

玉素甫·艾白都拉
(新疆哈密电视大学)

吾守尔·斯拉木
(新疆大学计算机科学系)

摘要: 本文论述了中心语驱动文法的维语句法分析器结构, 以及维语词法、句法自动分析试验系统的实现方法。根据维语句法特点修改和扩充HPSG文法并将扩充后的文法应用到维语句法描述和分析中。之中详细介绍了维语的句法(The Syntax Of Uighur Structure)分析方法, 中心语驱动文法的维语句法分析器结构。例如: 机器字典的构造、维语句法树的定义、句法范畴属性和规则、语义层次结构的定义, 上下文相关规则处理、分析器分析算法的定义、以及分析器的性能和几种句型的描述和分析过程。

关键字: HPSG 中心语驱动句法分析 维语。

The Design and Implementation of the HPSG-based Syntax Analyser of Uighur

Yushup. Abaidula Ushur. Slam
(Xinjiang Hami TV University) (Department of computer Science Xinjiang University)

Abstract: This talk introduce the Head-Driven Phrase Structure Grammar(HPSG) and the imlemention of the parse and syntax auto-analysing of Uighur. The modified and extended HPSG is discussed depending on the syntactic features of Uighur, and is applied to typical sentence model (assertive sentence) of Uighur in this paper. At the same time, the definition of analysislogrizim, syntax, semantic analysis and machine dictionary structure are introduced. As an example, some sentence models are described and analysed.

key words : HPSG, syntactic Analysis,Uighur.

一、 引言

本文国家自然科学基金项目维语自然语言理解课题研究中的成果。在中心语驱动文法的维语句法分析器的设计和实现中, 根据维语句法特点修改和扩充HPSG文法, 扩充后的文法应用到维语中心语驱动句法分析器中的典型而常用句型分析方法中去。研制了分析算法的定义、语义分析、上下文相关处理、机器字典的构造和几种句型的分析过程等。在分

析器中引入HPSG理论,是因HPSG的基于复杂特征集合—运算和句法局部树的合格性条件检查方法,对描述维语句法范畴、句法规则是合适的,严谨的,具有操作性和有效性。它能对像HPSG这样的属性文法进行高效的句法分析。从已实现的系统来看,这些方法在维语句法分析中产生了较好的效果,达到了高效的,分析正确的预期研究目的。

二、维语句法树、范畴属性、规则和语义层次结构

1、维语句法树

根据现代维语句子结构和特点可以提出维语陈述句的(人脑造句模型)完全语法树(Uighur integrated syntax tree),并且可以推广到维语其它句型中。维语陈述句的完全语法树的定义〔1〕中。

2、句法范畴属性、规则和语义层次结构

根据现代维语特点定义句法范畴属性,分析维语句子的构造规则集是进行句法分析的前提。对于自然语言描述和分析依靠句法、语义等知识将进行输入句子“分解”,从而确定输入句子的句法结构。其结构能说明该句子的词与词(词组与词组)之间的关系。自然语言分析目的是把自然语言句法的合法性予以确定,并且建立它的合法句法结构。笔者考虑用基于合一文法来描述和分析维语句子,是因为这种文法突出了自然语言句子的核心成分,以核心成分作为句子中心,使各个成分之间的信息传递匹配。保证了信息的传递,区分成分匹配的合法性。其中HPSG是很有代表性的一种基于合一的属性文法。这种文法不能完全适合描述和分析维语句子。笔者对HPSG文法进行扩充和修改后维语的分析。定义了句法结构、句法规则、语义层次结构和机器字典的构造〔1〕。

三、中心语驱动文法维语句法分析器的特点、结构和算法

1、中心语驱动文法维语句法分析器的特点

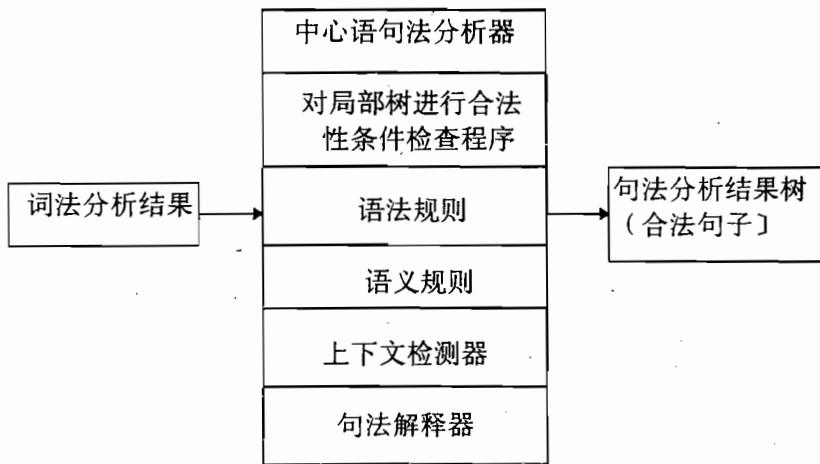
中心语驱动文法维语句法分析器能反映心理实现的句子分析模式;其程序是为后进先出自动机(Push down automation)设计的,并处于心理语言学家之手,担与其计算模式相比,这种分析程序更好的反映了人在分析句子时的心理过程。这是一个按照现代语言学X价标理论设计的句法分析器。中心语驱动文法维语句法分析器的主要特点是:

①、从顶向下分析(top-down parsing)和从底向上分析(bottom-up parsing)相结合,从而体现了人在分析句子时即根据规则推导预测又根据句子中每一个词分析归纳(both hypothesis-driven and data-driven),这是一个心理现象。

②、能够解释为什么某些句子结构(自嵌入[self-embedding或center-embedding]结构)会造成人分析时的困难。

③、记忆负荷方面也接近人的心理现象。

2、中心语驱动文法的维语句法分析器结构



中心语驱动文法的维语句法分析器结构图

3、中心语驱动文法的维语句法分析器的算法

从词法分析器传送来的信息，中心语句法分析器用对局部树进行合法性条件检查作为分析程序算法。在分析程序中，主要有四大步，首先进行查询句子中的谓语；第二步，进行查询句子中的主语，并使用合法性条件检查程序检查是否存在（受控者特征值是否控制者一致）合法主语；第三步，在对句子的连贯（局部树）成分进行合法性条件检查。首先，检查局部树父，子范畴属性值与语法规则的主范畴和子范畴对应范畴属性值是否一致；然后对局部树依次进行检查（受控者特征值是否控制者一致）规则匹配，以保证所要求的属性值同现；依据语法规则中所指定的需要进行合一的范畴属性，检查局部树中对应范畴属性间是否满足合一条件，包括父、子范畴属性之间或子、子范畴之间的合一。只要一处检查不合法，这个棵树是非法的，结束分析这棵树。否则这棵树是合法的，则被加入到句法分析树的相应位置上，称为句法树的一部分。第四步，句子的主语部分（主语局部树）和谓语部分（谓语局部树）连接在一起形成一个完整的句子（合法的句法树）。

四、中心语驱动文法维语句法分析器的性能

中心语驱动文法的维语句法分析器可以分析的短语有：名词短语，形容词短语，动名词性短语，动形容词性短语，动副词性短语。句型有：无主语，无修饰谓语成分的谓语句，无主语，有修饰谓语成分的谓语句，只有主语，谓语成分的，没有修饰主语和谓语成分的主谓句，有主语，有谓语修饰谓语成分的主谓句，有主语，谓语成分，只

有修饰主语成分的主谓句，有主语，谓语成分，又有修饰主语和谓语成分的主谓句等句型。另外又能分析部分复合句。在构造规则集的策略上，采用了即考虑常用的，通用的句型规则，又考虑了特殊的句型规则。

在研制句法分析器的过程中，遇到了一些困难，主要是怎样分析一些较特别的语言现象，通过维语语法的不断的深入学习和对语料各种现象的统计、分析，部分困难得以解决。现举例如下：

在并列主语句和并列谓语句情况下：

① ئىشچى، دېھقان، ئەسكەر ۋە زىيالىلار ئەمگە كەكە ئاكتىپ قاتناشتى.

[工人，农民，解放军战士和知识分子积极参加了劳动]。在这个句子中， ئىشچى [工人] دېھقان [农民] ئەسكەر [解放军战士] زىيالى [知识分子] 等四个名词的地位是平等的，都是第三人称，句子谓语要求也是第三人称，这是一个歧义问题。

② ئۇ كېرىلدى، ئەسنلى.

[他伸了个懒腰，打个呵欠。] 在这个句子中، كېرىلدى [伸了个懒腰] 和 ئەسنلى [打个呵欠] 等谓语动词都有逗号和句号。它们在句子中的地位是平等的，都是谓语，这也是一个歧义问题。句法规则不能解决，只能使用上下文相关处理规则得以解决这个问题。

综合起来中心语驱动文法的维语句法分析器的主要性能是：

- (1)、能分析对维语中存在的所有短语，大部分简单句和部分复合句等句型。
- (2)、句法分析结果信息丰富，以提供给后续转换使用，能满足各种需求的接口。
- (3)、规则维护方便。
- (4)、本分析器的句型覆盖率已经达到了90%以上。
- (5)、分析器每秒能分析10~15个单词。

五、结论

以上笔者讨论了描述和分析维语句子的方法，在此基础上建立了中心语驱动文法的维语句法分析器。机器上实现结果表明，本系统这方面取得了较好的效果，并且达到了预期的目的。本系统作为使研究维语自然语言理解和维汉（汉维）机器翻译系统的主要组成部分，单词频率统计、文本分析、汉维机器字典的构等语言学领域和维语计算机辅助教学（CAI）系统的研究等打下了基础。维语自然语言理解和机器翻译方面的工作是刚刚起步，涉及学科领域较广，进一步研究完善本系统需要作很多工作。

- 1、理论研究的提高和系统知识的扩展；
- 2、维语语义研究进一步深化；
- 3、机器词典的扩充和管理。
- 4、特殊语言现象的处理；
- 5、加强智能化功能，强化纠错功能等等。

参考文献

- 〔1〕。玉素甫·艾白都拉，姚天坊 “维吾尔语句法分析器的探讨”，第界中国青年计算机工作者论文集<<中国青年计算机研究新进展 '96>>，1996年10月。
- 〔2〕。现代维语（维吾尔文）哈萨克苏维埃科学院维吾尔学学部1966年阿拉木图用斯拉夫文出版，新疆教育出版社翻译出版 1985年8月。
- 〔3〕。哈密提·铁木尔，“现代维语语法”（维吾尔文）民族出版社出版 1987年6月
- 〔4〕。玉素甫·埃白都拉，吾守尔·斯拉木，姚天坊，“维吾尔语句法描述和分析方法的探讨”，<<中文信息>>学报，1996年4月。
- 〔5〕。吴蔚天，罗建林，“汉语计算语言学”，电子工业出版社，1994年7月。
- 〔6〕。Hsi-Jian Lee and Pei Rong Hsu Parsing Chinese Sentences in a Unification-Based Grammar Computer processing of Chinese & Oriental Language Vol. 5. No5. November 1991