

# 基于句类分析的功能块和事件小句标注

韦向峰

中国科学院声学研究所

wxf@mail.ioa.ac.cn

**摘要:** 本文介绍了 HNC (概念层次网络) 理论的概念基元符号、句类和语义块等概念, 说明了句类分析中的概念类别感知、特征语义块构造和假设检验等处理步骤, 在句类分析结果的基础上给出了标注汉语功能块和事件描述单元的方法, 标注结果在开放测试中分别取得了第二名和第一名的好成绩。

**关键词:** HNC; 句类; 语义块; 句类分析

## Labeling Functional Chunk and Event Sentence Based on the Analysis of Sentence Category

Xiangfeng Wei

Institute of Acoustics, Chinese Academy of Sciences

wxf@mail.ioa.ac.cn

**Abstract:** This paper introduces the symbolic system of conceptual primitives, sentence category and semantic chunk in HNC (Hierarchical Network of Concepts) theory. It shows the recognition of conceptual categories, constructing eigen chunk and hypothesis testing in the analysis of sentence category. It also proposes a method of labeling the functional chunks and event description unit based on the analysis of sentence category. The results with this method ranked first and second respectively in the open test of CIPS-ParsEval-2009.

**Keywords:** HNC (Hierarchical Network of Concepts); Sentence Category; Semantic Chunk; Analysis of Sentence Category

### 1 引言

HNC (概念层次网络) 理论[1,2]中的“语义块”与语法功能块有相似之处, “小句”则与“事件描述单元”的概念类似。因此, 我们参加了中文信息学会句法评测(CIPS-ParsEval-2009)中的任务 3 和 4, 分别是汉语功能块识别和汉语事件描述单元识别。

Abney 于 1996 年提出了组块的概念[3]。周强等引入成分组和词界块的概念, 提出了汉语组块的描述体系[4]。罗雪冰在分词和词性标注过的语料基础上, 结合 SVM、CRF、HMM 三种统计学习方法识别组块[5]。浅层语义分析中的语义角色标注采用“谓语动词-角色”的结构形式, 标注句法成分为给定谓语动词的语义角色, 一般用最大熵等统计学习方法进行标注[6,7]。HNC 认为句子由语义块构成, 句子的语义概念类别 (句类) 决定了语义块的个数和类型。从字词的概念基元符号感知概念类别, 通过构造特征语义块获得假设的句类, 依据句类知识检验假设的句类和语义块是否合理, 最后对语义块内部进行分析, 这一过程称为句类分析。

按句号等分隔的句子是“大句”, 按逗号等分隔的则是“小句”。一般而言, 小句与事件描述单元或 HNC 的句类较为接近。但在实际处理的过程中, 还需要句类分析结果和领域句类表示式的指导, 使得小句的边界识别更加精确。

本文以句类分析为基础, 识别汉语句子中的事件描述单元和功能块, 在开放测试中分别取得了第一名和第二名的好成绩。

## 2 概念基元及其符号

人类在理解语言时主要依靠概念以及概念间的关系和约束。自然语言中的文字符号只是概念的表现形式之一。为了从概念层面去描述和处理自然语言，HNC 建立了概念基元符号体系，通过概念基元及其组合描述各种各样的概念。这样，自然语言中的字、词和词组可以映射为概念符号，它们之间的概念联想脉络关系可以由概念基元符号来揭示。

概念可以分为三类：抽象概念、具体概念和两可概念。具体概念主要用于描述人（p）和物（w,jw），可通过挂靠抽象概念来描述；两可概念主要是物性（如颜色、性别等），通过具体概念符号的变化（把 w 改为 x）来表示。因此，概念基元基本上由抽象概念构成，抽象概念主要分为以下几大类：

表 1 抽象概念基元的分类

符号	描述的概念
0~5	作用、过程、转移、效应、关系、状态
71,72,73	心理、意志、行为
8	思维
a,b,d	专业活动、追求活动、理念
q6~q8	劳作、交往娱乐、想象与信念
j	基本概念（时间、空间等）
l,j,l,s	语法逻辑、基本逻辑、综合逻辑
f	语习概念（称呼、插入语等）

概念基元符号可以逐层往下延伸，表达更加细致更加具体的概念。例如“a”表示专业活动，往下延伸可以得到“a1”（政治）、“a11”（政权活动）、“a113”（最高领导人（国家或地方政府）更迭）。符号的延伸具有层次性，分为高、中、底三层；高层符号必须要有，高层可以直接进入底层（即可以没有中层符号），进入底层后又可以进入中层或底层，形成中底层的多重符号表示。中层符号（c,d,e 等）用于表示概念之间的对偶、对比和包含关系，底层符号（t,\k,i 等）用于表示概念之间的交织性、并列性和定向性。

词语表达的意义或概念可以映射为概念基元符号，词语之间的关系可以通过概念基元符号显式的计算出来。例如，“法院（pea56）”和“宣判（vc232|ra56）”都属于审讯与判决领域（a56），其中“pe”表示“组织”，“v”表示动态概念，“c”表示社会性概念，“|”表示位于其右边的概念是左边概念的作用内容，“r”表示效应概念。在判断两个概念是否相关或两个概念的相似程度时，只需计算概念基元符号的相近程度即可。

对于常用的字和词语，已经建立了它们的常用义项的概念基元符号表示，并存储在字知识库和词知识库中。字知识库和词知识库的数据项大致相同，包括字（词）形、音调、义项数、义项号、义况、独立性指数、组合代码、联想方向、义类、概念类别、HNC 概念符号、句类代码等等。目前，字知识库的义项数为 2 千多个，词知识库的义项数为 4 万多个。字词知识库是进行句子分析或句类分析的基础。

## 3 句类与句类分析

句类指的是语句的语义类别。句类由语义块构成，语义块可以是词、短语或句子。句类及句类知识指示了构成句类的语义块、构成语义块的概念以及语义块之间的语义关系。句类分析的目的就是分析出自然语言语句的句类、语义块以及语义块的内部构

成，揭示出句子的语义构成和概念联想脉络。

### 3.1 句类和语义块

句类是语句的语义类别，主要来源于概念基元中的作用效应链（0~5），分为广义作用句和广义效应句。在自然语言中，广义作用句允许出现规范格式（如“把”字句），广义效应句则不允许。广义作用句包括作用句、转移句、关系句和判断句，广义效应句包括效应句、过程句、状态句和基本判断句，即细分为八大句类。根据语义块个数和描述对象的不同，这八大句类又细分为 57 组基本句类。57 组基本句类可以两两互相混合形成混合句类，共 3192 组混合句类。在混合句类中，特征语义块是两类概念的混合（如作用效应概念），广义对象语义则取自基本句类中的广义对象语义块。

语义块是语句的下一级语义单位。语义块分为主语义块和辅语义块，主语义块是语句必要的、主干的成分，辅语义块是次要的、可选的成分。辅语义块主要用于描述时间、空间、条件、手段等背景信息，分为以下七类：条件 Cn、方式 Ms、工具 In、因 Pr、果 Rt、参照 Re、途径 Wy。

主语义块又分为特征语义块和广义对象语义块，特征语义块（E）表达了句类的基本语义概念，类似于句法分析中的述语块；广义对象语义块包括作用者（A）、对象（B）和内容（C），是基本语义概念的发出者、承受者和作用内容。例如“昨天我给他一本书”，整个语句的句类是一般转移句 TOJ，“昨天”是辅语义块，“我”是转移的发出者 TA，“他”是转移的承受者 TB，“一本书”则是转移的内容 TC。

句类的一般表示式如式子 1-1 所示，其中  $n=2-4$ ， $m$  的取值不定，但一般不大于 5。由于辅块弱依赖于句类，因此  $fK$  可以没有。在某些句类中，特征语义块  $EK$  也可以没有。E 为句类代码，用于表示是 57 组句类里的哪一组句类。

$$EJ=[fK_m]+GBK_1+EK+GBK_n \quad (1-1)$$

一个句类要么没有特征语义块  $EK$ ，要么只有一个  $EK$ 。即特征语义块不可能包含句子，否则该句子一定可以继续分解为特征语义块加广义对象语义块。对于广义对象语义块，可以包含句子也可以不包含句子。对于包含句子或句子变形的情况，称为句蜕（句子蜕化为语义块）。句蜕有三种类型：原型句蜕、要素句蜕和包装句蜕。原型句蜕是指语义块由一个句子充当；要素句蜕是指句子经过变形（如加“的”）后，其中的某个语义块或其要素仍然是上一级句子的广义对象语义块要素；包装句蜕是指一个句子加上前包装成分或后包装成分，然后作为上一级句子的广义对象语义块。在某些特定的句类中，发生在  $GBK_n$  中的原型句蜕称为块扩。块扩的特点是：当句类确定时， $GBK_n$  中的某个语义块必然由一个句子构成。由于句蜕或块扩中的语义块又可以出现句蜕或块扩，从而形成多级嵌套的块扩或句蜕现象，这是语义块构成分析的难点之一。

### 3.2 句类分析的基本步骤

句类分析的步骤可以简单地概括为“分词分段-感知概念类别-构造  $EK$ -假设检验-语义块内部分析”。在这一过程中，“分段”是指把具有模糊（歧义）切分的词合成一个字段，保留所有可能的切分，其模糊的消解留到假设检验阶段。由于评测任务 3 和任务 4 的输入都是分好词和标好词性的句子，因此“分词分段”的步骤可以跳过，直接进入后面的步骤。

感知概念类别是指获得词语的语义概念类别，特别是有助于语义块切分的“ $l$ ”和“ $v$ ”类概念。“ $EK$ ”代表特征语义块，主要根据“ $v$ ”概念及其前后词语构造。有了  $l$  和  $EK$ ，就可以假设出辅语义块、广义对象语义块，根据假设的语义块和“ $v$ ”概念的句类代码及句类知识，可以检验语句格式、语义块个数、语义块要素概念的合理性，

获得句子的句类和语义块划分。对于不含“v”概念的句子，按无特征语义块句类进行假设检验。语义块内部分析主要是指语义块内部含“v”概念时，按照块扩、原型句蜕、要素句蜕和包装句蜕进行处理，需要重新构造块扩或句蜕的EK，然后进行假设检验。语义块的内部分析实际上又是一个句子的分析处理，形成了一个递归处理过程。

### 3.3 感知概念类别

概念类别主要来源于五元组 (v,g,u,z,r)、概念基元类别、特征语义块构成成分和专有名词等，其目的是服务于特征语义块、广义对象语义块和辅语义块的感知。表 2 列出了句类分析中常用的概念类别符号。

表 2 概念类别符号

类别符号	意义或用途
l0	主语义块的指示标志
l1,lq,h	辅块的指示标志
p,fp	人、专有名词
v	动态概念（特征语义块的核心）
vv,qv,hv,QE,QEu,Eu	特征语义块的上下装、修饰
j1~j3	时间、空间、数量概念
s	综合类概念
f	语习概念（称呼、插入语等）

对于一般字词的概念类别符号，由于字词知识库中都标有字词的概念类别符号，可通过简单的转化规则得到。例如“l01”（GBK1 的指示标志）被感知为“l0”，“l6”（时态概念）被感知为“QE”。对于汉语中常用的一些特殊单字，如“把、被、向、对、在、的、了”等，需要进行特殊处理和单独感知。对于含“v”概念的字词，鉴于其对于特征语义块构造的重要性，也需要特殊处理和单独感知。

概念类别感知的难点是一个字或词有多个概念类别时的处理。这里有两个处理策略，一是根据词语的上下文，运用优先规则和排除规则等获得最优的概念类别并去除不合理的概念类别，如非 v 类概念前的 QE 或 Eu 可以去除；二是保留次优概念类别，当最优的概念类别不能通过假设检验时，需要回溯到这里重新获得概念类别。

### 3.4 构造特征语义块

特征语义块可以简单地表示为  $EK=QE+Ek+HE$ ，QE 表示特征语义块的上装修饰部分，HE 表示特征语义块的下装修饰部分，Ek 为特征语义块核心。Ek 主要有以下五种构成类型：(1) $Ek=E$ ；(2) $Ek=EQ+EH$ ；(3) $Ek=EQ+E$ ；(4) $Ek=E+EH$ ；(5) $Ek=EQ+E+EH$ 。(1)表示 Ek 的简单构成，即仅有一个 v 概念词。(2)表示 Ek 的并列构成，即 EQ 和 EH 都是 v 概念词。(3)表示高低概念搭配或“vv+v”构成。(4)表示动静概念搭配。(5)表示 Ek 的高低动静概念搭配，是高低概念搭配和动静概念搭配的交叉组合。对于 Ek 简单构成的第(1)种情况，以感知到的 v 概念为中心，向前寻找 qv、Eu、QE 类概念，合并为 EK；向后寻找 hv，合并为 EK。对于复合构成的第(2)~(5)种情况，需要根据感知到的 vv、EQ 和 EH 的字词与 v 概念字词进行合并处理，得到特征语义块 EK。

汉语句子常常出现多个动词，即多个 v 概念，这时候构造出的特征语义块 EK 就有多个。而句子多数情况下只有一个全局特征语义块（类似于句子第一级的述语块），因此必须对多个 EK 进行排序，这也有利于后面的假设检验步骤给出准确的句类结果。EK 排序的方法是根据 EK 前后是否有特殊字“的”或“是”、是否有 l0 或 l1 概念，EK 中的 v 概念字词是否为单字、是否为纯 v、是否有上下装、是否含块扩句类代码等计算

出权重得分，分值越小说明作为全局特征语义块的可能性越大，排序位置越靠前。分值相同的 EK 按在句子中的出现顺序排序。

### 3.5 假设检验

句子中的 EK 形成了对句子的一个初步切分，即可以是“GBK1+EK+GBK2”、“EK+GBK2”、“GBK1+EK”等情况。句子中的辅块 fK 通常带有指示标志如“在...中”、“根据...”、“...以来”等 I1 类概念，而 EK 和广义对象语义块指示标志 (I01 或 I02) 也是辅块的边界，并且现代汉语中辅块只出现在 EK 之前。不带指示标志的辅块一般是时间类概念，识别时间类概念则较为简单。因此，根据 EK、I0 和 I1 类概念以及辅块的特点，可以给出句子的句类表示式假设： $EJ=[fK_m]+[GBK_1]+EK+[GBK_n]$ 。根据自然语言空间的句子实际情况，GBK1 可能不出现，m 和 n 的个数也各不相同，各个 GBK 的位置也不一样。实际句子中 GBK<sub>n</sub> 的出现位置以及带指示标志 I0 的不同构成了不同的语句格式，如“GBK+I02+GBK+EK+GBK”可以假设为规范格式“!1111”。而根据 EK 的 v 概念字词，可以从字词知识库中获得具体的句类代码，进一步可以得到实际句子的句类假设，如“!1111T0J=TA+I02+TB+EK+TC”。

假设的句类及其表示式需要结合句类知识、句子自然语言空间的内容进行检验。检验包括格式检验、语义块个数检验、语义块要素概念检验。在语句格式检验中，如果句类代码所对应的句类为广义效应句，那么语句格式假设为规范格式时不能通过检验；如果字词知识库中约定不能出现某种语句格式，那么该语句格式假设也不能通过检验。在语义块个数检验中，如果假设的句类表示式的语义块个数与自然语言空间中句子的语义块个数不一致，那么不能通过检验。语义块要素检验包括 E-GBK 检验和 GBK1-GBK2 检验，E-GBK 检验是指在基本句类知识和字词知识库中给出了特征语义块 EK 对于 GBK<sub>n</sub> 的预期概念约束，不满足这些概念约束的不能通过检验；GBK1-GBK2 检验是指某些句类中两个 GBK 的概念必须有足够的对仗性，否则不能通过检验。简单的说，要素检验就是把句子中语义块要素（字词）的概念与知识库中的约定概念进行相似性匹配，匹配成功则通过检验，否则不能通过检验。

如果所有的假设都不能通过检验，或者句子没有特征语义块，那么需要进入无特征语义块句类的假设检验。简明状态句是其中的一种，其句类表示式为  $S04J=SB+SC$ ，特点是 SC 为 u 类概念或基本概念短语，像“今天的天气很好”、“一斤白菜五毛钱”都属于这种句类。对于通过检验的句类表示式，还需要对句类表示式中的语义块的内部构成进行分析。

### 3.6 语义块的构成分析

语义块的内部构成有简单构成和复杂构成。简单构成是指语义块不包含 v 概念，只由一个词语或几个词语组合构成。复杂构成是指语义块包含 v 概念，往往是一个句子或句子的变形。这种复杂构成在汉语句子中经常出现，形成“多动词”（句子中出现多个 v 概念字词）现象。多动词可以简化为两动词关系的处理，两动词关系主要有以下三种情况：（1）E<sub>p</sub>-E<sub>r</sub>；（2）E<sub>g</sub>-E<sub>l</sub>；（3）E<sub>1</sub>-E<sub>2</sub>。第一种情况称为块扩，即动词 E<sub>r</sub> 构成了一个句子，而是 E<sub>p</sub> 引导的句子的一个语义块，例如“我希望他能来”。第二种情况称为句蜕，分为原型句蜕、要素句蜕和包装句蜕（参见 3.1 小节），即动词 E<sub>g</sub> 是全局特征语义块而动词 E<sub>l</sub> 是局部（块内）的特征语义块。第三种情况即连动句，两个动词 E<sub>1</sub> 和 E<sub>2</sub> 分别引导两个句子，形成两个句类，例如“邓小平前往广西领导起义”。

块扩是特定句类（共 10 类）的特定语义块中出现句子的情况，具有明确的句子预期知识，因此块扩语义块可以按句子再次进行假设检验。原型句蜕与块扩类似，由句

子原形构成，因此也可以直接按句子进行假设检验。要素包装通常是句子的变形，一般都是在句子某个语义块如 E1 前面加“的”，因此去掉“的”后可按句子进行假设检验。包装句蜕分为原型包装句蜕和要素包装句蜕，其中的包装品是一些特殊概念如“r,g,s,djl,x,jw, jgw”等，如果语义块末尾为这些特殊概念，则把包装品去掉再按原型句蜕或要素句蜕进行假设检验。总之，对于含有 v 概念的语义块，如果是块扩则按块扩处理，否则若有包装品则按包装句蜕处理，否则根据句子形式按原型句蜕或要素句蜕处理，最后按照两个句类（E1-E2）的情况处理。

在语义块的构成分析中，首先需要构造出一个新句子的句类假设，即确定新句子的特征语义块 EK、句类代码、语句格式等，然后按原来的检验方法对新句子又进行假设检验，对于新句子的语义块内部仍然如此处理，在程序中是一个递归的处理过程。

### 3.7 句类分析的特点

句类分析从字词的概念类别入手，以字词知识库和句类知识库为基础，由 v 概念构造特征语义块，由 l 概念假设广义对象语义块，通过特征语义块假设句子句类，然后根据知识库的概念和句类知识进行检验，给出符合检验的句子句类、语义块和语义块的内部构成。句类分析具有以下特点：(1) 从概念切入，直接进入句子的语义切分；(2) 以语义块为处理单位，向上组成句子，向下形成嵌套结构；(3) 以特征语义块为中心，突破了述语动词的限制，形成了多动词处理策略；(4) 以句类和句类知识为指导，运用假设检验机制，消解处理过程中的模糊；(5) 给出了句子类型和语句格式的完备描述，有利于句子的理解和分析。

## 4 功能块标注

在评测任务 3（汉语功能块分析）中，给出的功能标记集合为：S—主语块、D—状语块、P—述语块、O—宾语块、C—补语块、J—兼语块、A—定语块、H—中心块、T—独立块、X—其他特殊块。输入为：经过正确词语切分和词性标注处理汉语句子（Gold-standard 数据）；输出为：不同层次的功能块组合形成的线性序列。

通过研究句类分析与任务 3 的输入输出，确定功能块的标注过程如下：在词语切分和词性标注的基础上进行概念类别感知，构造特征语义块并记录内部结构，通过假设检验得到句类表示式和语义块内部构成结果，根据句类分析结果转换为功能块标注序列。具体的转换标准如表 3 所示。

表 3 句类分析结果转换为功能块

句类分析结果	功能块	备注
GBK1	S	第一个广义对象语义块
Ek	P	特征语义块的核心
HE	P	特征语义块的下装
GBK2	O	第二个广义对象语义块
GBK3	O	第三个广义对象语义块
QE	D	特征语义块的上装
fK	D	辅块
HBZP	H	后包装品

从表 3 可以看出，句类分析结果经过简单的成分转换，已经可以标注出主要的主语块(S)、述语块(P)、宾语块(O)、状语块(D)和中心块(H)。对于其它的功能块，在分析训练语料和总结规则的基础上分别单独处理，各个击破。对于兼语块(J)，主要是寻找

词性为“/vJY”的特征语义块，将其后的广义对象语义块 GBK 标注为兼语块(J)。对于独立块(T)，主要是通过统计训练语料得到独立块的词或短语集合。以此为基础，结合序号、标号、引句、称呼等特点标注独立块(T)。对于其他特殊块(X)，主要是经过句类分析后不能得到句类结果的多个词组成的短语，一般都不以标点符号结尾。对于补语块(C)，主要是寻找特征语义块后的词性为“/vB”的词语，以及特征语义块的由“以、来、于”等引导的语义块。对于定语块(A)，没有找到很好的规律，只是把包装品中的数量概念作为定语块(A)。

从评测结果(表4)来看，独立块和述语块的标注效果比较令人满意。定语块的召回率则极低，这是因为没有得到识别定语的覆盖性规则。尽管如此，在总成绩上功能块标注结果还是取得了开放测试第二名的好成绩。

表4 任务3的评测结果

Label	Precision(%)	Recall(%)	F1-Measure(%)
X	93.79	90.67	92.20
P	80.28	85.20	82.67
D	76.54	82.49	79.41
S	77.03	80.84	78.89
J	71.20	76.72	73.86
O	70.28	75.98	73.02
C	80.00	43.48	56.34
H	58.05	47.94	52.51
T	56.35	27.84	37.27
A	82.46	9.02	16.26
<b>Avg.</b>	<b>75.94</b>	<b>78.24</b>	<b>77.07</b>

## 5 事件描述单元标注

在评测任务4(汉语事件描述单元识别)中，输入是经过正确切分和词性标注处理的完整汉语句子，以句号、问号和叹号等分隔，而输出是句子中各个事件描述单元的左右边界。即把句号、问号和叹号等分隔的句子划分为更小单位的句子。

在概念层次网络理论中，以句号、问号和叹号等分隔的句子称为大句，以逗号、分号、冒号、破折号等分隔的句子称为小句，一个小句必然对应于一个句类。显然，由于逗号的多功能性，被逗号分隔的语串不一定是小句，还可以是：(1)主块；(2)辅块；(3)句间逻辑概念(lb)；(4)语习类概念；(5)语义块的一部分；(6)其他。大句是句群的最小单位，句群的三要素是领域、情景和背景。句群的领域可以通过词语或短语的概念基元符号激活，用句类分析可以得到词语的语义角色，根据词语的语义角色、频次和位置信息可以确定句子的领域。句群领域确定之后，就可以得到句群的领域句类表示式，它通常由多个句类组合叠加构成。句群的领域句类表示式(包括情景和背景信息)又可以指导小句句类的判定和分析。

在处理过程中，首先把大句按逗号、分号、冒号、破折号等切分为“小句”，对“小句”进行句类分析可以得到如下结果：(1)完整的句类表示式；(2)省略语义块的句类表示式；(3)辅块；(4)语习概念；(5)句间逻辑概念；(6)其他。由于第(1)和(2)种结果包含一个句类，可以作为一个事件描述单元；而后面四种结果则需要合并到下一个事件描述单元中。通过这种处理方式，事件描述单元识别的正确率达到了75.37，召回率达到77.82，F1值为76.58。

经过对错误结果的分析,发现主要错误原因之一是在前面的第(2)种结果中,省略掉的语义块可能是前面或后面的小句,这在块扩句、是否判断句、因果果因句中出现较多。因此,在句群领域分析的基础上,重点改进了这些类型句子的处理,结果正确率提高到 79.80,召回率达到 81.91,F1 值为 80.84,在开放测试结果中取得了第一名的好成绩。

## 6 结论

汉语功能块分析是汉语理解分析中的任务之一。除了运用分类器和统计学习方法之外,也可以根据词法、句法和语义的分析构造得到功能块。本文在 HNC 概念基元符号的基础上,通过感知字词的概念类别,构造出句子的特征语义块、广义对象语义块和辅语义块,在句类知识和句类表示式的基础上检验假设的语义块,然后对语义块的内部进行逐级的句类分析,最后把句类的结果转换为功能块标注结果。测试数据结果表明,这一方法并不亚于统计学习方法。如果改进不基于句类分析结果的其他类型功能块的识别性能,则测试结果还会大幅度提升。

事件描述单元类似 HNC 中的小句处理,结合句群大句的处理结果可以给出小句的更加精确的边界,同时小句和非小句结构的合并也是影响事件描述单元边界识别结果的重要因素。本文在句类分析的基础上,结合领域句类表示式的知识,标注事件小句的边界,取得了开放测试第一名的好成绩。

## 7 致谢

本文得到了国家 973 项目“自然语言理解的交互引擎研究”(2004CB318104)、中科院声学所知识创新工程项目“句群理解处理理论及其应用”(O654091431)、中国科学院声学研究所“所长择优基金”(GS13SJJ04)、中国科学院青年人才领域前沿项目(O754021432)的资助,特此致谢。

## 参考文献

- [1] 黄曾阳. HNC(概念层次网络)理论—计算机理解自然语言的新思路[M]. 清华大学出版社, 1998.
- [2] 黄曾阳. 语言概念空间的基本定理和数学物理表示式[M]. 北京: 海洋出版社, 2004.
- [3] 向毅. 基于 Abney 的组块识别方法的研究[J]. 重庆科技学院学报, 2008,10(4):111-113,146.
- [4] 周强,孙茂松,黄昌宁. 汉语句子的组块分析体系[J]. 计算机学报, 1999,22(11):1158-1165.
- [5] 罗雪冰,黄德根. 汉语组块识别的研究[D]. 大连: 大连理工大学硕士学位论文, 2007.
- [6] 刘怀军,车万翔,刘挺. 中文语义角色标注的特征工程[J]. 中文信息学报, 2007,21(1):79-84.
- [7] 丁伟伟,常宝宝. 基于最大熵原则的汉语语义角色分类[J]. 中文信息学报, 2008,22(6):20-26,68.
- [8] 苗传江. HNC(概念层次网络)理论导论[M]. 清华大学出版社, 2005.
- [9] 晋耀红. HNC(概念层次网络)语言理解技术及其应用[M]. 科学出版社, 2006.
- [10] 韦向峰. 基于 HNC 理论的扩展句类分析平台研究[D]. 北京:中国科学院声学研究所博士学位论文, 2005.