

# 国内外隐喻知识库建设综述<sup>\*</sup>

张霄军<sup>1, 2a</sup>, 曲维光<sup>2b, 3</sup>

(1. 陕西师范大学 外国语学院, 西安 710062; 2. 南京师范大学 a. 文学院; b. 数学与计算机学院, 南京 210097; 3. 北京大学 计算语言学研究所, 北京 100081)

摘 要: 就国外的几个隐喻知识库, 包括 Master Metaphor List、Sense-frame、MetaBank、Metalude、Hamburg Metaphor Database、ATT-Meta 以及国内的隐喻句库的建设进行综述, 并对上述隐喻知识库作了简单评述。最后提出了一种建立面向计算的大规模汉语隐喻知识库构建的设想。

关键词: 隐喻; 知识库; 自然语言理解

中图分类号: TP391 文献标志码: A 文章编号: 1001-3695(2008)09-2561-05

## Introduction of metaphor corpus

ZHANG Xiao-jun<sup>1, 2a</sup>, QU Wei-guang<sup>2b, 3</sup>

(1. School of Foreign Language, Shaanxi Normal University, Xi'an 710062, China; 2a. School of Art & Literature, b. School of Mathematics & Computer, Nanjing Normal University, Nanjing 210097, China; 3. Institute of Computational Linguistics, Peking University, Beijing 100081, China)

**Abstract:** This paper introduced and reviewed six foreign language databases, it included Master Metaphor List, Sense-frame, MetaBank, Metalude, Hamburg Metaphor Database, ATT-Meta and one Chinese metaphor database. Finally, it designed a large-scale Chinese metaphor knowledge database oriented to computation.

**Key words:** metaphor; knowledge database; natural language understanding

### 0 引言

隐喻理解的计算模型离不开隐喻知识库。这是由隐喻的认知性决定的。隐喻涉及本体概念与喻体概念之间的对比, 因此隐喻计算模型的前提是要求系统具有概念描述和推理能力<sup>[1]</sup>。随着自然语言理解从字面义(literal meaning)到非字面义(non-literal meaning)研究的转变, 隐喻作为非字面义的典型, 越来越受到计算语言学家、认知语言学家和心理学家的高度重视, 隐喻计算也成为自然语言处理的一个热点问题。国内对隐喻计算的研究越来越深入, 主要集中在对隐喻计算的模型方法的介绍和研究上<sup>[2, 3]</sup>, 但对隐喻知识库建设的讨论不够深刻。目前, 应用于隐喻计算的模型方法主要有以下四种:

- a) 基于优先语义的方法。这种方法认为理解隐喻义的方法不同于理解正常字面义的方法, 突出语义知识的描写而不强调隐喻知识在隐喻理解中的应用。当句子中字面意义出现冲突(即选择限制发生冲突)时, 就需要进行隐喻处理。比较有代表性的系统有 Fass<sup>[4]</sup>提出的可以处理隐喻、转喻、字面义反常表达的隐喻理解模型 Met5 系统。
- b) 基于知识表示的方法。与基于优先语义的方法不同, 该方法强调隐喻知识的表示以及隐喻知识在隐喻理解中的作用。其包括隐喻突显理论<sup>[5, 6]</sup>和基于实例的模型, 如 Martin<sup>[7]</sup>设计的用来解释常规隐喻和识别新生隐喻的 MIDAS 系统, 以

及人工神经网络模型, 如 Veale<sup>[8]</sup>的 Sapper 模型。

c) 基于类比推理和逻辑推理的方法。隐喻的理解需要人们运用常识进行推理, 因此, 许多学者采用逻辑方法进行隐喻的机器理解。典型模型包括 Stainhart<sup>[9]</sup>提出的隐喻的结构理论和隐喻逻辑论、Gentner 等人<sup>[10, 11]</sup>的结构映射引擎 SME、Holyoak 等人<sup>[12]</sup>的 ACME 隐喻分析模型。

d) 基于语料库统计机器学习的方法。随着语料库语言学的发展, 利用统计方法处理语言信息得到了广泛的关注, 出现了许多基于统计学习方法的隐喻识别模型, 如 Kintsch 等人<sup>[13, 14]</sup>的基于潜在语义分析方法的模型, 以及 Mason<sup>[15, 16]</sup>提出的基于语料库的隐喻句抽取引擎 CorMet。

从上面的计算模型来看, 隐喻表达往往涉及到概念的推理关系, 因此隐喻表达的识别、解释、提取往往需要富含隐喻理解信息的隐喻知识库的支撑。具有代表性的国外隐喻库有 Master Metaphor List、Sense-frame、MetaBank、Metalude、Hamburg Metaphor Database、ATT-Meta。国内的隐喻库建设起步较晚, 目前公布的只有厦门大学的汉语隐喻标注句库。

### 1 国外隐喻知识库建设综述

#### 1.1 Master Metaphor List(MML)

MML(详见 <http://cogsci.berkeley.edu/lakoff/>)是加利福尼亚大学 Berkeley 分校 George Lakoff 等人搜集的英语常规隐

收稿日期: 2007-11-19; 修回日期: 2008-02-19 基金项目: 国家自然科学基金资助项目(60773173); 国家哲学社会科学基金资助项目(07BYY050); 国家博士后基金资助项目(20060400027); 江苏省社科基金资助项目(06JSBY001)

作者简介: 张霄军(1978-), 男, 陕西岐山人, 博士, 主要研究方向为计算语言学(andy\_zxj@126.com); 曲维光(1965-), 男, 副教授, 博士, 主要研究方向为隐喻理解与计算。

喻表达的一个在线知识库;他们从出版的隐喻文献、加利福尼亚大学 Berkeley 分校的学生写作及研究生论坛中收集隐喻用例,手工编辑而成。后来由 George Lakoff 等人完成了第二版。该知识库包含了隐喻映射和每个隐喻的隐喻实现,并在互联网上公布。数据库中的隐喻分为 event structure(事件结构)、mental events(心理事件)、emotions(情感)和 other(其他)四个范畴,共涉及词条 69 条;每个词条又分为不同层级的隐喻类别,共约 200 多个不同层级的隐喻类别。每种隐喻的类别都包含了各种隐喻方式的源域(source domains)和目标域(target domains)的描述及隐喻例句;每一类隐喻都有一个简要分析。MML 依照专家们的直觉分成不同的概念隐喻类别,不去考虑这些隐喻是否已经词汇化。

例如,在词条“force”下“Force is a substance contained in affecting causes”隐喻类别示例如下:

Force is a substance contained in affecting causes

Put more force into your punches.

He said some forceful words.

His punches carry a lot of force.

Each sentence contained the force of an order.

His punches have a lot of force.

Her death hit us all forcefully.

Her death hit us with a lot of force.

The force of the blow knocked me over.

The force added to the swing was enough to send the ball over the fence

Related metaphors: related to Causes are Forces

Source domain: sustance, contents, container, hitting

Target domain: force

Note:

The amount of force determines the degree of change in the affected party

Either the force or the causal-container can cause the effect.

Force is placed in the container by the person who wants the change in the affected part

Force is the ability to cause

示例给出了该类别的九个相关例句,有对该类别的简要分析(note),指出了隐喻的源域为“sustance, contents, container, hitting”,目标域为“force”,而且还给出了与词类别相关的隐喻类别“Related metaphors: related to Causes are Forces”。

MML 第一版完成于 1989 年,第二版完成于 1991 年。很快,这 200 多个隐喻类就成为各类形式化的、为隐喻计算服务的隐喻知识库建设的主要资源。

### 1.2 Sense-frame

Sense-frame 是一个词例化的语义知识库,共包含 500 条词条。Dan Fass 根据 Wilks 的优先语义学理论的优先中断思想提出了修正语义学的方法,在一个统一的语义框架内对转喻(metonymy)、隐喻和词义异常使用进行解释,并给出了实现程序 Met5。Sense-frame 就是为该程序的实现而建设的词例化隐喻知识库。在该知识库中,每个词义框架包含 arcs 和 node 两部分。Arcs 部分包含了该词条类属条目(一个词条和它的意义描述),语义框架所有的 arcs 构成了词语深层结构化的语义网络;node 部分包含了被定义的语义框架的差异。

名词“animal”在 Sense-frame 中的描述如下:

```
sf( animal1,
  [ [ arcs, [ [ supertype, organism1 ] ] ],
  [ node0,
    [ [ biology1, animal1 ] ],
```

```
  [ composition1, flesh1 ] ,
  [ it1, drink1, drink1 ] ,
  [ it1, eat1, food1 ] ] ] ) .
```

其中:[ supertype, organism1] 是一个语义网络体系;node0 表示该词条是一个名词;[ biology1, animal] 和[ composition1, flesh1] 为该词条的语义优先特征;[ it1, drink1, drink1] 和[ it1, eat1, food1] 为该词条的句法组合模式;it1 指该词条,即 animal1 在[ it1, drink1, drink1] 中能够被 it1 代替,在[ it1, eat1, food1] 中也能够被 it1 代替。该句法组合模式在 Sense-frame 中被称为 cell。

形容词“yellow”在 Sense-frame 中的描述如下:

```
sf( yellow1,
  [ [ arcs, [ [ superproperty, coloured1 ] ,
    [ property, yellow1 ] ] ]
  [ node1,
    [ [ preference,
      [ [ bounds1, bounded1 ] ,
        [ composition1, physical1 ] ,
        [ extent1, [ not1, zero_dimensional1 ] ]
        [ animacy1, nonliving1 ] ] ] ] ]
  [ assertion,
    [ [ color1, yellow1 ] ] ] ] ) .
```

其中:[ superproperty, coloured1] 和[ property, yellow1] 是一个语义网络体系;node1 表示该词条是一个形容词/副词;preference 部分为该词条的语义优先特征;assertion 部分为该词条在命题中的断言。

动词“eat”在 Sense-frame 中的描述如下:

```
sf( eat1,
  [ [ arcs, [ [ supertype, [ ingest1, expend1 ] ] ] ]
  [ node2,
    [ agent, [ preference, animal1 ] ]
    [ object, [ preference, food1 ] ] ] ] )
```

其中:[ supertype, [ ingest1, expend1]] 是一个语义网络体系;node2 表示该词条是一个动词;[ agent, [ preference, animal1]] 和[ object, [ preference, food1]] 为该词条的语义优先特征。

利用上述对名词、动词和形容词/副词的不同形式化描述,Fass 用程序 Met5 实现了对隐喻理解的模型解释。例如对“car drinks gasoline”,Met5 系统就是利用如下知识表示实现的:

```
sf{ drink1,
  {{ arcs,
    {{ supertype{ ingest1, expend1 } } } } .
  { node2,
    {{ agent,
      { preference. animal1 } }
    { object,
      { preference. drink1 } } } } }
sf{ animal1,
  {{ arcs,
    {{ supertype, organism1 } } } ,
  { node0,
    {{ biology1, animal1 } ,
    { it1, drink1, drink1 } .
    { it1, eat1, food1 } } } }
sf{ car1,
  {{ arcs,
    {{ supertype. motor_vehicle1 } } } .
  { node0,
    {{ it1, carry1, passenger1 } } } }
```

### 1.3 MetaBank

MetaBank 是 J. H. Martin 在 20 世纪 90 年代为对自然语言中习惯性隐喻给出语义解释而开展的隐喻研究项目。该项研究将经验诱导与理论驱动相结合,在三个方面探索了隐喻知识库的构建问题:文本语料的收集与语言知识的泛化;资源的拓展方法;知识库的构建。这三个方面的综合为直接使用隐喻知识对语言中的隐喻理解提供了一个整体框架。



在文本语料的收集与语言知识的泛化方面, MetaBank 采用了 MML、一个包含 150 000 词次的 UNIX 操作系统用户邮件语料库以及 Wall Street Journal 语料库。其中比较有特色的是 UNIX 操作系统用户邮件语料库。这一语料库的选择为 Meta-Bank 研究特定领域相关语料库提供了必要的资源。

资源的拓展包括两个方面: 如何分析和识别单独句子所包含的习惯性隐喻; 如何在大规模语料中高效搜索习惯性隐喻。MetaBank 接受了 Reddy<sup>[17]</sup>、Lakoff 等人<sup>[18]</sup> 给出的习惯性隐喻定义, 即习惯性隐喻是一些词或短语在其常规用法中脱离字面意义, 使用一种概念类型表达另外一种概念类型的一种语言现象。示例如下:

I can 't kill 'em-eclipse says they aren 't mine.  
其中: kill、says、mine 的使用即是习惯性隐喻的实例。在大规模语料库搜索方面, MetaBank 采用了直接搜索已知隐喻和随机抽取的方式来拓展资源。隐喻搜索依赖于目标域中的关键词进行。搜索完成后搜集的信息包括该种隐喻的出现频率、该类目标域中各关键词使用的频率、该类隐喻中源语义场有多少关键词与之同现、对该种隐喻目标域和源域进行抽象时的正确程度等。

MetaBank 认为隐喻知识的表示包括三个方面: 目标域的知识表述、源域的知识表述以及隐喻本身的知识表述。为此, MetaBank 采用了 Gruber<sup>[19]</sup> 提出的 OntoLingua 知识表述语言 (ontolingua knowledge representation language), 认为这种语言能够提供隐喻的独立的规范的表示方法, 并能为该知识库的潜在使用提供支持。

1. 4 Metalude

Metalude( 详见 [http://www.ln.edu.hk/1le/cwd03/Inproject\\_chi/home.html](http://www.ln.edu.hk/1le/cwd03/Inproject_chi/home.html)) 是一种已经词汇化的英语隐喻交互语料库, 该语料库从词汇上考查隐喻, 并非单纯的认识或知觉的研究。这一点与 Berkeley 的 MML 有明显的不同。目前 Metalude 包含 9 000 个英语词条以及它们的字面义、隐喻义、词类、实例等属性信息。其所收的隐喻均根据概念隐喻或构成隐喻的基本类来进行分类, 因此一个词形往往有多个词条。例如 lexical term = fleet ( 港湾) 有 “Activity is boat travel; Organisation is ship; Traffic is Liquid/blood” 三个根隐喻, 因此设定三个词条。

Metalude 对于它所收的词汇以及所假定的概念隐喻/基本类均有更严格的词汇学标准, 所收的词汇必须有某个隐喻义项见诸于某一本当代英语词典中。该语料库中设置的每一个基本类都必须出现在至少 6 个词汇上, 而且这些词汇在 Cobuild 英语在线词库中出现的频率不能低于 200 次。例如:

- a) 类: Achievement/success is high( 成就或成功即高)  
词条 “hang in there ” ( 字面意义 “悬挂 ”\_\_ ) idi ( vi + adv + adv ) , 隐喻意义 “尽管困难也要避免失败 ”。例句: When the marathon gets tough just hang in there ( 在剧烈的马拉松比赛中只有坚持才能避免失败) 。
- b) 类: Activity is boat travel ( 活动即划船)  
词条 “anchorman ” ( 字面意义 “接力棒 ”\_\_ ) n. , 隐喻意义 “新闻节目主持人 ”。例句: The early morning news programme has a new anchorman( 早晨的新闻节目换上了一个新的节目主持人) 。

上述两例中, “in there ”和 “man ”不是词项的隐喻部分, 因

此在圆括号中只译出 “hang ”和 “anchor ”, 后加空格。如果一个复合词或习语中的所有词语都有属于同一类的隐喻意义, 而且这个复合词或习语的字面意义很少用或从未用过, 那么这个字面意义也将放于圆括号中, 但是不加空格号。例如:

- c) 类: Activity is boat travel( 活动即划船)  
词条 “stick/put your oar in ” ( 字面意义 “摇桨, 推桨 ”) idi ( vt + nphr + adv ) , 隐喻意义 “恼人地打断 ”。例句: He kept on putting his oar in until I lost my temper( 令人气恼的是, 他一而再、再而三地打断我的讲话, 直到我发了脾气为止) 。

对于词性发生变化的词条, 中间用一条竖杠分隔开, 圆括号中的词性与圆括号外的词性相对应。前者为表示字面意义时的词性; 后者是该词做隐喻用法时的词性。例如:

- d) 类: Activity is boat travel( 活动即划船)  
词条 “harbour ”( 字面意义: 海港) ( n ) |vt, 隐喻意义 “庇护 ”。例句: The taleban were accused of harbouring Osama Bin Laden( 塔利班受控庇护本 · 拉登) 。

1. 5 Hamburg Metaphor Database

汉堡隐喻数据库 ( hamburg metaphor database, HMD, 详见 [http://wwwl.uni-hamburg.de/metaphem/index\\_en.html](http://wwwl.uni-hamburg.de/metaphem/index_en.html)) 是一个可供在线搜索的德语、法语隐喻数据库, 于 2002 年面向公众开放。其中数据以来自一些特定领域的大众传媒语料库中的隐喻为主。数据库中的隐喻标注包括词汇和概念两个层次。词汇标注所采用的资源是 EuroWordNet; 概念标注所采用的资源是 Berkeley 主要隐喻列表 ( master metaphor list) 。

HMD 中的隐喻标注的格式大致如表 1 所示。

表 1 HMD 中的隐喻标注示例	
数据	示 例
例句	Le parti de Helmut Kohl qui doit sortir demain comme le seul et le grand triomphateur( 科尔的党派明天将是惟一和最大的胜利者). C'est par centaines de milliers que les Berlinoises de l'Est ont afflué au cours du week-end à l'Ouest( 东柏林的居民在那个周末以数以万计的数量涌向西方).
词条	triomphateur ( 胜利者) affluer ( 涌向)
隐喻意义	vainqueur:1 triomphateur:1 gagnant:1
同义词集	affluer:2 enir en masse:1
字面意义	affluer:1 déferler:1
同义词集	
德语源域	KAMPF ( 战斗) FLÜSSIGKEIT ( 液体)
德语目标域	POLITIK ( 政治) BEWEGUNG ( 运动)
英语源域	FIGHT ( 战斗) MOTION ( 位移)
英语目标域	POLITICS ( 政治) CHANGE ( 改变)

注: 表 1 括号中内容为笔者所加, 感谢南京大学德语系陈民老师细心审读其中德文内容的翻译。

可以看出, 数据由八个部分组成。首先是包含隐喻表达的例句; 然后列举出具有隐喻性的术语, 如表 1 中的 triomphateur 和 affluer。同义词集 ( synset) 信息来自于词网 ( WordNet) 。例如在第一个例子中, 可在法语 EWN 中找到如下同义词集:

- { vainqueur: 1 triomphateur: 1 gagnant: 1}
- 这个同义词集可以用来标注第一个例句中的术语 triomphateur, 或具有能够解释该例隐喻的意义。但是在有些情况下, 这些词语并不具有一个具体的目标, 这样就需要一个字面意义上的同义词集。这里使用了两套源域和目标域标签。一套使用德语, 根据内容创建; 另一套与 Berkeley 提供的隐喻概念标签相一致。

HMD 的建设采用了人工标注与机器辅助标注相结合的方法。2002 年该库开始提供在线检索时有 160 多条隐喻实例;其后采用语料库辅助发现的方法,新增 1 000 多条实例。隐喻的机器辅助发现首先需要建立经过标注的语料库;然后需要确定目标域和源域,并借此建立目标域词汇表和源域词汇表。HMD 充分利用了语料库检索功能及频率、词语同现信息等,在仅已知目标域的情况下,通过以下三个步骤获取隐喻实例:

- a) 确定目标域词汇,如 Europe, 获取经过统计加权处理的同现列表( lists of collocates) 作为可能隐喻实例;
- b) 对获取的同现列表,搜索语料库以区分不同用法,并给出同现列表的源域;
- c) 使用 EuroWordNet 中的同义词集和词义关系,对源域的词汇进行扩展。

经过上述处理,隐喻实例、目标域及其词汇列表、源域及其词汇列表都可在语料库帮助下获得。

### 1.6 ATT-Meta

ATT-Meta( 详见 <http://www.cs.bham.ac.uk/~jab/ATT-Meta/Data bank>) 是英国 Birmingham 大学开发的一个基于规则的可程序化计算的信念和隐喻模拟推理系统。项目因以命题态度和基于隐喻的推理为主要研究内容而被命名为 ATT (itude) -Meta( phor)。ATT-Meta 同样以 Berkley 的 MML 为框架,集中讨论其中与大脑状态、大脑处理过程以及心智相关的隐喻。为此,ATT-Meta 也构建了一个隐喻语料库。其中主要包含真实话语中描写心智状态和运行的隐喻,此外也包含一些描写心智状态的暗喻。隐喻实例取自一个包含1 070个文字文本和 65 个演讲记录文本的语料库。

ATT-Meta 隐喻数据库按照隐喻本身类型进行组织( 图 1)。与心智状态和运行相关的隐喻首先被分为 17 类( 包括暗喻)。每一个类别存在两个链接: 一个链接指向对该类隐喻的解释;另一个链接指向隐喻实例页面。在实例页面中,又包含另一个链接,给出该实例的详尽上下文。

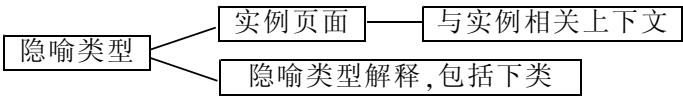


图 1 ATT-Meta 组织构架图

ATT-Meta 对文本中的隐喻实例进行了标注,下例取自于物质实体型主意/情绪比喻( ideas/emotions as physical objects):

例句 The fact that she had not yet seen her step-mother made no difference; she still ( \* felt her dislike of her RISE UP\* ) just as intensely as at first, and it ( \* BLOTTED OUT \* ) all her other feelings.

{ First segment: embedding within feeling. Also, example of "feel" as metaphoricity signal. }

{ " BLOTTED OUT": parallel mixing with COGNIZING AS SEEING. }

[ [ Text source] ]

此例中,ATT-Meta 首先给出这一类型的定义。文本中的隐喻部分被置于( \* ... \* ) 内,相关隐喻词汇大写表示。[ [ Text: source] ] 部分给出该语句的出处链接,已备查询。此外,其后还有对交叉性隐喻等其他情况的说明。

ATT-Meta 隐喻数据库已被尝试用于与心智相关的隐喻识

别和解释研究,并取得了一定成效。

## 2 国内隐喻知识库建设

厦门大学认知与计算研究中心构建了一个具有 1 万句规模 62 万字的汉语隐喻标注句库以及一个面向隐喻识别的汉语常用动词搭配库<sup>[20]</sup>。该隐喻句库采用文本方式进行存储,涵盖了极为丰富的汉语隐喻现象,并具有可扩展性。在该隐喻句库的基础上,进行句法分析和标注。带有句法分析标注的隐喻句库通过哈尔滨工业大学信息检索研究室依存句法分析系统<sup>[21]</sup>生成初步结果,最后由人工校正来完成。这样,带有句法分析标注的隐喻句库中,每个隐喻句采用依存方式表示句子结构以及句内每个成分之间的关系。

句库的结构设计如下: 每一个隐喻句显示为三行,第一行是无标记的生隐喻句;第二行为句法成分的切分与词性标注后的隐喻句,句子中每个词及词性的前面加上序号,句子末尾的句号由 EOS 标志;第三行是隐喻句的依存句法关系。依存关系中,每个关系以一个依存对表示,依存对中的第一个词是核心,支配第二个词,如 “[ 2] 公司\_[ 1] 我( ATT) ”这个依存对表示“我”和“公司”存在依存关系 ATT。其中:“公司”是这个关系的核心成分;“我”依存于“公司”。依存对之间以两个 tab 制表符相隔,整个句子的依存中心单独列出,由标记( HED) 标志,并与句尾句号构成一个依存对,如下所示:

雷霆无情地怒吼,

[ 1] 雷霆 /n[ 2] 无情 /a[ 3] 地 ui[ 4] 怒吼 /vg[ 5] , /wp[ 6] EOS / EOS

[ 3] 地\_[ 2] 无情( DI) [ 4] 怒吼\_[ 1] 雷霆( SBV) [ 4] 怒吼\_[ 3] 地( ADV) [ 6] EOS \_[ 4] 怒吼( HED)

句库的标注标准如下: 词性标注使用的是“863”标注体系的词性标准,包括 28 个词性标记;依存关系标记 24 个,如表 2 所示。

表 2 汉语隐喻句库标记集

标 记 集	标 记 集
定中关系 ATT( attribute)	数量关系 QUN( quantity)
并列关系 COO( coordinate)	同位关系 APP( appositive)
前附加关系 LAD( left adjunct)	后附加关系 RAD( right adjunct)
动宾关系 VOB( verb-object)	介宾关系 POB( preposition- object)
主谓关系 SBV( subject-verb)	比拟关系 SIM( similarity)
核心 HED( head)	连谓结构 VV( verb-verb)
关联结构 CNJ( conjunctive)	语态结构 MT( mood-tense)
独立结构 IS( independent structure)	状中结构 ADV( adverbial)
动补结构 CMP( complement)	“的”字结构 DE
“得”字结构 DEI	“地”字结构 DI
“把”字结构 BA	“被”字结构 BEI
独立分句 IC( independent clause)	依存分句 DC( dependent clause)

## 3 余论

一个隐喻知识系统的建立主要包括以下三方面的内容: 语料库及其类型;隐喻知识表述;隐喻识别算法及隐喻理解模型。语料库存在平衡语料库和受限领域语料库,不同类型语料库中隐喻的使用类型和分布存在不同,因而会对后续处理产生影响。隐喻识别与隐喻知识表示是两个紧密联系的方面。对隐喻概念的不同认识,如隐喻与词汇之间的关系问题的认识导致了不同的隐喻知识结构,因而也影响到隐喻识别与理解方法



的不一致。同时,采用何种知识表述语言来表述隐喻知识也是需要考虑的问题。在大规模的隐喻识别方面, Metabank、Hamburg Metaphor Database 都已经进行了有益的尝试,其中能够利用的关键信息是与源域、目标域相关联的词汇。

此外,需要注意的是 Berkeley 的 MML 为诸多隐喻知识库提供了理论框架。其他隐喻知识库其实就是对这一资源的衍生和拓展。而在国内的隐喻知识库似乎并没有利用这一成果,而且在隐喻知识的表征方面重视程度不够。汉语作为一种重“意合”的语言,非字面义的研究尤为重要。对汉语的隐喻自动理解与计算的研究在国内刚刚起步,亟待建立一个面向隐喻计算的隐喻知识库。

#### 4 工作设想

综上所述,本文提出了一个大规模汉语隐喻知识库的构建设想:隐喻知识库由句库和词库两部分构成。句库主要内容包括汉语隐喻句获取(前期可以采用检索的方式从语料库中抽取带标的隐喻句,即有比喻词的句子;中期可以根据本体研究成果内省式收集无标的隐喻句;后期可以在隐喻类的帮助下利用机器学习的方法自动获取隐喻句)、句法分析(不同于其他计算语言学应用领域的句法分析,隐喻知识库的句法分析要能够体现出句法依存关系,尤其是本体、喻体与比喻词之间的依存关系。当然,词法分析也是必不可少的)、语义标注和隐喻标注。词库主要内容包括词例化隐喻类(以词例为研究对象归纳隐喻的概念类别,概念类别类似于 MML 的分类;词例化类似于 Metalude 或 HMD 中对词例的描写)、典型例句、隐喻域、隐喻类体系。

该隐喻知识库的构建目标为句库规模达到 10 万句,约 200 万字;词库规模达到 5 000 词条。最终成果为一个检索系统:用户可以输入整句检索其隐喻类,可以输入具体词汇检索隐喻句,也可以输入隐喻句检索其隐喻类等。

参考文献:

[1] 黄孝喜,周昌乐.隐喻理解的计算模型综述[J].计算机科学,2006,33(8):178-183.

[2] 王治敏.隐喻的计算研究与进展[J].中文信息学报,2006,20(4):16-24.

[3] 张霄军.隐喻与换喻计算综述[C]//第四届全国认知语言学研讨会.南京:南京师范大学,2006.

[4] FASS D. A method for discriminating metonymy and metaphor by computer[J]. Computational Linguistics, 1991, 17(1): 49-90.

[5] RESNIK P. Using information content to evaluate semantic similarity in a taxonomy[C]//Proc of the 14th International Joint Conference on Artificial Intelligent. Montreal: [s. n.], 1995: 448-453.

[6] WEINER E J A. Knowledge representation approach to understanding metaphors[J]. Computational Linguistics, 1984, 10(1): 1-14.

[7] MARTIN J H. A computational model of metaphor interpretation[M]. New York: Academic Press, 1990.

[8] VEALE T. Metaphor, memory and meaning: symbolic and connectionist issues in metaphor interpretation[D]. Dublin: Trinity College, 1995.

[9] STEINHART E C. The logic of metaphor: analogous parts of possible worlds[M]. Dordrecht: Kluwer Academic Publishers, 2001.

[10] GENTNER D. Structure-mapping: a theoretical framework for analogy[J]. Cognitive Science, 1983, 7(2): 155-170.

[11] GENTNER D, FALKENHAINER B, SKORSTAD J. Viewing metaphor as analogy[M]//HELMAN D. Analogical reasoning. [S. l.]: Kluwer Academic Publishers, 1998.

[12] HOLYOAK K J, THAGARD P. Analogical mapping by constraint satisfaction[J]. Cognitive Science, 1989, 13(3): 295-355.

[13] KINTSCH W. Metaphor comprehension: a computational theory[J]. Psychonomic Bulletin & Review, 2000, 7(2): 257-266.

[14] KINTSCH W, BOWLES A. Metaphor comprehension: what makes a metaphor difficult to understand[J]. Metaphor and Symbol, 2002, 17(4): 249-262.

[15] MASON Z. Computational, corpus-based metaphor extraction system[D]. [S. l.]: Brandeis University, 2002.

[16] MASON Z. CorMet: a computational, corpus-based conventional metaphor extraction system[J]. Computational Linguistics, 2004, 30(1): 23-44.

[17] REDDY M J. The conduit metaphor[M]//ORTONY A. Metaphor and thought. Cambridge: Cambridge University Press, 1979: 284-324.

[18] LAKOFF G, JOHNSON M. Metaphors we live by[M]. Chicago, Illinois: University of Chicago Press, 1980.

[19] GRUBER T R. Ontolingua: a mechanism to support portable ontologies[R]. Palo Alto: Knowledge Systems Laboratory, Stanford University, 1992.

[20] 李剑锋,杨芸,周昌乐.面向隐喻计算的语料库研究和建设[J].心智与计算,2007(1):142-146.

[21] 尤昉,李涓子,王作英.基于语义依存关系的汉语语料库的构建[J].中文信息学报,2002,17(1):46-53.

### 下期要目

- 基于进程行为的异常检测研究综述
- 卫星对地观测任务规划问题简明综述
- 数据流系统降载研究综述
- 一种面向实时应用的时态数据库系统体系结构
- 基于起点预测的不连续十字形快速搜索算法
- 一种用于求解多目标组合优化的混合遗传算法
- 基于信息扩散原理的关联规则自动评价方法研究
- 群体智能混合算法优化 XML 数据查询策略
- 面向可信软件的风险管理模型研究
- 构件行为模型的构造研究
- 基于多级分类的大类别人脸识别研究
- 国际主流视频编码标准优化设计代码的对比测试
- 基于双线性对的门限代理盲多重签名方案
- 基于追踪部署的着色包标记算法的研究
- 利用关联和奉贤评估方法减少误报和漏报
- 基于规范的入侵检测技术在无线传感网中的研究
- 无线传感器网络中的拓扑控制
- 模极大值与阈值决策融合的小波语音数据去噪方法
- 结构化多点协作 P2P 系统研究
- 基于局域波分解的雷达辐射源信号时频分析
- 局域网流量的多重分形谱分析及关联性研究
- 变频多联机空调模糊逻辑控制及仿真研究
- 面向农产品安全生产的空间决策系统研究与实现
- 基于达芬奇技术的日面活动图像处理与识别系统