名词转喻的自动理解*

李斌[1] 曲维光[2,3] 陈小荷[1]

1. 南京师范大学文学院 南京 210097；2. 南京师范大学计算机科学与技术学院 南京 210097

3. 江苏省信息安全保密技术工程研究中心 南京 210097

E-mail: libin.njnu@gmail.com

摘 要：转喻是汉语文本中常见的语言现象。在主宾语位置上，名词会出现转喻用法。该用法往往是用凸显特征转喻本体，而凸显特征则蕴含在世界知识和人类主观体验之中，计算机自动识别转喻特别是找出转喻的本体难度很大。对此，我们提出了两点策略来识别转喻的本体，首先利用聚类搜索引擎获取和喻体词语高度相关的词语列表，以弥补世界知识和主观体验的不足；然后使用动词对名词的语义选择限制计算词语相似度，根据分值高低锁定相关词，以得到转喻的本体。在五个典型的转喻实例上，实验取得了较为理想的结果。

关键词：转喻，选择限制，聚类引擎，词语相似度，相关词

The Automatic Understanding of Noun Metonymy

Li Bin[1] Qu Weiguang[2,3] Chen Xiaohe[1]

1. School of Chinese Language and Literature, Nanjing Normal University, Nanjing 210097

2. School of Computer Science and Technology, Nanjing Normal University, Nanjing 210097

3. Jiangsu Research Center of Information Security & Confidential Engineering, Nanjing 210097

E-mail: libin.njnu@gmail.com

Abstract: Metonymy is a common phenomenon in Chinese texts. Noun metonymies may occur on the subjects or objects of verbs under the mode of "profiled feature instead of tenor". Profiled feature is embodied in real-world knowledge and subjective experience. Thus, the auto recognition of metonymies and detecting of tenors are difficult for the computer. Here, we suggested a novel model for metonymy understanding. First, clustering engine is queried to get the related words of the vehicle word. Second, word similarities between the verb-noun selectional restrictions and the related words are calculated to locate the tenor words. The experiments on 5 typical metonymies achieved good results.

Keywords: Metonymy, selectional restriction, clustering engine, word similarity, related words.

1 引言

转喻（metonymy）在传统修辞学中被称为“借代”，指用凸显特征或相关事物指代某一事物。比如下面例句 1-5 中 “红领巾”转喻“少先队员”，“贝多芬”转喻“贝多芬的音乐”等。随着认知语言学的兴起，Radden & Kevecses（1999）在借鉴前人的基础上，提出了一个广为人们接受的新定义：转喻是在同一理想化认知模型中，一个概念实体（即源域）为另一概念实体（目标域）提供心理通道的认知操作过程。因此，转喻与人们日常的认知体验密切相关。由于计算机缺乏世界知识和人类的日常体验，对“红领巾”和“少先队员”这类关系是很难把握的。如果给出“红领巾”，想让计算机自动得出“少先队员”来显然也是比较困难的。

例 1 红领巾拾金不昧。

例 2 大盖帽出示了自己的证件。

例 3 他爱听贝多芬。

*本文承国家社科基金项目“汉语词语搭配获取与语义特征分析的相互关系研究”（07BYY050）和国家自科基金项目“汉语隐喻理解关键技术研究”（60773173）的资助。

例 4 他一天到晚听耳机。

例 5 他整天吃食堂。

从词类上看，转喻中最主要的类型是名词的转喻，也是语言学界研究最多的转喻类型。因此，本文主要从动词对名词选择限制的角度出发，分析宾语名词转喻的机制并给出自动理解的方法。

2 名词转喻和选择限制

名词转喻的现象是较为普遍的，语言学界对转喻的机制主要有两种看法。Lakoff & Johnson（1980）认为转喻的主要功能是指称，沈家煊（1999）也以此观点讨论了“壶开了（壶里的水开了）”之类的转喻。这种观点基本上在名词本身的层面上论述转喻的本体和喻体之间的关系（部分-整体关系等）和凸显度问题。Langacker（1984，1990:189）则认为转喻并不是名词本身引起的，在句子中凸显出不同的侧面（facets），是谓词的含义决定的，并提出了“活跃区域分析法（active zone analysis）”。如，“She heard the piano（她听见了钢琴）”中，听钢琴实际上是听钢琴的声音，认为 hear 在句中的意思是“主语听宾语的声音”，是动词 hear 促成了人们对隐含的“声音”的理解。这种观点使用了动词对论元成分的语义选择限制来解释转喻的生成和理解机制。

在计算语言学界，Fass（1991）提出了转喻的计算模型，认为应通过动词对主宾语的语义约束来识别转喻。之后在主流的研究中，名词转喻的识别都在动词选择限制的框架下操作（Mason，2002）。转喻的本体识别也引起了一些学者的关注，Nissim（2005）和 Peirsman（2006）通过总结本体和喻体之间的关系，对不同类型的命名实体（国名、公司名、产品名），从语义相似度和机器学习的角度来计算转喻。2007 年的 Senseval 竞赛上的举办了一个专名转喻识别的比赛项目[1]。不过这种训练、测试的方式抛弃了语言知识及推理方式，完全依靠上下文的信息来进行自动分类，虽然可以得到较高的精度，但其缺陷是面对新类型的命名实体或其他词语时依然需要新的训练语料，缺乏可扩展性。

综合前人的研究，我们可以看出，转喻的理论问题主要集中在两点上：一是本体和喻体之间的存在什么关系，喻体相对于本体的显著度如何；二是转喻的理解机制问题，即谓词是如何凸显出名词的不同侧面的。从认知的角度看，任何一个事物都可以从不同的角度进行观察和言语陈述，比如“钢琴”是商品（买钢琴），是乐器（弹钢琴），有外观（钢琴很漂亮），有质量（钢琴很好）等等。这些不同的方面，都是和谓词紧密相关的。我们赞同 Langacker（1984）的分析，不存在纯粹静态的转喻。如例 1 中，“红领巾”是在句子中转喻，离开了语境，特别是动词的语义选择限制，“红领巾”就只是“红领巾”，无法转喻为“人”了。转喻的计算则主要集中在转喻的探测以及和隐喻的区分，而对转喻中的本体的确定还缺乏有效的研究。

3 转喻本体的自动理解

对于本体和喻体问题，前人总结出“整体和部分相互转喻”、“显著特征转喻事物”等关系。但是对于缺乏世界知识和人类经验的计算机来说，“红领巾”和“人”之间的显著度是不太好衡量的。如果用基于义类体系的计算方法，两者的语义距离是比较远的，相似度差别大，隐喻和转喻是难以计算的。这里，我们讨论了转喻理解的两种策略，并给出了实验流程。

3.1 自动理解的两种策略

第一种策略是，使用语义类的下位词语和搭配限制进行相关度计算，找出转喻的本体。如例 1 中，根据动词的选择限制“人”这个语义类，选取它的典型类别“婴儿、儿童、少年、中年、老年、职务、男、女”等下位词语作为中介，通过基于普通搜索引擎的互信息算法，依次计算其

1 网址：http://nlp.cs.swarthmore.edu/semeval/tasks/task08/summary.shtml。

相关度，则“红领巾”和“人”的中间媒介“儿童”和“少年”的相关度会比较高。但是这种方案存在三个困难：一，如果依靠人的内省，则很难穷尽某个语义类的所有“下位”词语；二，如果采用《同义词词林》等语义体系，则下位词语数量很多，计算量太大；三，即使找到“少年儿童”也并不准确，因为人们心目中的答案是“少先队员”。

图 1　bbmao 自动聚类结果：红领巾、大盖帽、贝多芬、耳机、食堂

注：搜索日期是 2009 年 2 月 19 日，标签后边的（）内是网页数量。

第二种策略是，使用喻体词语在文本中同现的相关词和搭配限制进行相似度计算，确定转喻的本体。在一个句子中，一般只出现转喻的喻体。而在一个篇章中，本体和喻体却经常同现。通过大量的语篇，检索和喻体（如“红领巾”）同现的词语，进行频率统计和排除停用词后，就可以得到相关词的集合（如，“小学”、“儿童”、“旗帜”等）。对于这些相关词，与搭配限制的语义类别逐个进行相似度计算，从而得到候选本体词语的排序。语料的规模越大，种类越多，越能

够凸显出关系强的相关词。该策略的难点就在于大规模语料的相关词获取技术。而大规模的相关词计算正是目前自然语言处理的一个热门课题，也出现了自动查询相关词的聚类搜索引擎，英语的如 Clusty（http://clusty.com/），中文如 bbmao（http://www.bbmao.com/）等。因此，我们采用这种方法来自动理解转喻。

3.2 自动理解流程

搜索聚类引擎可以获得我们需要的相关词列表。我们通过 bbmao 给出了学界常讨论的转喻词语的聚类结果（见图 1）。图中可以看出，聚类的结果主要有四类：一类是我们期望得到的本体词语，如“少先队员”，在图中用圆圈标出；一类是喻体的扩展短语，如“红领巾广播”；一类是其他的相关词语，如地点、事件等；最后一类是几乎每个查询词都会得到的“视频、图片”两个词语。我们过滤掉第四类词语，把前三类词语和搭配限制“人”做相似度计算后，加上网页数量作为加权系数，可以很容易得到“少先队员>小学生>孩子们>儿子>老师”的序列，这样使得本体的自动发现成为可能。

4 实验结果及分析

计算的对象是例 1 到 5 中的五个词语，“红领巾”、“大盖帽”、“贝多芬”、“耳机”和“食堂”。参照《知网 2000》的语义体系，这些动词对喻体成分的语义限制分别为“human|人”、“human|人””、“sound|声-information|信息”、“sound|声-information|信息”、“food|食品”。为了计算的方便，我们直接使用了义类的代表词作为相似度计算的依据，即转化为“人”、“声音-信息”和“食品”。

我们使用的聚类引擎是 bbmao，采用的词语相似度计算工具是刘群（2002）介绍的基于《知网 2000》的相似度工具包[2]，因此，我们采用的语义体系也是《知网 2000》[3]。下面分别来计算语义类的代表词和聚类结果词语的相似度。

表 1 “红领巾”聚类词语和“人”的相似度

聚类词	相似度	聚类词	相似度
少先队员	1.0	早上	0.139181
小学生	0.722222	学习	0.074074
儿童	0.722222	作文	0.074074
日子	0.7	少年先锋队	-1
视频	0.678451	胸前的红领巾	-1
妈妈	0.661111	原创	-1
儿子	0.640741	工委	-1
来源	0.622222	孩子们	-1
价格	0.598878	红领巾公园	-1
图片	0.186047	红领巾广播	-1
知识	0.186047	红领巾是红旗	-1
学校	0.152047	江苏省	-1

注：-1 表示《知网 2000》没有收录的词语

（1）红领巾——人

2 下载地址：http://www.nlp.org.cn。

3 下载地址：http://www.keenage.com。

表 1 给出了语义类代表词“人”和聚类词语的相似度的降序排列结果。“少先队员”赫然列于榜首，非常符合人的日常经验。而相似度最高的另外个名词“小学生”和“儿童”也都接近我们需要的答案。

（2）大盖帽——人

“大盖帽”的计算结果，得分最高的词语依次排列如下：

特征:0.755518，交警:0.722222，警察:0.722222，威望:0.691407，男人:0.661111。

我们看到排在前三位的比较符合要求，“特征”一词比较例外。不过没有关系，我们只要稍加调整即可得到满意的结果。

“特征”是属性类的词语，明显不在“人”所在的“实体”类中。根据《知网》自身的义原体系，我们把知网的 7 大类义原做一个简单归并，把“数量”和“第二特征”归入“属性”，“数量值”归入“属性值”，从而形成“实体、事件、属性、属性值”四大类别。然后对相似度计算的结果增加一条规则，惩罚掉不在一个大类中的词语。这样，我们可以成功地惩罚掉“特征”、“威望”，留下“交警”、“警察”和“男人”。

“城管、执法人员”两个词语是《知网》没有收录的，只能采用专名识别等技术来处理，在本文中，我们暂不关心像这样的未登录词问题。

（3）贝多芬——声音、信息

我们采用惩罚措施后，继续来看“贝多芬”的聚类词语和“声音”的聚类结果，相似度最高的四个词语依次如下：

全集:0.242424，歌曲:0.186047，协奏曲:0.186047

这个结果还是比较令人满意的。不过，用“声音”聚类的得分都比较低，我们再来看和“信息”的相似度计算结果：

歌曲:0.615385，协奏曲:0.615385 交响曲:0.369231

这三个词语更加合乎人们的直觉。

（4）耳机——声音、信息

下面我们继续用“声音-信息限制”对“听耳机”中的“耳机”进行计算，得出的结果也比较令人满意，得分最高的是“音乐”，其他的词语普遍得分较低。

声音——音箱:0.242424，东莞:0.208696，音乐:0.186047，接口:0.186047

信息——音乐:0.615385，音箱:0.186047，市场:0.171429

（5）食堂——食品

最后，我们来计算一下语法学界经常讨论的“吃食堂”的转喻本体，用“食堂”的聚类词语和“食品”进行相似度计算，看看结果如何。

饭菜:1.0，食品:1.0，小吃:1.0，美食:0.406838，菜谱:0.171429

“饭菜”、“食品”和“小吃”三个词语排在了第一位，这个计算结果同样令人满意。

我们在使用聚类引擎、语义类限制和相似度计算的时候，仅仅采取了两个小策略进行简单过滤，对这五个典型名词转喻的实验，都得到了令人满意的结果。分值最高的词语都是人们心目中比较理想的答案；排在前三位的其他词语也基本接近于理想答案。该方法可以很方便地实现对更多的名词转喻实例进行本体的识别。

5 结论及未来工作

对于人来说，转喻的使用是生动的、形象的，转喻的理解是自然的、微妙的，然而，在自然语言处理过程中，转喻因其复杂性而带来了诸多问题。本文对汉语中常见的名词转喻现象进行自动地计算，综合使用了两种策略，首先采用聚类引擎来获得喻体的相关词，这样就弥补了计算机在世界知识和人类经验方面的不足；其次，以动词对名词的语义类限制为中介，与相关词进行

相似度计算，根据分值高低确定本体，从而成功地找出了“红领巾”、“大盖帽”、“贝多芬”、“耳机”、“食堂”五个名词所转喻的本体。

转喻的类型较多，比如汉语中“端水”之类的转喻，由于“水”和“盛水的容器”之间的关系不是一种凸显关系，而是一种动态的“内容-容器关系”，用本文提出的方案是难以识别的，需从其他的角度去研究。

对于具有凸显关系的名词转喻，本章的工作仅仅是个案的、实验性的，在今后的工作中，我们还需要在以下几个方面进行拓展：（1）对更多名词的转喻现象进行计算；（2）尝试自动识别出本体和喻体之间的关系，比如“整体-部分”关系、“对象-凸显特征”关系等等，从而达到转喻的完整理解；（3）研究如何自动地发现转喻，从而使得本体的识别模块有一个良好的触发机制；（4）进一步研究字面义和转喻、隐喻的区分，建立较为完整的转喻识别和理解系统。

参 考 文 献

[1] Fass, D., met*: A Method for Discriminating Metonymy and Metaphor by Computer [J].Computational Linguistics, 1991, 17(1):49-90.

[2] Lakoff, G. Johnson, Mark. Metaphors We Live By [M]. Chicago: University of Chicago Press, 1980.

[3] Langacker, R.W. Active Zones[A], Proceedings of the 10th Annual Meeting of the Berkeley Linguistic Society[C], 1984:172-188.

[4] Langacker, R.W. Concept, Image, and Symbol: the Cognitive Basis of Grammar[M]. Berlin: Mouton de Gruyter 1990; reprinted 2002.

[5] Markert K. and Nissim M. Metonymy Resolution as a Classification Task[A]. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, Philadelphia, Penn., 6-7 July, 2002: 204-213.

[6] Mason, Z. A Computational, Corpus-Based Metaphor Extraction System [D]. Brandeis University, 2002.

[7] Nissim M., Markert K. Learning to buy a Renault and talk to BMW: A Supervised Approach to Conventional Metonymy[A]. In Proceedings of the Sixth International Workshop on Computational Semantics (IWCS-6), Tilburg, Netherlands. January 12-14, 2005.

[8] Peirsman Y. Example-based Metonymy Recognition for Proper Nouns[A]. In Proceedings of the Student Research Workshop of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006), Trento, Italy, 2006:71-78.

[9] Radden G., Kövecses Z. Towards a Theory of Metonymy[A]. In Panther K. U., Radden G. (eds.) Metonymy in Language and Thought[C]. Amsterdam: Benjamins, 1999:17-60.

[10] 刘群，李素建，基于《知网》的词汇语义相似度计算[A]，第三届汉语词汇语义学研讨会论文集[C]，台北，2002.

[11] 沈家煊. 转指和转喻[J]. 当代语言学. 1999（1）.