

一种新的处理汉语动名超常搭配的方法*

汪梦翔

北京联合大学师范学院, 北京 100010

wmx1314-@126.com

摘要

动名超常搭配由于一般带有成分的省略或隐喻, 给中文信息处理带来了难度。以往一般是通过整体打包进词库的形式来直接处理, 而本文的处理思路是对汉语动名超常搭配进行解构还原的方式来进行处理。具体做法是, 依托西方生成词库理论的思想构建一套专门的汉语词项描述体系, 这一知识表示体系可以较为清晰的还原因省略或隐喻而造成的非常规搭配, 进而解读出它们的组合机制和生成过程。然后本文通过补缺和替换的形式把动名非常规搭配还原为常规性搭配来进行处理。实验表明, 这种思路处理动名超常搭配切实有效。

关键词: 超常搭配; 生成词库论; 知识库

A New Method for the Processing of Chinese Verb-Noun Anomalous Collocations

Mengxiang Wang

Teachers' college of Beijing Union University, Beijing 100010

wmx1314-@126.com

Abstract

Verb-Noun Anomalous Collocations are generally accompanied by the omission or metaphor of components, which bring difficulties to the Chinese processing. In the past, it is generally classified the anomalous collocations into the lexicon. While this paper is to deconstruct or restore the anomalous collocations. Firstly, we build a set of special word-description system based on the Generation Lexicon theory. This knowledge representation system can restore Anomalous Collocations which caused by omitting or metaphor and interprets their combination mechanism and the generation process. Then we restore the the anomalous collocation to normal collocation by adding and replacing ingredients to process it. The experiment shows that the method is effective in the Processing of Chinese Verb-Noun Anomalous Collocations.

Keywords: Anomalous Collocation, Generation Lexicon Theory, Knowledge Base

1 引言

我们先看几个例句:

A类

(1)两个黄鹂鸣翠柳。(在翠柳上鸣)

©2021 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

基金项目: 北京社科基金青年项目 (20YYC021)

(2)今晚我要赶稿子，就不陪你了。（赶着写稿子）

B类

(3)乘帆起航了。（乘船）

(4)七连只花了一天功夫就吃掉了敌军。（消灭了敌军）

(5)电脑开机后不读硬盘了。（获取信息）

这几个句子划线的部分，它们有个共同特征，就是动词所接的名词宾语按照动词的常规选择性限制条件（Selectional Restriction, SR）来说都不是动词的常规宾语，但人们都认可和接受这种搭配，我们称之为动名超常搭配。动名超常搭配的内部成因和表现形式比较复杂，孙梦询（2012）按照成因归纳出了8类（1.轻动词成因；2.介词省略成因；3.关键信息组合成因；4.种类高度概括成因；5.古代结构残留；6.古今词义变迁；7.对举和用典；8.义场系统成因）。不过我们从他们的分析再结合总体成因，认为可以归纳为两类现象：省略和隐喻¹。其中（1）、

（2）可以分为一组，因为其中并非是动词直接作用于后接名词，而都含有某种成分的省略：有的是省略了谓词，有的是省略了介词。比如（1）句里的动名关系可以理解为动作和场景的关系，我们可以通过补充介词并进行移位的方式来还原，而不需要改变名词或动词来替换，如“鸣翠柳”就是“在翠柳上鸣”，此外还有“晒太阳（在太阳下晒）”、“走西口（向西口走）”、“跑资金（为资金跑）”。这种介词省略型动名超常搭配，传统的语义角色标注方法就可以处理，比如“鸣翠柳”可以标注为“鸣+地点”。而（2）句虽然“稿子”前是和动词“赶”搭配，但是真正发生关系的是和缺省的动词“写”，而“赶”和“稿子”是没有语义关系的，因此这类谓词省略型的动名搭配不适合用现有语义角色标注理论处理，我们需要寻找新的方法。

（3）、（4）、（5）则是另一类，其动因是由隐喻造成的，或者说动词或名词需要替换成其他相关的词才能符合一般常规的动宾搭配规则，如：（3）的“乘帆（乘船）”是名词发生隐喻，这样的搭配还有“吃肯德基（吃快餐食品）”、“考鸭蛋（考零分）”；而（4）的“吃掉敌军（消灭敌军）”是动词发生隐喻，此外还有诸如“挖真相（探求真相）”、“打副手（担任副手）”、“咀嚼文意（体会文意）”、“品味生活（体验生活）”等；（5）的“读硬盘（获取信息）”是名词和动词都发生隐喻，我们可以分别进行替换，类似的还有诸如“喝西北风（收获了没用的东西）”、“捞油水（获得好处）”、“吃老本（依靠基础）”。

这些超常搭配不仅是在实际语料中尤其是歌词和诗词中普遍存在，而且在句法分析中也经常出现，因为句法分析的核心是抽取核心动词和其搭配成分，这期间由于信息的缺省往往会造成动词和所搭配成分间形成超常搭配（如“哥抽的是寂寞”中核心是超常搭配“抽寂寞”）。本文就是想为这些常见超常搭配提出一种新的处理思路和方法。

2 相关研究及数据来源

动名搭配的分析和处理一直是关系到句法分析和语义提取的重要环节，通常我们主要关注符合语法规则的常规搭配。实际上，语言中还存在着大量“语法上符合规则而语义上不符合常规认知的超常搭配现象”，而这样的现象与语言的隐喻表达和省略思维有着密切的联系，对自然语言理解将产生重要的影响。对于动宾超常搭配，长期以来学界基本遵循着“大词库、小语法”的思想，也就是不管是省略式还是隐喻式，都是把这些非常规的搭配固化成词汇进行收录。这种方式，从工程角度讲，确实简单有效。但是这样一来，一方面我们被迫需要构造越来越大的词库，另一方面其搭配结构内部的生成机制和描写规则仍然得不到解决，从而导致句法语义分析技术长期停滞不前。20世纪90年代美国计算语言学家Pustejovsky（1995）提出了针对超常搭配问题的生成词库理论（Generative Lexicon Theory, GLT），开始通过针对性的描述词项的语义知识（主要从词语的物性角色来描述本体概念）来达到还原和处理超常搭配的目的。比如针对英语出现的“begin the novel”，Pustejovsky通过描述“novel”的物性角色特征来为短语中所省略的成分圈定可选范围。（比如，“novel”的功能角色有“read”，施成角色有“write”，那么“begin the novel”省略的部分在“to read”或“to write”范围内选择）²。

¹这里的省略和隐喻是广义的说法。省略有时候和隐含纠结，汉语学界曾经把它和隐含做过区分，吕叔湘（1979）指出必须可以添补才叫省略，类似“你一言，我一语”这种只能称为谓词隐含。而隐喻有时候和借代纠结，比如“考了个鸭蛋”和“吃肯德基”这里的“鸭蛋”借代“零分”，“肯德基”借代“西式快餐”，因为没有比喻的性质。本文对这些概念不做区分，只要有成分缺省的，不管是否能添补都称之为省略；只要是出现语义映射关联，需要借助替换概念来帮助理解的都看做是隐喻。

²功能角色和施成角色都是物性角色的一种。Pustejovsky当初把物性角色分为四类，即是：形式角色（回答词项是什么的问题）、构成角色（回答词项由什么组成的问题）、功能角色（回答词项能干什么的问题）、施成角色（回答词项怎么产生的问题）。而且他指出一般谓词隐含的超常搭配，其省略的谓词多为功能角色或施成角色。

生成词库理论的出现为搭配问题的深入处理提供了新的思路，把超常搭配的处理过渡到词语知识表示和描述的建设当中。这也为我们处理汉语动名超常搭配提供了借鉴，但是他所构建的知识表示方式对汉语很多隐喻型的超常搭配就解构不了。（比如，“喝西北风”、“抽寂寞”所隐含的信息，在以往的四物性角色中都没有涉及，按照Pustejovsky（2006）更新的知识描述方法，无法还原“喝西北风”这类隐喻型的超常搭配）。

目前国内对动名超常搭配的处理可以说大部分都集中于超常搭配的认识（主要是隐喻），对动名搭配的认识还不够全面，而且较少涉及到超常搭配识别后的理解和处理。黄洁（2009）、吴琼（2014）认为动名超常搭配是因隐喻而产生，不是动词就是名词发生有隐喻导致。对于隐喻的处理，学者多依托Lakoff（1980）的概念隐喻理论，他们认为，隐喻（metaphor）是两个概念之间的特征映射，两个概念能达成映射，必须具有类比关系，也即是相似性；通过相似性，我们可以把某个事物的特征映射到另一个事物身上，从而找到一种相关。基于这一理论思想，我们对隐喻的处理就落在描述两个概念的相似性和解读两个概念的语义关联上了。目前在隐喻计算方面很多人都是基于这种思路，比如Zachary J.Mason（2004）所构建的CorMet系统，戴帅湘（2005）隐喻形式化方法，王治敏（2008、2009）的汉语名词隐喻识别策略，李剑锋（2008）的隐喻识别模型研究都来源于概念隐喻理论，不同的是他们在提取隐喻实体和知识获取的方式和方法上有所差别。对于隐喻的判别学界一般利用语义选择限制条件来判断，Wilk早在1975年就以违反语义选择限制作为隐喻的判定条件，只不过语义选择限制需要依托一定量级的语义知识。国内贾玉祥（2011）就是从人民日报中提取动词的语义选择限制条件来判断隐喻。不过目前学界主流识别方法一般基于机器学习，主要代表如张威（2004）、游维（2007）、杨芸（2008）、白振凯（2016）等。王贵荣（2021）依托大规模标记语库来从常规搭配对中发现偏离值进而识别隐喻。还有一部分学者依托相当规模的知识库中的语义分类，通过构建语义映射库的方式来进行，如王金锦（2009）。虽然该方法基本思路是通过构建语义知识库的方式来进行隐喻识别，但和本文在语言知识描述的维度和方式上有所不同。

我们处理汉语动名超常搭配的总思路也是基于构建语义知识库的形式，利用一些词项特征，根据省略型和隐喻型超常搭配内部不同的形成机制进行针对性处理。另外，本文在构建汉语生成词库过程中借鉴了北大的《语法信息词典》(Grammatical Knowledge-Base of Contemporary Chinese, GKB)、中文概念词典(Chinese Concept Dictionary, CCD)和“汉语名词物性结构知识库”的部分特征，并结合自身已有前期研究成果，构建了汉语生成词库(Chinese Generative Lexicon, 简称CGL)，通过知识库词项特征的叠加来生成有针对性的、能够解构超常搭配的语义知识，从而完成动名超常搭配的认识和处理。

3 面向动名超常搭配处理的汉语生成词库的构建

3.1 汉语生成词库的总体框架

我们汉语生成词库主要还是围绕动词和名词语义特征的描述来进行，然后通过语义特征的组配规则来针对性处理不同类别的动名超常搭配。知识库的总体搭建框架以及具体的语义特征的内涵和功能如下：

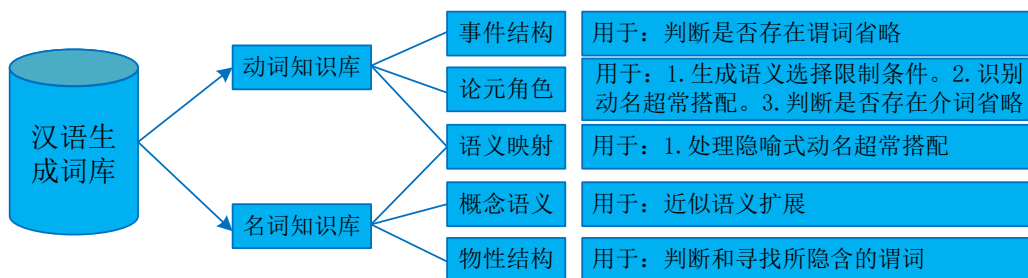


图 1. 汉语生成词库的框架搭建图

3.2 动词的语义特征内涵及针对性功能

专属动词的语义特征主要是论元角色和事件结构。

动词的论元角色，指的是动词所接名词的语义角色。我们除描述动词所能接的语义角色类型外，还进行了基本式和扩展式的区分，一般扩展式是超常搭配的常见语义角色框架。比如“吃饭”是“吃”的基本式“V+受事”，而“吃食堂”则是扩展式“V+地点”。扩展式主要是场景类角色（地点、方式、工具、缘由、时间等）充当。此外针对动名搭配中还对比语义角色标记进行了归纳和整理形成了一个格标库（语义角色标记库）。按照Fillmore“格语法”理论思想，每种语义格都有专属格标记“K”（K可以为零标记）以及优先常规句法位置，比如“地点”的格标记有“在、向”等，且常规句法位置是状语，而“吃食堂”如果确定为“吃+地点”则省略的介词实在是格标记，其常规表达应该是“在食堂吃”，反过来，我们对“吃食堂”这种介词省略型的动名搭配只要判定了语义角色关系，外加确定所添加的格标记，就能准确还原。因此在语义角色特征中，我们重点描述语义关系和每一类语义角色的格标记库。

另外我们利用动词的论元角色特征和名词的概念语义特征组合，可以生成动词的语义选择性限制条件（Selectional Restriction, SR）。本文认为，如果动词所接的名词在SR规则内的则为常规搭配，不在SR规则内的则可以判别为超常搭配。

动词事件结构，主要是为判定动名搭配中是否有谓词省略而设定的，我们主要描述这个动词是否为事件动词。这里事件动词，主要是能否接动词性短语（VP）。如果一个动词不是事件动词，不能触发事件，那么后面即使接了名词，也不具备谓词隐含的可能。因此这一特征就是判断一个动名搭配是否为省略型搭配的重要依据。在事件动词的提取上，我们依托《语法信息词典》关于后接成分的特征（比如是否后接VP、小句）分类，最终从3000个常用动词中抽取了213个事件动词。具体方法可参考汪梦翔（2014）。

3.3 名词的的语义特征内涵及针对性功能

专属名词的语义特征主要是概念语义和物性角色。名词的概念语义实质就是名词的分类，我们采用了中文概念词典（Chinese concept dictionary, CCD）关于名词的概念语义分类。而物性角色的描述，我们借鉴了Pustejovsky的描述模型，并增加了名词的处置角色。其基本内涵，可以由下表所示。

表 1. 本文所描述各物性角色内涵

物性角色	形式角色	描写词项有什么区别于其他对象的属性	一般用名词描述
	结构角色	描写词项所指的对象由什么构成	一般用名词描述
	功用角色	描写词项能够有什么用途和功能	一般用动词描述
	施成角色	描写词项由什么产生	一般用动词描述
	处置角色	描写人对词项的常规性处置动作	一般用动词描述

3.4 语义映射特征

语义映射特征指的是动词或名词的关联转指性特征或喻体性特征，这种特征一般是词项所能比喻或转指的对象。这一特征的描述主要是为了解决隐喻型的动名超常搭配问题，或者是为了寻找动词和名词的合适替换对象。

其中名词的语义映射特征，一般来源于名词所能比喻的对象或者隐含意义，是直接在语料中进行提取。比如“狼”的语义映射特征是“恶人”，“书”的语义映射可以是“阶梯、镜子、宝库”等，具体提取方法可以参考汪梦翔（2017）。

而对于动词的语义映射特征，我们则根据人工的归纳对动词的语义映射对象进行了聚类的限定，这主要是因为动词本身是一种抽象的表达逻辑关联的词语，而且能够带非常规宾语的动词极为有限（我们发现在常见的2568个动词中，只有200多个能够接非常规宾语，而且还相对集中，比如打”、“吃”类所接的非常规宾语比一般动词的种类要多得多）。我们的做法是把动词所能映射的对象限制在九类代表性动词中，每类选一个动词作为代表：

- (1) 从事类Assume（“从事、打、当、承担”等可映射此类，如：“打副手”=从事副手）
- (2) 生成类Aassemble（“制作、拿、组装、装、合”等可映射此类，如：“拿主意”=生成主意）
- (3) 拆分类disassemble（“拆解、分析、解剖”等可映射此类，如“剖析问题”=拆分问题）
- (4) 动作类Act（“打、跑”等可映射此类。这些动词可以理解为展现，一般映射为动作的，映射对象为do或动词本身”，这里的“do”可以理解为一种将后接成分谓词化的虚拟标记，

如打瞌睡= 瞌睡)

(5) 获得类Get (“获得、取、吃、喝、拿”等可映射此类，如：吃亏=获得不好事情)

(6) 给予类Output (“给、送、传、递”等可映射此类，如送温暖=给予温暖)

(7) 消灭类Destroy (“打、吃、救、灭、耗”等可映射此类，如：救火=消灭火)

(8) 依靠类Depend (“依靠、依托、凭借、吃、啃”等可映射此类，如：啃老=依靠老人)

(9) 感知类Feel (“感知、体会、理解、揣摩、品尝”都可映射入此类，如“咀嚼文意”，这里的“咀嚼”相当于“感知”或“体会”)

这种粗线条的划分一方面有利于名词对动词组配的高效率筛选，另一方面映射对象所属等级越高，则包含的继承性特征越多，这有利于提高词项与语义映射对象的匹配度。比如“打水”到底是映射“捞”、“提”还是“取”，无论你选哪个则其他映射的词项语义特征可能就没有。如：选了“捞”可能“提”的某些特征就没有了。那么如果都归结为“获得”，这一特征都被其他词项所隐含和继承。

此外，有些动词可以同时映射多个类，比如“吃”在“吃亏”里是“获得”，在“吃（掉）敌人”里是“消灭”。而且根据“吃”后接非常规宾语的频次，我们得出“吃”的映射对象依次倾向于“获得类”、“依靠类”、“消灭类”，排在前面的可以获得名词的优先选择。

我们选取了两个词作为描述样例：

(1) 书<名词>

概念语义：物/具体物/人工物

物性角色：

形式：有形物质、人工物；

构成：白纸、化石、石头、竹子、竹片、竹筒、纸张、纸、布、丝绸、电子...

施成：写、写成、编著、编写、出版、刻、雕...；

功用：读、朗读、学习、认识、遮、用、赚钱、宣传、当官、教育、关联、铺路、铺墙、垫、推荐...；

处置：撕、扔、丢、抛、甩、烧；

语义映射：知识、智慧、权威、阶梯、力量、朋友

(2) 读(动词)

事件结构：

事件动词：否

后接VP:无（非事件动词后面不能接VP）

论元角色：

【基本式】受事：读+{书、报、杂志、论文、诗、信。。}

【扩展式】结果：读+{博士、硕士、研究生、博士后、本科、专科。。}格标：
成(了)

地点：读+{大学、中学、小学、幼儿园。。}格标：在

时间：读+{一辈子、一天、一个小时。。}

语义映射：获取、动作

4 处理汉语动名超常搭配总体思路

研究表明，动名超常搭配一般含有隐喻和省略两种情形，我们要做的就是如何对这两者进行区分，并分别进行处理。对于省略的超常搭配，我们通过补充成分来还原，而对于隐喻的超常搭配我们通过替换来进行还原处理。

具体处理思路如下图：

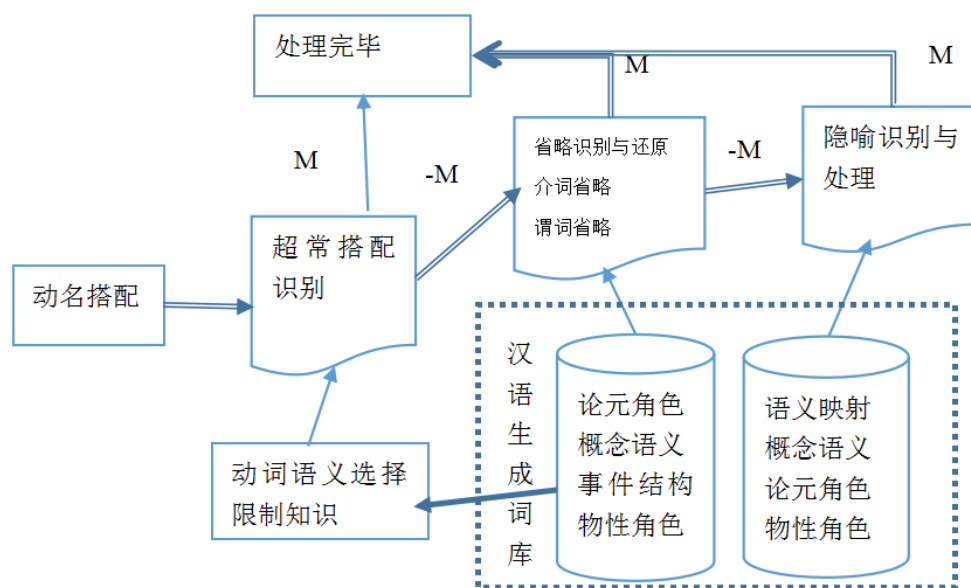


图 2. 汉语动名超常搭配处理总体思路图

处理过程主要包括三个模块：超常搭配识别、省略识别与还原、隐喻识别与处理，这三个呈线性序列。同时，动词语义选择限制知识获取为超常搭配识别提供规则，汉语生成词库为省略型超常规搭配和隐喻型的超常搭配分析提供知识依据。不同的是，隐喻识别和处理是依托词项语义映射等特征的描述，而省略识别与还原则分为介词省略和谓词省略两个流程，介词省略侧重利用论元角色和概念语义特征的组配，而谓词省略则侧重事件结构和物性角色特征的描述和组配。另外汉语生成词库中动词的论元角色和名词的概念语义特征可以帮助生成语义选择限制条件。

5 汉语省略型动名超常搭配的处理

5.1 谓词省略型动名超常搭配处理

首先我们从动词入手，如果该动词是事件动词就具有后接VP的能力，才能有谓词省略的可能以及获得动词所接VP的知识，从而形成潜在性的VP搭配集合。接着就要从名词入手，利用名词物性角色中对相关动词的描述与中心动词的VP集合进行匹配。我们可以以“弹钢琴”和“学钢琴”为例，“弹钢琴”的分析模式如下图：

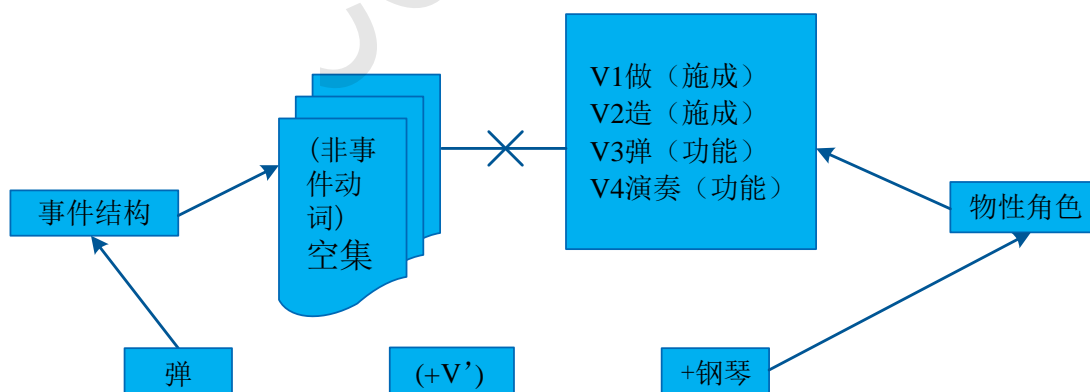


图 3. “弹钢琴”的分析流程图

这里，“弹”通过动词知识库识别出非事件动词，因此不具备后接VP的能力，因此不会存在谓词省略。

而“学钢琴”的分析模式如下图：

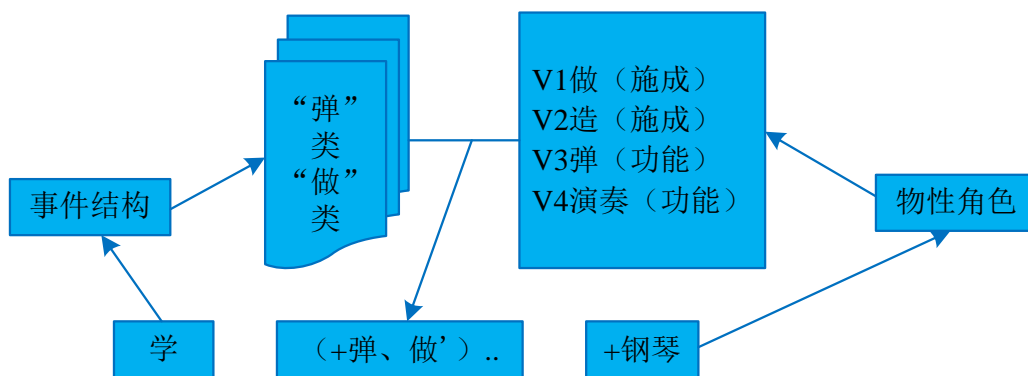


图 4. “学钢琴”的分析流程图

这里“学”通过动词知识库识别出为事件动词，根据后接VP词集知识，再和“钢琴”的物性角色中的施成角色和功能角色的描述匹配，可以得出两类动词省略。

5.2 介词省略型动名超常搭配处理

首先我们依托名词的概念语义特征，对照该动词的语义角色框架来判断语义关系类型，然后参考改语义关系所涉及的格标，如果有对应格标，则进行添加还原，如果没有则不视为带有介词省略。

我们以“洗温泉”和“洗衣服”为例。“洗”的基本式为“施事（人）+V+受事（具体物、人造物等）”，扩展式为“V+地点”、“V+时间”、“V+方式”等，按照名词“温泉”和“衣服”的概念语义特征，可以判断“洗温泉”为扩展式“V+地点”。而“洗衣服”为基本式“V+受事”，基本式为常规搭配关系，所接名词处于最优势句法位置，一般格标记为零形式，不存在介词省略，故不需要添加还原。因为“地点”的格标记为“在”，则我们可以将“洗温泉”处理为“在温泉（里）洗”。具体方位词是“里”还是“上、下”，需要进一步借助名词是否具有容器性特征来判断，这个可以包括在名词的概念语义特征中。具体流程如下图所示：

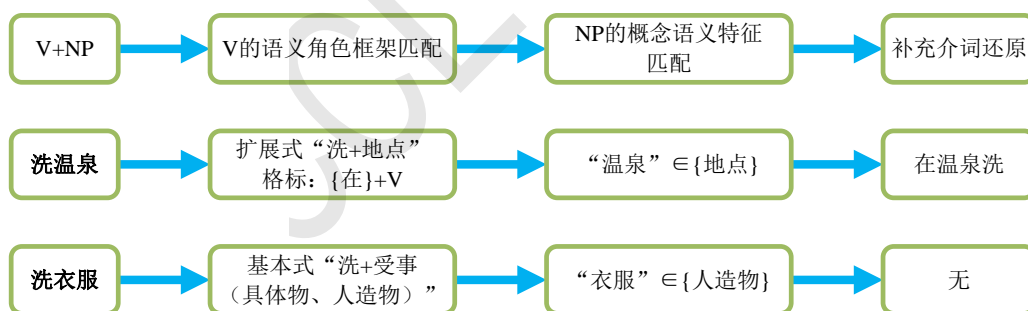


图 5. 介词省略型处理流程图

6 汉语隐喻型动名超常搭配的处理

对于一个隐喻型超常搭配“V+NP”，我们先从动词的语义映射特征着手，找出动名搭配中动词所对应的映射特征V’（主要表现的语法形式是动词）。接着我们从名词出发，找出名词对应的语义映射特征NP’（这时主要语法表现形式也是名词），然后再通过映射对象中名词的物性角色分析（尤其是功用角色和施成角色，因为这两个角色的表现形式主要是动词）看是否能够蕴涵动词V’，如果蕴涵，则可以利用一种新的常规的动宾搭配来进行解读。

具体处理思路，如下图所示：

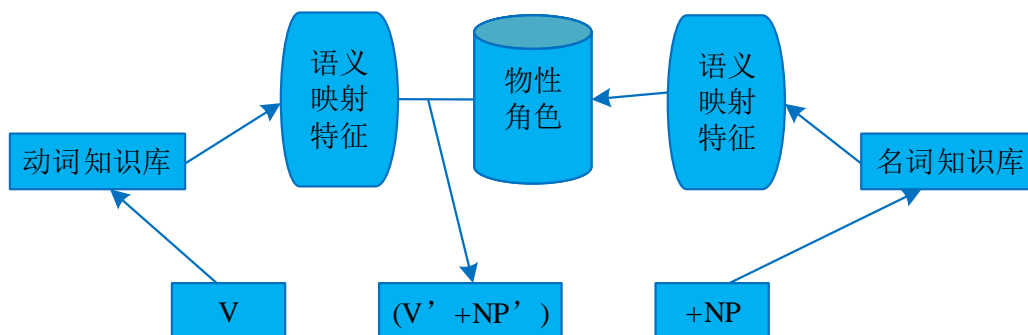


图 6. 隐喻型超常搭配分析流程图

我们以“读硬盘”为例：

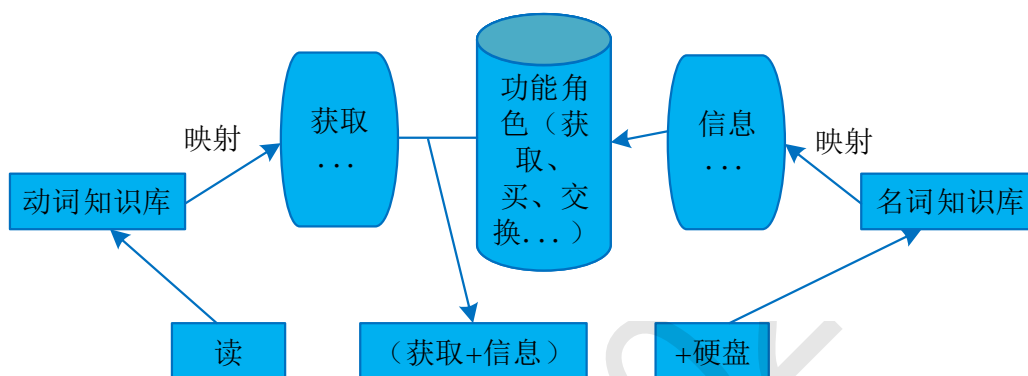


图 7. “读硬盘”分析流程图

这里动词“读”和“获取”可直接通过语义进行关联，从而建立语义映射；但是名词间的语义映射关联，主要还是通过物性角色。比如“硬盘”和“信息”的关联，要靠功用角色。“硬盘”功能是“存储”，而“存储”又可以激活名词“信息”或“数据”。这种动词的激活性特征已经得到了心理实验的验证。如Warren&McConnell（2007）、Altmann（1999）。实验表明，有些动词自身就包含有一些特定的名词信息，在激活某一个动词时，自然会激活相应的名词。比如“悼念”自然会机会其对象“死人”。所以，人们由“硬盘”激活“信息”或“数据”并建立映射关联是合乎心理的。换句话说“读硬盘”，我们可以理解为“读硬盘存储的信息”，而“读”之前我们已经通过语义关联映射为“获取”，所以“读硬盘”解读为常规搭配的形式就是“获取信息”。

7 实验验证

我们利用上述处理思路，从北京大学CCL语料库现代汉语分库中随机抽取了5000个动名搭配，根据动词的语义选择限制条件，筛选出425个动名超常搭配，经过人工判断其中410个识别正确，在剩余的4575个常规搭配中，又发现50个动名超常搭配，总体上超常搭配的所占比例不到10%，利用语义选择限制所识别的超常搭配识别的正确率为86.3%。正确识别并处理的省略型超常搭配30个，正确识别并处理的隐喻型动名超常搭配是320个。同时，我们增加了句法分析中所提取的动名超常搭配，将动名超常搭配基数扩展到了1000个。在处理过程中进行了省略型和隐喻型的区分处理，具体数据如下：

表 2. 省略型动名超常搭配和隐喻型动名超常搭配处理结果

	准确率	召回率	F值
省略型	94.26	94.53	94.51
隐喻型	86.5	86.84	86.72
全部	90.38	90.685	90.62

从这个结果中，我们可以看到省略型的处理效果要好于隐喻型，主要有两方面原因，一是因为，从数量上看，隐喻型动名超常搭配所占比例比省略型动名搭配高的多，近90%的动名超常搭配是隐喻型的，处理基数多会影响准确率。二是因为对于隐喻类的动名超常搭配处理，所涉及的动词和名词的语义关联特征较多，且都是一对多的关系，我们在构建汉语生成语义知识库时，对词项知识的描述难免会存在偏差，最终影响了处理结果。而省略虽然关联特征较多，但是其中又大多数是介词省略型的超常搭配，所依据的语义特征仅仅需要靠动词的语义角色框架信息和名词的概念语义信息匹配就可以了，这些语义信息都是传统语义资源经常构建的信息，因此数据资源较好，所以其处理的准确性也较高。另外，虽然目前信息学界大都没有将动名超常搭配进行分类处理，但是在CCL2018隐喻测评任务中，第一名所使用的HQU模型对隐喻的识别F值为83.27，这说明本文的方法还是具有一定优势。

部分动名超常搭配还原和处理结果如下：

表 3. 部分动名搭配处理结果

隐喻型动宾超常搭配处理				省略型动宾超常搭配处理			
原搭配	处理后	原搭配	处理后	原搭配	处理后	原搭配	处理后
吃救济	依靠救济	拉家常	获取日常信息	踢正步	用正步踢	禁止烟火	禁止（产生）烟火
吃回扣	获得回扣	露心声	展现想法	奔月球	向月球奔	避免病毒	避免（感染）病毒
吃掉敌军	消灭敌军	拿主意	做决定	唱高调	用高调唱	准备桑拿	准备（洗）桑拿
打瞌睡	do瞌睡	剖析问题	分析问题	鸣春涧	在春涧鸣	推迟运动会	推迟（举办）运动会
打副手	从事副手（职务）	送秋波	给予爱意	抽烟斗	用烟斗抽	学电脑	学（修/用）电脑
打主意	获得想法	送温暖	给予关怀	坚持中药	坚持（服用）中药	学挖机	学（开/修）挖机
读硬盘	获得（硬盘里的）信息	讨生活	获得生路	哭长城	为/在长城哭	要求正装	要求（穿）正装
读眼神	感知（眼神里）信息	挖数据	获得数据	跑资金	为资金跑	预料大雪	预料（下）大雪
扶贫	消除贫困	洗清秋	消除秋意	走西口	向西口走	赶材料	赶（写）材料
咀嚼文章	感知（理解）内容	攒人品	获得运气	战长沙	在长沙战斗	渴望金钱	渴望（获得）金钱
啃老	依靠长辈	啃老本	依靠原有条件	睡公园	在公园睡	开始论文	开始（写）论文
拥抱阳光	感知阳光	饮尽风雪	获得苦难	洗温泉	在温泉里洗	开始电话	开始（打）电话

8 结语

我们通过构建汉语知识库的方式还原汉语动名超常搭配的处理方法，是一种解构分析的处理思路，不同于以往把超常搭配整体纳入到词库的思路。本思路基于生成词库理论框架，除了能处理隐喻型超常搭配，还能够处理省略型超常搭配，并且再下分为介词省略和谓词省略两种类型进行针对性描述和分析。对动名超常搭配的处理和还原有利于展示动名搭配更细致的语义隐含信息，为诗词和歌词的处理，以及句法分析提供参考。实验表明，在具体超常动名搭配中，隐喻型的动名搭配所占比例较高，识别和处理的准确率不及省略型动名搭配，不过整体还

原处理效果较好,这一方法的切实有效是以词库中高质量的词项语义知识描述为前提,因此对构建汉语生成词库提出了更高要求。

参考文献

- Ekaterina Shutova, Lin Sun, Anna Korhonen. 2010. Metaphor identification using verb and noun clustering. In *Proceedings of the 23rd International Conference on Computational Linguistics*, 1002–1010.
- Gerry Altmann. 1999. Thematic role assignment in context. *Journal of Memory and Language*, 41(1):124–145.
- George Lakoff, Mark Johnson. 1980. *Metaphors We Live by*. Chicago: The University of Chicago Press, (1):85–90.
- James Pustejovsky. 1995. *Generative Lexicon*. Massachusetts: MIT Press, Cambridge.
- James Pustejovsky. 2006. Type Theory and Lexical Decomposition. *Journal of Cognitive Science*, 39(6):39–76.
- Tessa Warren, Kerry McConnell. 2007. Investigating effects of sectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic bulletin & review*, 14(4):770–775.
- Yorick Wilks. 1975. A Preferential Pattern-seeking Semantics for Natural Language Inference. *Artificial Intelligence*, 6:53–74.
- Zachary J. Mason. 2004. CorMet: A Computational, Corpus-Based Conventional Metaphor Extraction System. *Computational Linguistics*, 30(1):23–44.
- 白振凯. 2016. 基于主体模型的汉语动词隐喻识别方法研究. 杭州电子科技大学硕士论文.
- 戴帅湘, 周昌乐, 黄孝喜, 杨芸, 王雪梅. 2005. 隐喻计算模型及其在隐喻分类上的应用. *计算机科学*, 32(5):159–163.
- 黄洁. 2009. 动宾非常规搭配的转喻和隐喻透视. *同济大学学报 (社科版)*, (1):85–90.
- 贾玉祥, 俞士汶. 2011. 基于词典的名词性隐喻识别. *中文信息学报*, (2): 99–104.
- 李剑锋. 2008. 面向隐喻计算的汉语语义超常搭配识别模型研究. 厦门: 厦门大学硕士论文.
- 吕叔湘. 1979. *汉语语法分析问题*. 北京: 商务印书馆.
- 孙梦荀. 2012. 论现代汉语中词语的超常搭配及对外汉语教学. 苏州大学硕士论文.
- 汪梦翔, 王厚峰, 刘扬, 饶琪. 2014. 动词资源馆 (Chinese Verb Library) 的构建. *中文信息学报*, 28(6):85–94.
- 汪梦翔, 饶琪, 顾澄, 王厚峰. 2017. 汉语名词的隐喻知识表示及获取研究. *中文信息学报*, (6):1–9.
- 王贵荣, 饶高琦, 荀恩东. 2021. 基于大规模语料库的现代汉语动宾搭配知识库构建. *中文信息学报*, 35(1):34–42.
- 王金锦, 杨芸, 周昌乐. 2009. 隐喻字面语义表示与生成. *中文信息学报*, 23(3):95–102.
- 王治敏. 2008. 名词隐喻相似度及推理识别研究. *中文信息学报*, 22(3):37–43.
- 王治敏. 2009. 汉语名词隐喻的语义映射分析. *语言教学与研究*, 3:89–96.
- 吴琼, 周保国. 2014. 汉语非常规动宾搭配理解层级性的实证研究. *汉语学习*, (6):91–96.
- 游维, 周昌乐. 2007. 基于统计的汉语隐喻生成模型及其系统实现. *心智与计算*, (1):133–141.
- 杨芸, 李剑锋, 周昌乐, 黄孝喜. 2008. 基于实例的汉语语义超常搭配的自动发现. *计算机科学*, (9):195–197.
- 俞士汶, 朱学锋, 王惠. 2003. *现代汉语语法信息词典详解(第二版)*. 北京:清华大学出版, 19–136.
- 于江生, 俞士汶. 2002. 中文概念词典的结构. *中文信息学报*, 16(4):13–21.
- 袁毓林. 2014. 汉语名词物性结构的描写体系和运用案例. *当代语言学*, 16(1):31–48.
- 张威, 周昌乐. 2004. 汉语隐喻理解的逻辑描述初探. *中文信息学报*, 18(5):23–28.