

# 隐喻化新词的考察\*

李惠, 冯敏萱

南京师范大学文学院 南京 210046

E-mail: lh9743@sohu.com

**摘要:** 隐喻是通过一种领域内的词汇来描绘另一种领域内的现象, 目的是为了理解、认知新事物。它是创造新词新义最经济的手段。隐喻化新词是以一定的语音形式用已存在的概念去表达新事物的词语。随着社会的发展, 这样的新词大量出现并成为一种社会现象存在。这是一个以联想为基础, 由源域向目标域映射的过程。本文从隐喻角度出发, 基于搜狐语料, 对新词语的语义特点和结构特征进行分析, 并建立规则库, 提出未登录词的识别和获取方法, 设计实验, 制订隐喻新词词表并实现隐喻化新词的自动识别。

**关键词:** 新词, 隐喻, 自动识别

## Investigation on Metaphoric Neologism

Li Hui, Feng Minxuan

Department of Chinese Linguistics, Nanjing Normal University, Nanjing 210046

E-mail: lh9743@sohu.com

**Abstract:** Metaphor is widely used in our daily life to describe A for B via words and phrases. In order to recognize and understand the emerging things, the most economic way to create new words and new meanings, in my opinion, is metaphor. We apply the phrase "metaphoric neologism" into definition a unit with certain phonetic symbols, based on the existing concepts, to demonstrate the fresh. Currently there is no doubt that flourish metaphoric neologism will appear continuously, and we define this rising trend as a mapping process from source to target.

In this paper, based on the materials from Sohu, we intend to make full use of metaphor to discover the semantic characteristics, digest the structure of the novel words, built the corpus of rules, collect the neologism, and most importantly, burn the fire with the idea of automatic identification by our delicate experiment.

**Keywords:** neologism, metaphor, automatic identification

### 1 引言

2010年百度百科联合“南都周刊”发布十大新词“甲型H1N1流感、躲猫猫、钓鱼执法、临时性、被时代、蜗居族、压力差、70码、穷二代、翻墙”。在当今的信息时代, 新词语广泛地出现在聊天、论坛等各种网络交流场合, 并渗透到现实生活的方方面面。无论是衣食住行还是政策体制, 人们在创造了一批词语的同时, 也赋部分旧词以新义, 来表达思想。目前对新词语的研究主要集中在新词语的外形特点和语法构成的单一考察, 没有结合新词语的隐

---

\* 本论文得到南京师范大学特聘教授、高层次人才科研启动基金项目(编号:0609025)的支持和资助。

喻性语义特征。其实新词语的出现有一种很重要的渠道就是用“隐喻的手段赋原词以新义产生的，因而语义更容易扩展和泛化”（江傲霜，2004，pp54）。

研究新词语的隐喻性语义特征及影响，对于新词语的自动识别具有一定的意义。教育部在《中国语言生活状况报告》(2006)中，列出了171条汉语新词语选目，“奔奔族、独二代、国际高考移民、梨花体、晒客、学术超男”等词收录其中。以“奔奔族”为例，“族”原指具有共同起源和共同遗传特征的人群，现出现在新词“乐活族/急婚族/慢活族/洋漂族/装嫩族”等中指具有同种生活态度或生活方式的人，不一定聚集成群，根据这种隐喻特征（“一种生活态度/生活方式+族”），可以归纳出一定的结构，并据此进行识别可能出现的组合方式，并加以实验证明之。

## 2 隐喻化新词语的界定

人们在社会中运用语言，语言也从一定程度上反映着社会。当今社会的迅猛发展，体现在语言上最显著的就是新词语的产生和广泛使用。人们不仅创造了大量新词，也赋予了一部分旧词以新的含义，这就离不开我们赖以生存的隐喻。朱红雷（2008，pp72）谈到“隐喻认知从源域(具体域)到目标域(抽象域)的投射是构成词义的重要手段。翻开词源词典我们会找到很多词的理据。新词的构成绝大部分是通过这种直接或间接的投射得到的”。我们在阅读了大量关于隐喻和新词语的文献之后，给出了隐喻化新词语的定义：以一定的语音形式用已存在的概念去表达新事物的词语。这类词语一般都只有两音节或三音节，但形象生动，言简意赅，它们往往蕴含了丰富的信息，和当今的社会现象有着紧密的联系。隐喻的实质是借助一类事物理解和体验另一类事物。在日常生活中，人们往往参照熟悉的、有形的、具体的概念来认识、思维、经历和表达无形的、难以定义的概念，形成了一个不同概念之间相互关联的认知方式，也就是说，人们在隐喻的基础上进行再创造得到了新词。这些隐喻化新词不仅反映了语言学的经济原则，同时也反映了人们求新求异的语言文化心理和当今社会对幽默风趣的崇尚。

## 3 语料分析

### 3.1 语料资源介绍

在新词的识别中，需要的语料和汉语分词工具，分别介绍如下：

- (1) 语料库：从搜狐新闻网页下载的09年9月至10年2月半年的语料，大小为433K。实验以搜狗细胞词库2007-2009年全年新词，分词正确率达到100%，大小约为2.3M，作为提取新词识别规则的基础。搜狗的细胞词库涵盖了生活、娱乐、运动休闲、艺术、医学、自然科学、人文科学、社会科学等领域的大量新词，平时接触到的方方面面几乎都会有所涉及。
- (2) 分词工具：中科院汉语词法分析系统ICTCLAS 1.0 version，实验在经过ICTCLAS自动分词和词性标注的语料上进行。

本文所讨论的新词语的范围为近两年来出现的新词语。主要采取语料调查与实验分析相

结合的研究方法。通过语料收集词语，对之进行考察分析，探求其中的规律。通过搜狗细胞词库，笔者已经收集了 86,076 条新词语，从中发现了一些现象可以通过隐喻的知识来加以分析理解。

### 3.2 字族化现象

李宇明（1999, pp148）指出：“大多数新产生的词语，都有一个现成的框架背景，这一框架背景就像是造词的模式一样，能批量生产新词语，并使其所生产的新词语形成词语族”。宋作艳（2007, pp102）也认为“新词往往不是单个出现，而是围绕一个相同的字（核心字）成群出现，形成一个字族，字族的成员之间存在平行关系”。我们在分析语料的过程中，发现了很多“X族”、“X门”、“X奴”的字族化结构，举例说明：

门：原指建筑物的出入口，又指安装在出入口能开关的装置，现组合时多为贬义，表示一种不好的社会现象或风气，可以组合成“电话门/监控门/解说门/骷髅门/回扣门/鬼脸门/集邮门/录音门/扔鞋门/偷笑门/黑幕门/告白门”等等。

被：原用作助词、介词，无明显褒贬色彩，现组合时含有的是更多的无奈和对现实的调侃，例如：“被就业/被增长”等，这种以“矛盾修辞”方式组合而成的新词语清晰地传递出人们的复杂心态。

类推是语言发展和演变的一种重要机制。人们利用已有的字族类推出词语，这样既不会额外增加新的构词模式，又可沿袭已有字族词的影响，增强了新词的生命力。通过对这些字族的隐喻分析，我们可以得出它们的语义特征，对未来出现的新词，我们可以把字族的前缀和后缀的意义与字族的语义通过一定的模式相连，就可以得出新词的词义。

### 3.3 委婉语现象

语料中也出现了许多起委婉作用的新词，它们往往带有很浓厚的幽默意味。这可能有两方面的原因。“一方面因为随着社会的进步与发展，人们在语言表述上也更加追求文雅”（江傲霜，2004, pp55）；另一方面是对当前一些政策体制的一种不满的宣泄，他们无法直接地去批驳，只能通过谐音委婉的方式来表达。基于这两点原因，人们便将原有词隐喻化来满足表达上的不同需求。例如：

杯具：原指盛水的器具，后因与“悲剧”一词谐音，不少年轻人在网络上甚至生活中都常常用“杯具”来代替“悲剧”，形容人、事、物。该词现多用于主观地表示不如意，不顺心或者失败，或者是委婉地对别人表示某方面的不满，主要靠意会。一般戏谑的味道比较强。并由此引申出“洗具”、“餐具”等。

宅：原指住所或房子，后因当今社会的快节奏和高压力，很多人选择不与外人直接交流，待在家里通过网络虚拟交流，习惯性地逃避现实。

通过以上新词使用时种种形象化的比喻，不仅表述上更加委婉，而且生动形象。人们使用这种异化的生活用语进行交流，会有参与感和被认同感，同时也能感受到这种语言所带来的娱乐感。这也是隐喻化的委婉语迅速为人们所用的主要原因。

## 4 隐喻化新词的自动识别

## 4.1 识别方法

我们迄今收集的新词，大多数还是遵循构词法产生的，比如字族化现象，新词的构成方式都是围绕着一个中心词展开的。我们想使用统计和规则相结合的方法来进行新词识别实验。

### 4.1.1 统计的方法

#### (1) 语素生产率

由字族化现象，我们可以推测隐喻化新词中有些汉字有特有的构词用法，出现在某个或某几个特定位置上。像“族”通常出现在词末，“被”常常出现在词头。因此可以用语素生产率 (Morphological Productivity, MP) 来描述这一特点。计算公式如下：

$$MP = n1/N$$

“其中，N 表示一个特定范畴在特定文本中的符号数量；n1 表示文本中只出现一次的词形数量，就是那些仅仅用一个符号表示的词形，MP 值较高的词语容易派生出新词” (秦浩伟, 2004, pp370)。例如“X 虫”的结构在文本中共出现 200 次，但词频为 1 的词形只有 10 个，那它的语素生产率为：10/200=5%。本文中，我们将此公式中 n1 引申为文本中只出现一次的隐喻化新词的词形数量，这样可以更好地说明隐喻化新词的能产性。

#### (2) 互信息

互信息 (MI) 可以用来描述两个词之间的相互依赖程度。这里我们用互信息来反映词语间结合的紧密度，当紧密度 MI 大于某个阈值 t1 时，就判定两词结合为未登录词。

互信息的计算公式为：

$$MI(w1, w2) = \lg P(w1, w2) / (P(w1) * P(w2)) = \lg (f(w1, w2) / (f(w1) * f(w2)))$$

P(w1, w2) 表示由相邻两个碎片 w1 和 w2 在测试文本中出现的概率；P(w1) 表示碎片 w1 在测试文本中出现的概率；P(w2) 表示碎片 w2 在测试文本中出现的概率，在这里我们用频率来代表概率。MI(w1, w2) 值越大，紧密程度越大；MI(w1, w2) 值越小，紧密程度越弱。

### 4.1.2 规则的方法

基于上文所分析的新词隐喻特征，我们想建立隐喻化新词的规则库，下面是规则示例：

设 A、B、C、D 代表四个任意汉字，则二元组字串可表示为 AB、三元组字串可表示为 ABC、四元组字串可表示为 ABCD。

就字族化现象而言，很多新词都是以三元组的方式出现，所以我们的实验字串定为三元组：

```
① X 门: while(C=='门'){
    if((tagnum(AB)==1) new(ABC)=2;
    else if(wordtype(A)=='N' && wordtype(B)=='N')
        new(ABC)=1;
    else if(wordtype(A)=='V' && (wordtype(B)=='A' || wordtype(B)=='N'
        || wordtype(B)=='V')) new(ABC)=1;
    else if(wordtype(A)=='A' && wordtype(B)=='N')
        new(ABC)=1;}}
```

该规则表示：如果抽取的“X 门”中，AB 只有一个词性标记的话，就做个标记，

等到人工校对的时候一一筛选；A 为名词时 B 为名词，或者 A 为动词时 B 为名词或动词或形容词，或者 A 为形容词 B 为名词的话，则将该短语做标记，认为是隐喻化新词。

```

② 被 X: while (A=='被') {
    if((tagnum(AB)==1) new(ABC)=2;
    else if(wordtype (B)=='V' && (wordtype(C)=='N' || wordtype (C)=='V' ||
wordtype (C)=='A')) new(ABC)=1;}

```

该规则表示：如果抽取的“被 X”中 B 为动词时 C 为名词或动词或形容词，则将该短语作标记，认为是隐喻化新词。

我们也意识到，仅靠词性序列去判断新词是不够的，如防盗门，走后门等，符合词性序列的要求，但不属于我们的新词范畴了。于是我们同时将现代汉语词典（第 5 版）所收录词语组成词表（见附录），对九个字族进行过滤，如果规则判定的新词与现汉有匹配，则去掉它的新词标记。

## 4.2 实验结果

本文主要通过准确率（P）、召回率（R）和综合指标（F）来评测新词识别的性能，计算公式如下：

$P = \text{识别正确的新词数目} / \text{识别出的新词数目}$

$R = \text{识别正确的新词数目} / \text{语料中出现的新词数目}$

$F = 2 * P * R / (P + R)$

经过去标签，提取正文的预处理后，半年的新闻语料共 728,472 字，我们已通过 ICTCLAS 分词工具将其自动分词并进行词性标注，再使用 JAVA 程序根据正则表达式抽取，再根据上述相关规则筛选，人工校对，我们选取一部分得到的字族化新词列表所示（词后加\*为规则无法识别的隐喻化新词）：

表3. 2-1

X族	词频	X门	词频	被X	词频	X奴	词频
奔奔族	41	相声门	16	被残疾	3	房奴	3
快闪族*	9	电话门	14	被增长	1	车奴	1
啃老族	3	代言门	4	被代表*	1	卡奴	1
钟摆一族*	2	间谍门*	3	被就业	1	...	...
蚁族*	2	芯片门	2	被被捕*	1	...	...
80后一族*	1	牌坊门	1	被中产*	1	...	...
月光族	2	窃听门	1	被被了*	1	...	...
...	...	...	...	...	...	...	...
文本中“X族”的词语数目	64	文本中“X门”的词语数目	130	文本中“被X”的词语数目	351	文本中“X奴”的词语数目	6

在实验中，我们分别对字族化新词和谐音委婉词进行了统计计算（在这里语素生产率阈值为 0.2，互信息阈值取 2.50），得到以下结果：

实验条件	X族	X门	被X	X奴
语素生产率	0.078125	0.253846	0.931624	0.50
规则识别率（识别正确的隐喻化新词数目/文本中出现的隐喻化新词数目）	70.31% (45/64)	92.68% (38/41)	55.56% (5/9)	100% (5/5)
实验条件	甲流	杯具	洗具	剩女
互信息	3.9945	4.2164	3.2463	3.1137

可以看出，字族化新词的语素生产率除“X族”之外，普遍高于阈值，也证明了隐喻化新词的能产性；而互信息的测试结果，说明了谐音委婉语结合的紧密程度，我们可以将这些高于阈值的结构标记为一个隐喻化新词。

根据我们的规则匹配原则，得到以下评测结果：

识别正确的新词数目	系统判为新词的数目	文本中出现的新词数目	准确率	召回率	综合评价
97	155	118	62.58%	82.20%	71.06%

由上表可知，综合评价只有71.06%，不尽人意，因此我们又引入了上文所说的词典过滤的方法，根据词表和规则，我们的统计结果如下所示，准确率超过了85%，效率有了大幅度的提高：

识别正确的新词数目	系统判为新词的数目	文本中出现的新词数目	准确率	召回率	综合评价
97	109	118	88.99%	82.20%	85.46%

#### 4.3 错误分析

上述实验中，我们也发现了一些问题。例如 X 族的语素生产率比预期的要低，但“奔奔族”这个词出现了 41 次，占“X 族”总条数的 64.06%，而且所有出现的“X 族”均为隐喻化新词，可见 X 族的能产力不低。为了解释一点，我们又选取了百度搜索引擎进行了测试，我们对前 10 页的搜索结果中的 100 个新词进行了规则鉴别和词频统计（不包含字母组合），除了“恐归族/上班族/蚁族/淘乐族”等重复出现了两到三次，其余都只出现了一次，语素生产率达到了 80% 以上。可见只要将语料的范围扩大，不仅是新闻，加入论坛帖，广告或聊天记录，语素生产率可以大幅度地提高。

试验中，存在一些隐喻化新词，规则并没有识别出来。以“X 族”为例，在开放测试中，我们发现“X 一族”的形式经常出现，但在类似的“X 门”、“X 奴”等中却没有这种说法，这是一个很特殊的现象，“X”族的规则必须细化，我们可以将“X 一族”考虑成四元组，对第三位为“一”，第四位为“族”的词作标记，结合人工校注，取得较好的获取效果。再看“被 X”族结构，

严格说它是一个短语,但现在我们通常把它当一个词用,在开放测试中,我们发现了“被被 X”的结构,这种结构违背常理但真实存在,我们同样可用“X一族”的方法来细化规则。

对于一些谐音新词,我们无法预测它将以怎样的形式出现,但我们可以再下一步的工作中用语音语义结合的方法做一些实验。我们想构建一个汉字拼音对应的VFP数据库,基于王敬农(1999, pp7)的思想。该数据库包括三个表:拼音表,单音字表,多音字表。我们希望可以构建该数据库,查询所有的同音词,再通过词典过滤加规则的方法标注其中的隐喻化新词。

### 参 考 文 献

1. 陈保亚,宋作艳,邱立坤,陈泽浩,密雪飞. 网络文本中的词汇自动识别[A]. 科学中国人. 2003,(11).
2. 戴师湘,周昌乐,黄孝喜,杨芸,王雪梅. 隐喻计算模型及其在隐喻分类上的应用[A]. 计算机科学. 2005,(32).
3. 韩艳,姚建民,朱巧明,张晶. 不限领域的中文新词的识别研究[A]. 郑州大学学报(理学版). 2008,(9).
4. 贺敏,龚长春,张华平,程学旗. 一种基于大规模语料的新词识别方法[A]. 计算机工程与应用. 2007,43(21).
5. 江傲霜. 新词语隐喻化的特征及方式[A]. 红河学院学报. 2004,(12).
6. 李宇明. 《词语模》. 载邢福义《汉语语法特点面面观》[M]. 北京语言文化大学出版社. 1999.
7. 林玲. 汉语网络新词的判定及造词方式[A]. 成都大学学报(社科版). 2008,(2).
8. 刘红妮. “X奴”族新词及其社会心理[A]. 汉字文化. 2008,(4).
9. 刘晓梅. 当代汉语新词语研究[D]. 厦门大学博士学位论文. 2003,(6).
10. 刘宗保. “X门”词族探微[A]. 现代语文. 2006,(9).
11. 秦浩伟,步丰林. 一个中文新词识别特征的研究[A]. 计算机工程. 2004,(12).
12. 宋作艳. 字族化与汉语未登录词的自动提取[A]. 北京大学学报(哲学社会科学版). 2007,(3).
13. 苏东华. 当代新词语修辞现象词汇化研究[A]. 暨南大学硕士学位论文. 2006,(6).
14. 王敬农,杨威. 汉字的拼音自动转换[A]. 计算机时代. 1999,(3).
15. 杨绪明. 网络“客”族新词及其社会文化心理探析[A]. 西南民族大学学报(人文社科版). 2009,(1).
16. 叶蜚声,徐通锵. 语言学纲要[M]. 北京大学出版社. 1997.
17. 张谊生,许歆媛. 浅析“X客”词族——词汇化与语法化的关系新探[A]. 语言文字应用. 2008,(11).
18. 郑家恒,李文花. 基于构词法的网络新词自动识别初探[A]. 山西大学学报(自然科学版). 2002,(25).
19. 朱红雷. 网络新词的隐喻研究[A]. 郑州航空工业管理学院学报(社会科学版). 2008,(6).
20. Martin J H. A Computational Model of Metaphor Interpretation. Boston. Academic Press. 1990.
21. Nehaniv C L. Computation for Metaphors, Analogy, and Agent. Springer. 1999.