汉英词汇隐喻属性的对比分析与互增益技术*

匡海波[1]，李 斌[1,2]，王嘉灵[1]，王 帅[1]，陈小荷[1]
[1]南京师范大学 文学院，江苏 南京 210097
[2]南京大学 计算机软件新技术国家重点实验室，江苏 南京 210093
E-mail：mchypocn@hotmail.com

摘 要：本文基于隐喻认知观和词语属性分析理论，利用网络数据挖掘技术，构建了基于《知网》语义体系的汉英双语词汇隐喻属性知识库，进行跨语言系统对比分析词汇隐喻属性。通过研究跨语言词汇隐喻属性的异同，用量化统计和系统分析初步地回答了隐喻的否跨语言特点，本文同时进而提出了利用双语知识库，以一种语言的词语隐喻来增益研究对译词语隐喻属性的方法，为基于隐喻属性的语义分析计算打下了一定的研究基础。

关键词：隐喻；属性分析；映射关系；互增益

The Comparative Analysis and the Technology of Mutual-gain on Lexical Metaphor Properties of Chinese-English

Kuang Haibo[1], Li Bin[1,2], Wang Jialing[1], Wang Shuai[1], Chen Xiaohe[1]
[1] Department of Literature, Nanjing Normal University, Nanjing 210097
[2] Department of Computer Science & Technology, Nanjing University, Nanjing 210093
E-mail: mchypocn@hotmail.com

Abstract: Based on the metaphoric conition and feature analysis theory, we acquires data from the web to constructs the Chinese-English bilingual lexical metaphor properties knowledge base linked to "HowNet". By comparing the differences of the metaphoric properties, we get some answers to the core linguistic problem "is the metaphor universal?" Then, we put forward a novel method to gain the metaphoric properties of a word by its translation in another language. The paper lays a solid foundation for semantic analysis and computing of lexical metaphoric properties in the future.

Keywords: metaphor; property analysis; mapping relationship; mutual-gain

1 引言

词汇隐喻属性体现了语言使用者在特定环境中对词汇隐喻的认知体验，跨语言的词汇隐喻属性比较体现了不同语言使用者的词汇隐喻认知异同，这两者皆为语言学、认知科学等多学科的研究焦点。然而，传统研究往往囿于具体词汇的隐喻属性列举和比较，基于大规模语料库的词汇隐喻属性研究还不多见，系统的跨语言比对分析研究更为鲜见。

因此，本文将基于隐喻认知观和词语属性分析理论，以《知网（HowNet）》(董振东，董强)为语义支撑，利用数据挖掘技术，构建汉英双语词汇隐喻属性知识库，并对库中词汇及属性进行多角度、互增益分析。实验表明，汉英双语词汇隐喻属性知识库的构建，为系统的跨语言词汇隐喻属性对比分析提供了数据基础，为回答跨语言词汇隐喻认知机制的异同提供了量化统计。知识库的完善将有利于跨语言隐喻认知研究和隐喻语义计算等领域。

2 理论背景

在众多隐喻认知观学说中，Bowdle 与 Gentner 提出的隐喻生涯假说(the career-of-metaphor hypothesis)认为，词汇隐喻的认知机制往往表现为隐喻属性的规约化程度：规约化程度较低的新异

* 本文承国家社科基金 10CYY021、07BYY050，南京大学计算机系重点实验室招标课题 KFKT2011B03，国家自然科学基金 61073119 的资助。

隐喻的理解通过比较机制；规约化程度较高的公认隐喻的理解通过属性机(Bowdle&Gentner，2005；Gentner&Bowdle，2001)。因此，通过观察词汇隐喻属性的规约化程度，我们可以在一定程度上基于生涯阶段理解该隐喻的认知机制。

另一方面，词语属性分析在语义形式化的研究中扮演重要角色（陈小荷，2005），对词汇隐喻属性进行形式化分析也很有必要。我们也认为，利用较完备的语义体系对词汇隐喻属性进行形式化分析，方便了词汇隐喻属性间的映射及同义、近义、对义等关系的抽取，有利于基于词汇隐喻属性的语义计算。

我们将词汇隐喻属性定义为如下形式，并且和《知网》语义体系作了映射：{喻体（喻体概念表达式）：[映射关系]喻体-隐喻属性 1（隐喻属性概念表达式）；[映射关系]喻体-隐喻属性 2（隐喻属性概念表达式 2）；…….}

可具体举例如下：{actor/演员 *(DEF={human/人:HostOf={Occupation/职位},domain={entertainment/艺},{perform/表演:agent={~}}})*；[Ironic/反讽]actor-dumb/演员-说不出话的 *(DEF={disable/残疾:scope={MakeSound/发声}})*；[Ironic/反讽]actor-realistic/演员-真实 *(DEF={able/能:scope={fulfil/实现:patient={$}}})*；}…

3 词汇隐喻属性知识库的构建

词汇隐喻属性的采集是一项颇具难度的工作，不过已有学者寻找采集途径。Veale（2006；2010）认为存在着“从明喻机制到隐喻机制的进化性道路”，并通过搜索引擎采集了大量的英语明喻句，以 WordNet 为语义支撑，构建了“名词喻体-形容词隐喻属性”的英语词汇隐喻属性关系库。贾玉祥（2009）用相似的方式，采集了汉语的明喻句，以《同义词词林》为语义体系，构建了“名词喻体-形容词隐喻属性”的词汇隐喻属性关系库。

本文以《知网》为语义体系，采用指定明喻句式（“X 像 Y 一样 P”），通过搜索引擎挖掘汉语明喻句，构建“喻体-隐喻属性”的知识库，并与 Veale 等建立的英语关系库进行比对。实际采集过程中发现，收集到的句式却可能是反讽句（如“他像猪一样的聪明”），因此我们在采集过后，通过两位语言学专业本科生的人工校对，对词汇隐喻属性的映射关系加以区分（目前关系二分为“象征”、“反讽”）。

3.1 英语词汇隐喻属性知识库 sardonicus[1]

Veale 等建立的英语词汇隐喻属性知识库 sardonicus（Veale&Hao，2006）（以下简称 sardonicus）共收集了 74704 个明喻实例句，使 3769 个形容词隐喻属性映射到 9286 个名词喻体上（Veale&Hao，2006），并将词汇隐喻属性的映射关系进行了二值划分（即 Factual：horse-strong；Ironic：ant-strong 等）。

我们对 sardonicus 中的“名词喻体-形容词隐喻属性”词对进行过滤筛选，去掉单纯比较和错误条目，并将喻体及属性分别与《知网》（2007 版）进行连接映射2。统计结果显示，加工后的 sardonicus 在库词对共计 10411 个，其中喻体 3585 个，平均每喻体拥有词对 2.90 个；另外，51%的喻体对应 1 项属性，15.7%的喻体对应 2 项属性，13.3%的喻体对应 3 项属性，而 20%的喻体则对应了 4 个或 4 个以上的属性（见图 1）。由此可以观察到在库喻体属性数目的分布情况。

需要说明的是，我们基于词形将喻体及属性连接到《知网》时，并没有急于进行喻体及属性的“义项（概念条目）消歧”，并且给出了喻体属性词对的汉英语表达，以便跨语言对比分析。

---

[1] Sardonicus 是爱尔兰都柏林大学生成语言系统小组的一个项目，其网址为：http://afflatus.ucd.ie/sardonicus/tree.jsp。

[2] Veale 等将知识库连接 WordNet，我们与知网相连接的理由是其更有利于进行跨语言的比较，并且拥有更丰富完备的语义体系。

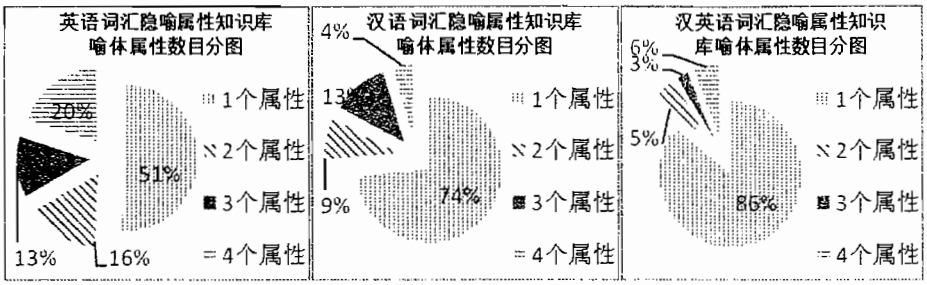

图1　词汇隐喻属性知识库喻体属性数目分布图

目前已连接到《知网》的英语词对共计 6083 个，映射概念条目达 312,042 个，平均每个词对拥有 50.5 个概念条目（具体统计数据见表 1）。实际上，3579 个喻体词型并未见于《知网》，但恰恰可通过喻体属性进行特征分析和概念描述。这与《知网》本身的理念“通过义原解释概念”有异曲同工之妙。

表1　已映射到《知网》的英语喻体属性词对概念条目数分布

概念条目数	0（喻体词型未登录）	0（属性词型未登录）	1~20	20~40	40~	合计
喻体属性词对数	3579	749	2883	1199	2001	312,042

3.2　汉语词汇隐喻属性知识库

为了进行跨语言的对比分析，我们借鉴 Veale&Hao（2006）和贾玉祥（2009）的做法，采用指定明喻句式（“X 像 Y 一样 P”），也进行了大规模汉语数据挖掘，使用规模为 2 万汉语词条的词表，在百度上查询，使用 ICTCLAS 进行分词和词性标注。然后对抽取到的“喻体-隐喻属性”词对也按照象征、反讽进行二值划分。与之不同的是，我们在挖掘过程中保留了词对的频次信息，以期深入对比分析。

需要说明的是，汉语挖掘并没有进行词性限定，而是收集了不同句法性质的词语、词组、甚至句子。例如“像【如何把大象放进冰箱】一样【简单】”这样的复杂喻体也予以收录。据我们统计，目前汉语知识库在库的喻体属性词对共计 4002 个。其中喻体“词型”（这里的“词型”包含了“词型”、“句型”“词组型”等）计 1908 个（如表 2），平均每“词型”采集到 2.09 个属性。

表2　在库词对型数共计 1908 个，词对例数 4002 个

喻体类型	词对型数	词对型数所占比例	词对例数	词对例数所占比例	样例
名词性	1258	65.93%	2779	69.44%	泥土
形容词性	412	21.59%	781	19.52%	以往
动词词性	178	9.33%	361	9.02%	吃苹果
其他词性	60	3.14%	81	2.02%	舒肤佳
合计	1908	100%	4002	100%	

汉语知识库中的喻体拥有属性数目统计见图 1。在汉语库中，81%的喻体拥有 1 项属性，10.18%的喻体有 2 项属性，13.82%的喻体有 3 项属性，只有 5%的喻体拥有了 4 个或 4 个以上的属性。与 sardonicus 比对，不难发现，汉语知识库中喻体属性确定性较 sardonicus 更强，这与我们挖掘范围和数据量稍小不无关系。

如上文所述，我们保留分析了汉语喻体属性词对的频数信息，其中 38%的词对仅出现 1 次，21.1%的词对出现 2 次，26.7%的词对出现 3 次，14.2%的词对出现 4 次或以上。高频词对显然更利于观察隐喻的生涯阶段及理解机制，但也不能只简单对应于基本词频信息。

例如，喻体“白开水”共有三组词对：“白开水-单纯”（频度 4）、“白开水-淡然”（词频 3）、“白

开水-索然”（词频2）。显然，我们不能断言属性“单纯”的隐喻生涯快于“淡然”、“索然”，但也不应忽视“单纯”已经逐渐作为“显著特征”，进入到“白开水”的隐喻属性项中。我们认为，隐喻的生涯阶段及理解机制判断必须与词频、各属性使用频、属性词义关系相结合，单一数据无说服力。

我们将汉语喻体属性词对与《知网》相连接，并给出相应的汉英语表达。统计结果显示，目前在库的词对概念条目达44,495个，平均每词对拥有概念条目36.10个，这里就不一一赘述。

4 汉英双语词汇隐喻属性平行知识库的初步整合

为回答隐喻是否跨语言的问题，我们考虑对已有知识库进行整合，构建新的汉英双语词汇隐喻属性平行知识库（以下简称“平行知识库”），同时保留汉英语知识库，以便跨语言对比分析及互增益研究。

通过比对汉语知识库和英语知识库中的喻体属性词对，我们得到共现喻体属性词对76个（其中包含喻体53个，各喻体拥有共现属性数目统计结果见图1）。显然，我们需要对喻体属性词对是否共现采取人工甄别的措施：如“【反讽】羽毛-重”与“【象征】feather-light”应作为共现词对。

不难看出，平行知识库中喻体的隐喻属性趋于一致（如表4所示）。我们设想，这部分隐喻可能是待证的跨语言交际中的“无障碍”认知成分。事实上，这部分隐喻的属性词逾95.3%属于《知网》基本义原范畴，这也佐证了我们的设想。我们选取了属性相似度最高的10个喻体，展示如表3。另一方面，我们对未收入库的在库喻体进行整理，分别统计出汉英语的独现喻体词表（见表4）。

表3 汉英双语喻体相似度最高Top 10

编号	汉语喻体	英语喻体	相似属性	属性义原举例
1	水晶	crystal	清，清澈，纯，纯净-pure；脆-clear	pure：{spotless|洁}
2	花	flower	新-fresh；甜-sweet；纯真-pure	sweet ： {sweet|甜}
3	妈妈	mother	好-good；温柔，柔和-gentle	gentle ： {gentle|柔}
4	蚂蚁	ant	慢-slow；渺小，小-tiny	small：{small|小}
5	蛋糕	cake	甜美，甜蜜-sweet，luscious	sweet ： {sweet|甜}
6	糕点	cake	甜美，甜蜜-sweet，luscious	sweet ： {sweet|甜}
7	糖	sugar	甜蜜，好吃-sweet，nice	sweet ： {sweet|甜}
8	婴儿	baby	裸-bare，naked	naked：{naked|赤裸}
9	海洋	ocean	宽广，大-broad	broad ： {broad|广}
10	针	needle	锋利-sharp，incisive	sharp：{sharp|利}

表4 汉英语隐喻属性词汇知识库喻体拥有属性数目Top 10

汉语独现喻体

编号	喻体	属性（包含象征、反讽）
1	天气	好，火热，糟阴,沉沉，阴冷，萧条
2	昨天	幸福，开心,无聊，美好，华丽
3	阳光	健康,绝望,明媚，忧伤,透明
4	心情	晴朗,乱,快乐，灿烂,平静
5	牛奶	美丽,嫩,可爱，白皙,漂亮
6	奥运	快，随意，畅通，
7	白纸	坦白，纯洁，清纯，渊博
8	莉香	洒脱，骄傲，美丽，勇敢
9	吃苹果	容易，简单，容易
10	鞭炮	连续，响

英语独现喻体

编号	喻体	属性（包含象征、反讽）
1	kitten	happy,ineffective unworldly
2	statue	hard deadly shapely stiff ...
3	tiger	alert energetic fierce
4	puppy	lovable sweet gentle cozy
5	snowflake	Intricate natural pure
6	root_canal	joyful enjoyable thrilling
7	tornado	capricious deadly violent cute
8	snail	responsive peaceful Lazy mindless
9	dolphin	sleek smart docile gorgeous
10	log	dumb unsettled sleepy heavy

我们设想，这部分隐喻可能是待证的跨语言交际中的 “有障碍”认知成分。

我们还在《知网》语义体系内试图抽取属性间的语义关系，促进隐喻属性规约化的研究及多喻体间的不同关系的抽取。目前基于平行知识库已经形成了一些新的映射关系（如表5，表6所示）。尽管列举的关系并不丰富，但通过《知网》语义体系大范围抽取隐喻喻体及属性间的映射关系是完全可行的。

表5　喻体属性间关系例举

喻体	属性	属性关系
蛋糕	甜美 sweet 甜蜜 luscious	同义关系
蚂蚁	快 fast 慢 slow	反义关系
	微小 tiny 小 small	近义关系
	强 strong 弱 weak	反义关系
海洋	宽广 broad 大 broad	同义关系

表6　多喻体间关系例举

类型	喻体集合
属性有同义关系	水-water，水晶-crystal，苹果-apple
	花-flower，太阳-sun，蛋糕-cake，糕点-cake：sweet 甜美
	妈妈-mother，羽毛-feather，气流-breath
属性有反义关系	狼-wolf，蚂蚁 ant
	针-needle，气流-breath

5　跨语言隐喻属性的互增益研究

跨语言隐喻属性的互增益研究，是本文的另一个焦点所在。这项工作实际包括两方面：跨语言属性比对挖掘和跨语言喻体“嫁接”。所谓跨语言属性比对挖掘，是对知识库中喻体进行针对性的跨语言数据挖掘，观察该喻体不同种语言属性间的比对关系，达到动态补充知识库的效能。我们从知识库中抽取了7个英语喻体进行汉语数据挖掘，统计结果见表7。

表7　英语喻体汉语挖掘效果

编号	喻体	英语属性	挖掘到汉语属性（词型_词频）	汉英语属性比对
1	abacus 算盘	primitive(不开化，原始，原始人)	死板_1，坚硬_2	隐喻属性类似
2	abattoir 屠宰场	bloody(浴血，血淋淋，嗜血，血腥)	性感_2，无奈_1，恶心_1	隐喻属性相异
3	Chef 厨师，大师傅，主厨	fastidious skilled expert (挑剔，熟练，谙练，熟，内行，专家，拿手，娴，里手，通，专才，行家)	创新_1，专业_2，出色_2	隐喻属性共现
4	Tuna 金枪鱼	intriguing (奇妙)	被捕杀_10	隐喻属性相异
5	Torrent 湍流	swift concentrated（迅疾，迅速，雨燕，湍急，专一，密集，集群）	汹涌_3，奔放_1,急_2	隐喻属性共现
6	waterfall 瀑布	natural dynamic magical lovely spectacular（自然，活跃，动态，能动，神奇，神异，可爱，妩媚）	飞泻_1，狂飙_2，义无反顾_1，漂亮_3	隐喻属性共现
7	Lynx 猞猁	fearless（无所畏惧，毫不畏惧，无畏）	浓密_1，神秘_1	隐喻属性相异

所谓跨语言喻体“嫁接”，是针对跨语言交际中隐喻用法不相同或隐喻属性相似度极低的喻体，进行类似“嫁接”喻体的工作，以达到更易理解的表达效果。例如 “像白纸一样清纯”，普通机器翻译结果为：“as pure as the white paper”。此翻译虽遵循原文，但可能带来一定的理解障碍。我们则通过知识库找到属性“清纯”（英语为 pure）映射喻体为 angel，baby，snowflake 等。我们设想，对翻译结果中的喻体（即 white paper）进行替换（如“snowflake”），更利于跨语言隐喻用法的表达。但是如何进行选择，选择效果如何，都需要进一步深究。我们以汉语隐喻用法翻译为英语为例，从知识库中随机生成了汉语比喻句7句，其分析结果可见表8。

表8　汉语比喻句英语翻译及待“嫁接”喻体表

编号	汉语喻体	汉语比喻句	英语翻译及待嫁接喻体
1	小强（蟑螂义）	像 *小强* 一样 顽强	as stubborn as*XiaoQiang*
			[待嫁接喻体，下同]ass badger bear bull ...
2	如何把大象放进冰箱	像 *如何把大象放进冰箱* 一样 简单	As simple as ***how the elephant in the fridge***
			child sunbeam ...
3	火锅	像 *火锅* 一样 热情	as warm as ***the hot pot***
			Chalet comforter hair_dryer ...
4	香烟	像 *香烟* 一样 寂寞	as lonely as ***cigarettes***
			mojave_desertlighthouse_keeper ...
5	足球	像 *足球* 一样 臭	as smelly as ***football***
			Not find
6	神仙	像 *神仙* 一样 快活	as happy as the ***gods***
			angel beggar cheerleader king ...
7	粽子	像 *粽子* 一样 可爱	as cute as ***dumplings***
			bunny candy_canedeer_mouse ...

6　总结及未来工作

通过上述工作，我们已经在《知网》语义框架内，建立了汉英语词汇隐喻属性知识库（包括3个子库）。词汇属性及比对结果多角度的描写了跨语言隐喻属性的风貌：隐喻属性确有跨语言成分，并且这部分隐喻属性可以通过数据挖掘比对结果体现，其语义特征往往类似于《知网》的义原范畴。同时，跨语言互增益研究为我们继续深入探讨提供了持续可能。

我们下一步的工作将继续扩充知识库，仔细研究隐喻属性规约的数据表现，并基于《知网》充分交互挖掘隐喻属性的多语言语义特征，以进行更深入的隐喻计算。

参 考 文 献

[1] Lakoff G & Johnson M. Metaphors We Live by[M].Chicago: The University of Chicago Press, 1980.

[2] B. F. Bowdle & Gentner D. The Career of Metaphor[J]. Psychological Review, vol.112, 2005.

[3] Gentner D & B. F Bowdle. Convention, Form, and Figurative Language Processing[J]. Metaphor and Symbol, vol.16, 2001.

[4] Tony Veale, Yanfen Hao. Learning to Understand Figurative Language: From Similes to Metaphors to Irony[A]. Proceedings of CogSci 2007[C], Nashville, USA, 2007.

[5] Yanfen Hao, Tony Veale. An Ironic Fist in a Velvet Glove: Creative Mis-Representation in the Construction of Ironic Similes[J]. Minds and Machines Vol.20, No. 4, 2010.

[6] Roncero, Kennedy, J. M., Smyth, R. Similes on the internet have explanations[J]. Psychonomic Bulletin and Review, vol.13, 2006.

[7] 董振东，董强.《知网》，http://www.keenage.com.

[8] 陈小荷. 属性分析说略[A]. 孙茂松、陈群秀主编，语言计算与基于内容的文本处理[C]. 清华大学出版社，2005.

[9] 贾玉祥，俞士汶. 基于实例的隐喻理解与生成[J]. 计算机科学，2009，36(3).

[10] 袁毓林，一价名词的认知研究[J]，中国语文，1994(4).