

语料库视野下的现代汉语“狮”字释义研究

THE INTERPRETATION OF THE MODERN CHINESE CHARACTER

“LION” FROM A CORPUS PERSPECTIVE

李卫东¹, 李仕春²

Weidong Li¹, Shichun Li²

^{1,2}泰国华侨崇圣大学中国学学院

^{1,2}College of Chinese Studies, Huachiew Chalermprakiet University, Thailand

Received: February 21, 2023 / Revised: September 12, 2023 / Accepted: October 27, 2023

摘要

伴随世界范围内的汉语学习热潮, 中国以外汉语学习者对汉语学习词典的编纂和功能提出更高的要求。对比英语词典的释义, 汉语词典在义项数量和新技术的运用上都存在差距。以动物名词“狮子”(Lion)为例, 从北京大学现代汉语语料库(CCL语料库)中搜索含有“狮”字的全部语料8787条, 逐条进行分析标注。借用物性结构理论, 通过对“狮”的每一类物性角色的分析, 重新对“狮”字的动物名词性义项释义进行修改, 使得“狮”的释义更加科学。同时还发现汉语词典可以为“狮”字增加2个名词性义项, 即名词“一种民间表演形式或民间表演用道具”和名词“形状像狮子的人造物”。通过物性结构理论, 为每一类名词构建相对统一的释义模式, 语料库分析发现名词释义新义项, 从而丰富现代汉语词典编纂手段, 使释义更加科学和规范, 提升词典的实用性和使用舒适感。

关键词: 名词释义 义项分布 义项描写 语料库技术 物性结构理论 狮

Abstract

Due to the worldwide boom in learning Chinese, foreigners studying Chinese are putting higher demands on the compilation and functions of Chinese learning dictionaries. Compared with English language dictionaries, Chinese dictionaries have gaps in the number of meaning items and the use of new technologies. Taking the animal noun “lion” (狮) as an example, we searched through all 8,787 items containing the word “lion” from the CCL corpus of Peking University, and analyzed and annotated them

one by one. Using the theory of Qualia structure, we analyzed each type of object role of “lion” and revised the meaning of the animal noun definition of “lion” so that the interpretation of the meaning of “lion” was made more scientific. It was also found that two noun meanings for the word “lion” could be added to the Chinese dictionary, namely, the definition “a form of folk performance or a prop for folk performance” and the definition “a man-made object shaped like a lion”. Through the Qualia structure, a relatively unified interpretation model was constructed for each type of noun, and corpus technology discovered new items of nouns interpretation, thus enriching the means of compilation of modern Chinese dictionaries, which will make the interpretation more scientific and standardized, and enhance the practicality and comfort of use of the dictionaries.

Keywords: Nouns Definition, Semantic Distribution, Semantic Description, Corpus Technology, Qualia Structure, Lion

引言

随着中国经济发展，汉语在国际上影响力不断加大，学习汉语者也越来越多。汉语教学的发展，迫切需要与之相配套的外向型学习词典。Xia (2009) 调查广州地区 CFL (Chinese as a Foreign Language) 使用汉语学习词典情况，结果显示：非中国人的汉语学习者有汉语词典使用的需求，但所调查的全部受试者中没有一人使用过任何一种类型的外向型汉语学习词典。

近年来，为了适应教学的需要，中国国内也相继出版了一些双语词典。但是我们对 10 本权威性中英文中型语文词典中 100 核心词义项数目进行比较，发现在 100 核心词的义项数目方面，《现汉》(第 6 版) 义项数目明显少于《牛津》(第 7 版)、《韦氏》，通过对同类词释义的比较发现汉语辞典漏收了不少义项。

词典释义是词典编纂中最核心的问题之一，词典释义研究也是词典学中热门研究领域之一，对外向型汉语学习词典释义原则进行专门研究的论文数量也比较多。Cui (2011) 通过分析对外汉语新词语词典释义原则和方法，提出可以通过圈重点、变句式和改词语三种方法来降低释义的难度；Zhang (2011) 发现对外汉语学习词典释义比语文词典释义，在内容上更丰富，语言更简短、通俗，更符合二语学习特点；Xie (2012) 提出科学性与实用性相统一、整体性、客观性和思想性兼顾三大释义原则；Xia (2013) 强调词典释义的重要性，指出要突出外向型汉语词典释义的可懂性；Weng (2014) 对比《现代汉语词典》和《商务馆学汉语词典》提出准确化、明了性、模式化三大释义元语言原则；He (2020) 通过对《商务馆学汉语词典》44 个职业义场常用词研究，提出在原型范畴下的释义模式可以为词典释义规范化、精细化、系统化提供新方向，有效解决性通行原则未贯穿词典编纂始终的问题。

学界对词典释义原则分析不仅只基于理论层面的描写,更注重与实际描写结合。大多数学者都注重词典释义系统性原则。为此,我们借用语料库技术和生成词库理论中的物性结构理论来重新分析和定义动物名词的释义。

用语料库编纂词典,比传统的人工卡片方式具有明显优势。占据现今英语辞书主要市场的 Collins, Longman, Cambridge, Oxford 等出版商开发的英语词典,都是在语料库基础上编纂而成。目前中国汉语类词典还没有大规模地运用语料库技术对汉语常用词义项进行丰富与补充,汉语类词典中常用词义项的编纂水准至今仍处在 20 世纪五六十年代靠语感编纂的水准。

美国教授 Pustejovsky (1991) 提出生成词库理论 (Generative Lexicon Theory, GLT), 创造性的将生成技术引入词义研究。其主要内容包括词项的词汇表征和句法层面的语义生成机制两部分,其中最有特色和代表性的就是引入物性结构。基于物性结构分析,借用 Aristotle “四因说”,提出了名词三大语义类型,即按名词所代表的意义内容分为自然类、人造类和合成类三种。词项物性结构主要包括构成角色 (Formal Role, CON)、形式角色 (Constitutive Role, FOR)、功用角色 (Telic Role, TEL)、施成角色 (Agentive Role, AGE) 四个层面的语义知识。

Yuan (2014) 借鉴 Pustejovsky 物性角色理论,结合汉语名词的特点和自己研究目标、应用需要,将论元结构、事件结构、词汇继承结构相关语义知识分别融入到有关物性角色的描述中,增加了单位 (Unite, UNI)、评价 (Evaluation, EVA)、材料 (Material, MAT)、行为 (Action, ACT)、处置 (Handle, HAN)、定位 (Orientation, ORI) 角色,将物性角色总数增加至 10 个,并提出了建构和标注物性角色的 5 条一般原则,形成了包括“物性角色”和“句法格式”的汉语名词物性结构描述体系。一般来说,不同名词会有不同的物性角色数量和种类,但是形式角色、构成角色是所有名词都具备的。

利用语料库技术,借用物性角色理论,可以更好的分析和描述“狮(子)”的释义。

内容

“狮”字在汉语词典、英语词典中的义项

1. 汉语词典中“狮”的义项

“狮”字在《国际中文教育中文水平等级标准》中属于七九级汉字;在《现代汉语频率词典》“使用度最高的前 8000 个词词表”中的排列顺序为 5507, 使用度为 11, 词次为 21, 频率为 0.0016 (Research Institute of Language Teaching, 1986), 是较为常用的汉字。

在《现代汉语词典 (第 7 版)》(以下简称《现汉》)、《现代汉语学习词典》(以下简称《现学》)、《商务馆学汉语词典 (双色本)》(以下简称《商学》)三部汉语学习词典中,“狮”字义项都只有一个,即“狮子”的动物名词性义项。

《现汉》释义“狮”时，物性角色排列顺序为“形式+构成+行为+评价”。其中，形式角色排列顺序为“分类属性+大小+四肢+尾巴+毛皮”，抓住了狮子最典型的特征体型大、四肢强壮，毛色棕黄色、雄狮颈部有长鬣，行为上扑食其他动物、吼声很大，人类对其有“兽王”的评价。选用的多字条目除了“狮子”以外，还有“狮子搏兔”“狮子大开口”“狮子狗”“狮子舞”“狮子座”。

《现学》释义与《现汉》基本一致，只是物性角色少了“评价”，“行为”角色中减少了“吼声”。选用的多字条目有“狮子”“狮子大开口”“狮子狗”“狮子头”“狮子舞”。

《商学》释义相对更简练，采用物性角色排列顺序为“形式+构成”，其中“形式角色”主要突出狮子皮毛为黄褐色，雄性的脖子上有长毛等最重要的特征。多字条目只有“狮子”一个。

2. 英语词典中 Lion 的义项

选取“狮子”对应的英语单词 Lion，分别查找其在《牛津高阶英语词典第 10 版》(Oxford Advanced Learner's Dictionary 10th Edition, 以下简称《牛高》)、《朗文当代高级英语辞典第 5 版》(Longman Dictionary of Contemporary English, 以下简称《朗高》)、《麦克米伦高阶英语词典》(Macmillan English Dictionary for Advanced Learners, 以下简称《麦高》)、《柯林斯词典（修订版）》(Collins Cobuild Dictionary Overhaul, 以下简称《柯氏》)、《韦氏高阶英汉双解词典（2019 版）》(Merriam-Webster's Advanced Learner's English-Chinese Dictionary, 以下简称《韦高》) 中的释义。

《牛高》中 lion 只有 1 个动物名词性义项，物性角色涉及“形式角色”(Large, Animal of the Cat Family, Yellow-brown Fur, Mane)、“构成角色”(Africa and Southern Asia, Male)、“行为角色”(Hunts in Groups)。

《朗高》收录了 2 个义项，一个为动物名词性义项，一个为专有名词。在动物名词后面还增加了 2 个常用短语释义。动物名词性义项中物性角色涉及“形式角色”(Large, Animal of the Cat Family, Gold-coloured Fur, Mane)、“构成角色”(Fur, Male)。

《麦高》收录了 2 个义项。一个动物名词性义项，物性角色涉及“形式角色”(Large, Wild Animal, Golden Fur, Thick Hair, Mane)、“构成角色”(Africa, Male)。一个为隐喻意义，“狮子”有 Golden Fur, Thick Hair, Large, 这些都彰显着力与美，隐喻到人类社会中，就用 Lion 来指称 Someone Who is Powerful, Impressive, or Brave。

《柯氏》只收录了 1 个动物名词性义项，物性角色涉及“形式角色”(Large, Wild Member of the Cat Family, Yellow Fur, Mane, Long Hair)、“构成角色”(Africa, Male, Head, Neck)。

《韦高》收录了 3 个义项。一个动物名词性义项，物性角色涉及“形式角色”(Large, Panthera Leo, Tawny, Tufted, Shaggy, Blackish or Dark Brown, Mane)、“构成角色”(Africa and Southern Asia, Body, Tail, Mane, Male)、“行为角色”(Heavily Built Social Cat)。此义项下，还解释了 LEO。一个为隐喻意义，指称那些和狮子一样具备 courage or ferocity 的人。第三个义项为专有名词，指狮子俱乐部成员。

对比汉语学习词典和英语学习词典中“狮 (Lion)”的释义,从收录义项数目方面,《现汉》、《现学》、《商学》等汉语学习词典都收录一个义项,即对动物名词本义的解释。英语学习词典中,在动物名词释义外,还增加了隐喻义,《牛高》收录 1 个义项,《朗高》2 个,《麦高》2 个,《柯氏》1 个,《韦高》3 个。英语学习词典收录义项数要多于汉语学习词典,这与我们之前的研究结论是一致的。

语料抽取情况

从北京大学 CCL 现代汉语语料库中搜索并下载“狮”字全部语料共计 8787 条,逐条进行标注分析。发现“狮”字用在人名、地名、品牌名称、公司名称或社会团体组织名称等专用名字中共计 4032 条,这一类语料中对于我们研究“狮”字义项划分和描写意义不大,可以忽略。后文中,对“狮”字的义项分布统计,均以剩余的 4755 条语料为计算范本。所选语料在 CCL 语料库中具体分布情况如下:

表 1 “狮”语料分布情况

二级分类	三级分类	四级分类	语料数量	有效语料数量
	CWAC		14	10
	口语		10	8
	史传		34	12
		中国政府白皮书	1	0
		健康养生	3	3
		法律文献	0	0
		社会科学	65	51
	应用文	自然科学	99	76
		药方	0	0
		菜谱	28	25
		议论文	2	0
		词典	39	39
		说明书	0	0
		1994 年报刊精选	714	171
		人民日报	1550	753
当代		作家文摘	107	70
	报刊	市场报	138	51
		故事会	3	3
		新华社	3022	1322
		读书	180	145
		读者	340	302
		青年文摘	2	2
	文学	台湾作家	321	117
		大陆作家	334	250
		香港作家	224	109
	电视电影	文艺	19	12
		非文艺	120	108
	相声小品		31	28
	网络语料		308	115
	翻译作品	应用文	253	219
		文学	714	655
现代	戏剧		2	2
	文学		110	97
总计		8787		4755

来源: 根据北京大学 CCL 现代汉语语料库中“狮”字语料统计得出。

汉语学习词典中“狮（子）”释义情况分析

1. 汉语学习词典中已有“狮（子）”义项情况

《现汉》、《现学》、《商学》3部汉语学习词典中，“狮”字都只有一个动物名词性义项，该类语料共计2890条，占全部下载语料60.77%。

在Yuan (2014)汉语名词物性结构的描写框架的基础上，Feng和Lu (2011)提出汉语动物词释义基本模式：“类义征（哺乳动物）+个性义征[状貌（大小+外形+毛皮+头+面+耳朵/眼睛/鼻子/牙齿/唇部/角+尾巴+四肢……）、性情、行为、感知觉能力、食物、生活栖息地、与人类的关系、种类]+又叫/俗称/通称。”结合二者的理论观点，我们对动物名词“狮（子）”描写如下：

狮 shī 名词。哺乳动物，身体长约3米，形体较大，全身毛棕黄色，雄狮的颈部有长鬣，尾巴细长，末端有一丛毛，四肢强壮，有钩爪，掌部有肉块。吼声很大，捕食羚羊、斑马等动物。生活在非洲和亚洲西部。有“兽王”之称。

物性角色：

形式 FOR: 动物、哺乳动物；

构成 CON: 是食肉目、猫科、豹属的大型猛兽；从性别分类，可以分雄、雌或公、母；从年龄分类，可以分为老、成年、幼；从生活区域分类，可以分为亚洲、非洲等等；

单位 UNI: 个体量词有“头、只、个”等，集合量词有“群”等，不定量词有“些”等；

行为 ACT: 群居、捕食、吼、跑等；

评价 EVA: 兽王、勇猛、漂亮等。

句法格式：

S1: CON+

如：雄～|雌～|公～|母～|幼～|亚洲～

S2: Num+UNI+

如：一头～（子）|一只～（子）|一群～（子）|一些～（子）

S3: EVA+的+

如：勇猛的～（子）|漂亮的～（子）

S4: +ACT

如：～（子）群居|～（子）伏击|～（子）捕杀|～吼|～跑

2. 通过语料库新发现“狮子”义项情况

1) 名词，一种民间表演形式或民间表演用道具。

Pustejovsky (1991)根据代表的意义内容将名词分为自然类 (Natural Type)、人造类 (Artifactual Type) 和合成类 (Complex Type)。自然类事物具有较多的自然属性，不涉及人的参与和创造，是与物性结构中形式角色或者构成角色相关的概念，是其他类的基础；人造类事物是由

人创造出来的，体现人的意志，具有较为明显的功能和用途，结合了施成角色或功用角色；合成类事物，在生成词库理论中又被称为 Dot Object（点对象），从 2 至 3 个自然类和人造类继承角色，同时具有自然属性和人造属性。常见合成类型有：

- phys_obj•info（物质和信息）：Book, Record
- event•info（事件和信息）：Lecture, Play, Seminar, Exam, Quiz, Test
- event•phy_obj（事件和物质）：Lunch, Breakfast, Dinner, Tea
- event•(info•sound)（事件、信息和声音）：Concert, Sonata, Song, Symphony

“狮（子）”的动物名词性义项明显是自然类，基本不涉及人类的参与和创造。但是在全部语料中还发现有 1865 条“狮”字语料，有明显的人类参与创造。如：其快乐程度不下于看狮子、龙灯、高跷、抬阁、和僧道齐全、六十四杠的大出丧。

“看”的对象“狮子”，可以是物质对象，即人造的、模仿动物狮子形状的表演道具；也可以是事件对象，即舞狮子表演的过程，应归纳为合成类 event•phy_obj（事件和物质）。

类似含有“舞狮”、“狮子舞”、“耍狮子”等词语的语料在有效语料中共计 1098 条，占 23.09%。

“狮（子）”的动物名词义和物质义、事件义的差别在物性角色上体现得更明显。

表 2 “狮”字动物名词义和物质义、事件义对比

物性角色	动物名词义	物质义	事件义
形式	事物、动物、哺乳动物、兽	事物、人工物、器具；形状像动物狮子的头部	过程、事件；表演
构成	雄狮、母狮；大狮子、小狮子；亚洲狮、非洲狮	文狮、武狮；南狮、北狮	音乐、舞蹈、动作
功用	——	表演用的道具	庆祝、祈福
施成	胎生	制作	编排
单位	个体量词：头、只 集合量词：群 不定量词：些	个体量词：条、个 集合量词：对、队	场、次、回
评价	兽王、勇猛、漂亮	精美、活灵活现、鲜艳	精彩、好看
材料	——	木头、竹子、布料、彩纸、 颜料	——
行为	群居、捕食、吼、跑	——	走、跳、翻滚、点头、眨眼
处置	——	制作、举、抬、舞、耍	看、欣赏
定位	——	——	——

来源：借用袁毓林设计的 10 种物性角色名称，作者自制表格。

2) 名词, 形状像狮子的人造物。

Song (2017) 指出“隐喻是人类用来组织概念系统不可缺少的认知工具, 是通过一种事物来理解另一种事物的手段。隐喻的运作方式是不同概念域之间的映射, 隐喻产生的基本条件是相似性, 物性结构可以揭示相似涉及的特征范围。”“形式角色因其空间性, 最易被感知, 因而最易作为理解其他陌生或抽象概念的基点。”下列句子中的“狮子”是模仿动物名词“狮子”的形式角色, 用石头或者金属、木头、泥土等物质雕刻塑造出来, 外形上具有极大的相似性, 带有人类强烈的装饰、审美或者精神寄托。与自然类的“狮子”不一样, 是典型的人造类。如: 到这时我才想起, 来到卢沟桥上, 岂能不还摸狮子的夙愿。那桥栏柱头的石狮子据说有五百多个。

类似含有“石狮子”、“石狮(非地名)”、“狮身人面像”等词语的语料共计 609 条, 占 12.81%。“狮子”的动物名词义和人造义的差别同样也在物性角色中有明显的体现。

表3 “狮”字动物名词义和人造义对比

物性角色	动物名词义	人造义
形式	事物、动物、哺乳动物兽	事物、人工物
构成	雄狮、母狮; 大狮子、小狮子; 亚洲狮、非洲狮	雕像、塑像; 石头的、金属的、木制的
功用	—	装饰、审美、精神寄托
施成	胎生	制作
单位	个体量词: 头、只 集合量词: 群 不定量词: 些	个体量词: 尊、座个 集合量词: 批 不定量词: 些
评价	兽王、勇猛、漂亮	精美、活灵活现、雄伟、巨大
材料	—	泥土、木头、石头、金属
行为	群居、伏击、捕杀、吼、跑	—
处置	—	制作、摆放、看
定位	—	—

来源: 借用袁毓林设计的 10 种物性角色名称, 作者自制表格。

3. 多字条目调整

综合《现汉》《现学》《商学》3部汉语词典中“狮”的多字条目, 分别在CCL现代汉语语料库中进行统计。其中含有动物名词“狮子”语料共计 3123 条; 含有“狮子搏兔”语料共计 8 条; 含有“狮子大开口”语料 39 条; 含有“狮子狗”语料 54 条; 含有“狮子舞”语料 67 条; 含有“狮子座”语料 98 条; 含有“狮子头”语料 149 条。其中, “狮子搏兔”、“狮子大开口”虽然语料数量不多, 但是这里“狮子”还是动物名词, 所以建议顺序还是放在多字条目“狮子”之后。

据此, 建议将“狮”字释义多字条目顺序调整为“狮子”“狮子头”“狮子座”“狮子舞”“狮子狗”“狮子大开口”“狮子搏兔”。

结论与建议

基于物性结构理论的分析,借鉴 Feng 和 Lu (2011) 提出的哺乳类动物词释义的基本模式,我们将“狮(子)”的动物名词性义项进行了重新排列;通过语料库技术分析和物性角色的对比,我们还发现了可以为“狮”字补充两个新义项:一种民间表演形式或民间表演用道具;形状像狮子的人造物。我们为名词“狮”增加了两个新义项。据此,以《现汉》释义为模板,建议将汉语词典中“狮(子)”的义项描写如下:

狮 shī

1) 名词。狮子。哺乳动物,身体长约3米,形体较大,全身毛棕黄色,雄狮的颈部有长鬣,尾巴细长,末端有一丛毛,四肢强壮,有钩爪,掌部有肉块。吼声很大,捕食羚羊、斑马等动物。生活在非洲和亚洲西部。有“兽王”之称。

2) 名词。一种民间表演形式或民间表演用道具。

3) 名词。形状像狮子的人造物。

释义多字条目为“狮子”“狮子头”“狮子座”“狮子舞”“狮子狗”“狮子大开口”“狮子搏兔”。

新释义模式也可以用在“狮”之外的其他兽类动物名词释义中,实际使用过程中可以有一定的灵活性,根据实际情况在释义表达方式和顺序上略作调整。同理,也可以参照该模式,并根据其他动物类名词(如禽类、虫类、水生动物类等)的特点,修订出该动物类名词的释义模式,进而实现名词释义的“同场同模式”原则。

运用语料库技术和物性结构理论,为每一类型动物名词的释义建立了统一的模式,可以促进名词释义的科学性、准确性和规范性。通过语料库技术分析和物性角色的对比还可以发现部分名词新义项。此项系统研究促使现代汉语词典编纂的手段更丰富,使得现代汉语词典释义更加科学和规范,从而提升词典的实用性和用户的使用舒适感。

References

- Cui, L. (2011). On the interpretation principles and methods of a dictionary of new words in Chinese as a foreign language. *Overseas Chinese Education*, 4(61), 18-25. [in Chinese]
- Feng, H. X., & Lu, D. M. (2011). An examination of the systematic interpretation of animal lexical items in language dictionaries and the optimization of interpretation patterns. *Lexicographical Studies*, (6), 15-26. [in Chinese]
- He, Y. (2020). A study on the interpretation of professional semantic field words in the commercial press learner's dictionary of contemporary Chinese. *Journal of Luohe Vocational Technology College*, 7(4), 4-7. [in Chinese]
- Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics*, 17(4), 409-441.

- Research Institute of Language Teaching. (1986). *Modern Chinese frequency dictionary*. Beijing Language and Culture University Press. [in Chinese]
- Song, Z. Y. (2017). The cognitive orientation of generative lexicon: from the perspective of linguistic relevance. *Journal of Peking University (Philosophy and Social Sciences)*, 54(2), 151-160. [in Chinese]
- Weng, X. L. (2014). The rhetorical guidelines of the meta-language of the Chinese learning dictionary: A discussion on the paraphrasing meta-language of the Chinese dictionary of business pavilion learning. *Contemporary Rhetoric*, 5(185), 84-90. [in Chinese]
- Xia, L. X. (2009). Research on the publication and users of CFL learner's dictionaries. *Publishing Journal*, 17(1), 65-69. [in Chinese]
- Xia, L. X. (2013). Definition-related problems in commercial press learner's dictionary of contemporary Chinese and possible solutions. *Journal of Yunan Normal University (Teaching and Research on Chinese as a Foreign Language Edition)*, 1(11), 44-48. [in Chinese]
- Xie, H. J. (2012). Several issues on the interpretation of Chinese cultural dictionaries. *Lexicographical Studies*, (3), 15-23. [in Chinese]
- Yuan, Y. L. (2014). On a descriptive system of qualia structure of Chinese nouns and its application in parsing complex Chinese grammatical phenomena. *Contemporary Linguistics*, 16(1), 31-48. [in Chinese]
- Zhang, Y. H. (2011). An investigation on the problems in the definitions in Chinese dictionaries for international learners: Based on a contrastive study of EFL and CFL lexicography. *Chinese Teaching in the World*, 1(25), 120-128. [in Chinese]



Name and Surname: Weidong Li

Highest Education: Doctoral Candidate

Affiliation: College of Chinese Studies, Huachiew Chalernprakiet University

Field of Expertise: Chinese Teaching



Name and Surname: Shichun Li

Highest Education: Doctoral Degree

Affiliation: College of Chinese Studies, Huachiew Chalernprakiet University

Field of Expertise: Lexicography, Lexicography, and History of Linguistics