

新编同义词词林语义分类体系

马永腾¹, 亢世勇²

(鲁东大学, 山东烟台 264025)

摘要: 适应信息技术的发展, 我们拟建立一个新的语义分类体系, 为开发编纂《现代汉语同义词反义词词林》建立语义分类体系基础。希望建成一个能够满足人机两用, 严格区分上下位关系和同义反义关系的层级语义体系。

关键字: 语义分类, 层级体系, 人机两用

A Newly Organized and Categorized Semantic system for Synonym Dictionary

MA Yongteng¹, KANG Shiyong²

(Ludong University, Shandong Yantai 264025)

Abstract: Meeting the development of the information technology, we plan to set up a new categorized system as the categorized system foundation of semantics to compile *Modern Chinese Synonyms Antonym Word Forest*. It was hoped to be satisfy man-machine dual-purpose, distinguish the relations and level of semantic systems against the justice relation of synonym of upper and lower location strictly.

Keywords: Semantic classification, level system, man-machine dual-purpose

1 建立汉语同义词语义分类体系的目的

随着现代信息技术, 特别是中文信息处理的发展, 原有的《同义词词林》《反义词词林》已经明显的不能适应社会的需要。

主要表现在: (1) 语义分类体系不够科学, 一些词语的归类也需要调整。我们随机抽取了“动物”中的“昆虫”小类, 发现归类失当的词为数不少。如 973 语义分类体系, 1.1.1.2.4 昆虫: “蚯蚓”; 《同义词词林》(梅家驹, 1996) B118 昆虫类收“蚯蚓、曲蟾”, “蚯蚓、曲蟾”应归为无脊椎动物类。另外, 有些收录的并不成词, 如“冻死骨” Bk01 等, 实际上混淆了词与非词的界限。(2) 语义分类体系粗疏, 人机两用的兼容度低。我们建立汉语同义词语义分类体系, 为开发编纂的《现代汉语同义词反义词词林》建立语义分类体系基础。

我们希望做到: (1) 既能满足人的需要, 也能满足机器的需要; (2) 建立科学的语义分类体系, 这个体系同样是一个层级体系, 能够体现词语的上下位关系和“同义”关系。(3) 严格区别词语的同义、反义关系。

2 建立语义体系的基本方法

(1) 严格描述词语的上下位关系及同义关系, 建立其联系和区别, 可以为计算机信息检索、机器翻译等广泛使用, 也可以为人学习、研究使用。

作者简介: 马永腾(1981-), 男, 山东高密, 05 级在读研究生, E-mail: flyinma@163.com

(2) 吸收和借鉴近年来国内外词网研究和开发的新成果，特别是语言信息处理领域的新成果，建立科学的语义分类体系。现有各家的语义分类体系因其面向实际的考虑不同而各有所长，也各有不尽完善的地方。905 立足于语义场却有些忽视了语义关系，知网抛开语法研究语义，973 对事物的分类太过简单。

(3) 我们以 ONTOL-MT 为蓝本，基于标注语料进行适当调整。吸收借鉴美国的 WORDNET、FRAMENET，国内的董振东的知网、鲁川的运动类语义体系、北大的 CCD、清华现代汉语述语动词词典、山西大学的 CFN 以及最近的“现代汉语语义词典”语义结构等各家语义分类体系和新成果，进一步细化，分为三个层级(大类—中类—小类)。

3 语料标注和语料库的建立

我们以 05 版《现代汉语》为标注语料，增加《新词语大辞典》中的新词新语，建立规模约为 13 万词的《新编同义词词林语义标注语料库》，以此为基础建立语义分类体系。

基于语料库，在标注时我们的具体做法有以下几个方面。

1. 有些类的内涵不明确，我们按照待标词语的实际情况，按照从典型到边缘化处理。如：Gp[机构]，政府，监狱，这是典型的机构。博物馆，歌厅，游泳池等相似度较大和边缘词语也划入这一类。

2. 兼属划类问题。有些词兼属几个词类或有几个义项，归为多个类。有些词义项单一，但归类不一，如“水库” Ch//Go，“矿物” Bb//Bg。在标注的时候，我们觉得这样的词语不应划为重复类。而“竹笋”等词可以同时归入 Bd[植物]和 Cd[食物]。“吃豆腐”“吃醋”“吃皇粮”等词，只标其比喻义。“马蜂窝”等词实义和比喻义都标。

3. 还有一些具体词语的调整：“保姆” Ai09[其他职业]，我们觉得跟“理发师”“导游”等词是一类的，所以归入 Ai08[服务业人员]。有的词语如“乘务员，职员”分在 Ah[成员]（党员），我们调整为“Ai[职业]-Ai08[服务业人员]（导游），Ai[职业]-Ai03[商业人员]（推销员）”。[团体]从 H.[数量]-Hs[群组 对]中调至 A 类，Ah02[团体]，表示“人”，如“少先队，歌舞团”等。

这样，我们建立了如下的语义分类体系，这是一个知识本体分类系统。（我们给出了标记和例词，例词限举。）

4 语义分类体系（我们只抽取 A 层的前 9 类和 B 层 f 类为例）

A. [人](human)

Aa[泛称]：指称各类人。

Aa01[复称]：例如，人类，人民，大伙儿，公民，群众，众生，黎庶，黎民

Aa02[双称]：例如，师生，父女，叔侄，舅甥，宾主，僧俗，党群，军民

Aa03[个称]：例如，匹夫，个人，小子，竖子

Ab[人称]：三种人称。

Ab01[第一人称]：例如，我，我们，俺，吾，予，吾侪，我辈

Ab02[第二人称]：例如，你，你们，尔，汝

Ab03[第三人称]：例如，他，她，他们，她们

Ac[自他]：指称自己或别人。

Ac01[自己]：例如，自己，本人，在下，鄙人，小可，晚生，不才，老朽，老身

Ac02[泛他]：例如，别人，旁人，他人，别个，旁个，人家

Ac03[面称]：例如，诸位，各位，大家，兄台，仁兄，阁下，同志，大姐

Ac04[某人]：例如，某男，某生，有人，某君

Ad[男女]：男子或女子。

Ad01[男子]：例如，男子，男人，男生，男子汉，男儿，汉子，爷儿，须眉

Ad02[女子]：例如，妇女，姑娘，处女，少女，妇女，丫头，千金，巾帼，闺女

Ad03[男女]: 例如, 儿女, 红男绿女, 男男女女, 倩男靓女, 叟姬
Ae[老少]: 不同年龄段的人。
Ae01[老人]: 例如, 老汉, 老太婆, 老翁, 长者, 耄耋, 老头, 媪, 老太太
Ae02[成人]: 例如, 壮夫, 大人, 丁壮, 丁男, 成年人, 大人
Ae03[老少]: 例如, 童叟, 老幼, 稚老, 老少, 老小, 黄发垂髫
Ae04[青少年]: 例如, 青年, 小伙儿, 初生之犊, 少年, 后生
Ae05[儿童]: 例如, 孩童, 孩子, 宁馨儿, 稚童, 小人儿, 伢子, 童蒙
Ae06[婴儿]: 例如, 婴孩, 宝宝, 襁褓小儿, 乳儿, 小毛孩, 幼婴

Af[体态]: 高矮胖瘦等形体不同的人。

Af01[高个子]: 例如, 高个子, 大高个儿, 大个子, 彪形大汉
Af02[矮子]: 例如, 小个子, 矮子, 小矮子, 矮个儿, 侏儒
Af03[胖子]: 例如, 胖子, 肥仔, 大块头, 胖敦子, 大胖子
Af04[瘦子]: 例如, 瘦子, 瘦长条, 骨头架
Af05[美人]: 例如, 美眉, 美女, 靓女, 索女, 帅哥, 靓仔, 型男
Af06[丑人]: 例如, 丑女, 恐龙, 丑男, 青蛙, 母夜叉, 丑八怪

Ag[籍属]: 国籍或籍贯不同的人。

Ag01[国籍]: 例如, 中国人, 韩国人, 美国佬, 英国人, 荷兰人, 西班牙人
Ag02[省市]: 例如, 云南人, 烟台人, 北京人, 香港客, 山东大汉
Ag03[其他]: 例如, 赵庄人, 辛县人

Ah[成员]: 某团体的成员。

Ah01[党派成员]: 例如, 党员, 团员, 会员, 议员, 委员
Ah02[团体]: 例如, 少先队, 共青团, 球队, 互助组, 互助会, 校友会
Ah03[运动员]: 例如, 运动员, 裁判员, 健将, 前锋, 选手, 二传手, 守门员
Ah04[身份]: 例如, 要员, 大员, 专员, 首领, 核心, 灵魂, 领军

Ai[职业]: 个人在社会中所从事的作为主要生活来源的工作。(冯志伟, 2005)

Ai01[专业技术人员]: 例如, 技师, 教师, 讲师, 工程师, 医生, 护士, 记者, 诗人, 语言学家, 主持人, 播音员, 歌手, 小提琴手, 艺术家, 舞蹈家, 工人, 建筑师。

Ai02[管理人员]: 例如, 公务员, 部长, 司长, 处长, 科长, 官僚, 高官, 雇员, 法官, 律师, 公证人, 参事, 秘书, 助手。

Ai03[商业人员]: 例如, 企业家, 老板, 商人, 奸商, 店员, 推销员。

Ai04[工农林牧副渔业人员]: 例如, 工人, 农民, 伐木工人, 渔民, 牧羊人, 猎人。

Ai05[运输业人员]: 例如, 司机, 船员, 水手, 车夫, 驾驶员, 飞行员, 宇航员。

Ai06[公安人员]: 例如, 警察, 保安人员, 门卫, 消防队员。

Ai07[军人]: 例如, 将军, 大校, 大尉, 中尉, 退役军人, 复员军人, 士兵, 水兵。

Ai08[服务业人员]: 例如, 理发师, 厨师, 导游, 按摩师, 鞋匠, 清洁工, 勤杂工。

Ai09[其他职业]: 例如, 学生, 研究生, 保姆。

具体物](concrete-things) 是有形、有色、有质量的物, 包括自然物、人造物、工具等。

B [自然物](natural-things)

.....

Bf[躯体]: 人和动物的身体。

Bf01[身体]: 例如, 身体, 肢体, 形体, 全身, 浑身, 胴体, 身子, 形骸

Bf02[尸体]: 例如, 尸首, 死人, 尸身, 尸体, 遗体, 亡人, 木乃伊

Bf03[头部]: 例如, 头, 脑壳, 脑袋, 脑瓜, 脑袋瓜儿, 脑壳, 首级

Bf04[面部]: 例如, 脸面, 面庞, 腮, 鸭蛋脸, 眼, 耳, 鼻, 口, 牙
Bf05[胸颈]: 例如, 颈, 脖梗儿, 肩, 胸, 乳房, 肩胛, 肩头, 胸脯, 胸腔
Bf06[背腰]: 例如, 背脊, 脊梁, 腰板儿, 柳腰, 蜂腰, 楚腰, 虎背熊腰
Bf07[腹臀]: 例如, 腹部, 肚子, 小肚子, 肚脐, 臀部, 屁股, 腓
Bf08[四肢]: 例如, 手, 脚, 腿, 臂, 胳膊, 肘弯, 臂膀, 手掌, 脚趾
Bf09[角翼类]: 例如, 触角, 犄角, 尾巴, 鳍, 翅膀, 鞘翅, 羽翼
Bf10[皮肤 肌肉]: 例如, 肌肤, 眼皮, 平滑肌, 骨骼肌, 胸肌, 腹肌
Bf11[毛发]: 例如, 头发, 眉毛, 睫毛, 胡须, 汗毛, 马鬃, 翎, 羽毛
Bf12[鳞甲类]: 例如, 指甲, 鳞片, 甲壳, 贝壳, 甲壳, 龟甲, 鱼鳞
Bf13[内脏]: 例如, 心, 肺, 胃, 肠, 肝, 脾, 心房, 肺叶, 心室
Bf14[腺类]: 例如, 淋巴腺, 肾上腺, 甲状腺, 腺体, 唾液腺, 乳腺
Bf15[膜类]: 例如, 胸膜, 膈膜, 瓣膜, 耳膜, 视网膜, 黏膜
Bf16[生殖器官]: 例如, 生殖器, 私处, 下体, 胎, 胚胎, 胞衣
Bf17[神经类]: 例如, 经络, 骨髓, 神经元, 运动神经, 感觉神经
Bf18[血液类]: 例如, 血液, 血水, 血小板, 红细胞, 淋巴, 淋巴结
Bf19[排泄物]: 例如, 尿, 屎, 粪便, 秽物
Bf20[分泌物]: 例如, 汗, 奶水, 头皮屑, 唾沫, 乳汁, 乳, 涎水

目前, 这个语义体系正在拟建中, 随着辞典的编纂工作的进行将不断修改完善, 限于时间和水平错漏之处在所难免, 欢迎各位有识之士批评指正!

参考文献:

- [1]冯志伟.基于 ONTOL-MT 的知识本体分类系统(修改稿)[J].2005,12,05
- [2]梅家驹等.同义词词林[M].上海辞书出版社.1996
- [3]王惠、詹卫东、俞士汶“现代汉语语义词典”的结构及应用[J].语言文字应用[M].2006,2
- [4]梅立军、周强、臧路,等.知网与同义词词林的信息融合研究[J].中文信息学报[M].第十九卷, 第一期