·临床研究·

汉语动词语义特征建模与分析*

高 婧[1] 林 枫[1] 江钟立[1,2] 卢红建[1]

摘要

目的:通过构建汉语动词语义特征常模,为临床言语治疗提供量化和可视化的语义特征数据库。

方法:选择30个日常生活常用动词作为刺激词,采集健康人的语义特征,并进行条目编码,然后根据汉语语义特征分型方案对其进行分类。统计软件采用R软件进行数据可视化和统计检验。

结果:①动词的语义特征以功能用途类显著。②动词首位秩次的语义特征以功能用途类显著。③一论元结构动词以内省特征显著。

结论:根据汉语语义特征数据建立的模型可以有效反映概念语义结构,有助于根据量化指标提取语义训练素材。

关键词 动词概念;论元结构;语义特征训练;言语语言治疗;康复

中图分类号:R493,R767 **文献标识码:**A **文章编号:**1001-1242(2016)-04-0381-07

The modeling and analysis of semantic features for Chinese verbs/GAO Jing, LIN Feng, JIANG Zhongli, et al.//Chinese Journal of Rehabilitation Medicine, 2016, 31(4): 381—387

Abstract

Objective: To explore the model of semantic features for Chinese verbs to provide quantitative and visual semantic database for clinical speech therapy.

Method: The semantic feature entries were collected from a total of 30 stimulating verbs in volunteers by feature nomination. These entries were classified into different feature types according to Semantic Feature Classification Scheme for Chinese. With the R statistical computing environment, the distribution of feature types was visualized and analyzed based on properties.

Result: ①Verbs nonliving domain had significantly more usage property. ②The frequency of first rank features were significantly higher than expectation in the function. ③Verbs with one argument structure had significantly more introspection property.

Conclusion: The model of semantic features for the Chinese verbs by feature nomination can effectively reflect the semantic structures of concepts and help to chose semantic training materials according to quantitative indicators.

Author's address Rehabilitation Department, The First Affiliated Hospital of Nanjing Medical University, Nanjing, 210029

Key word feature norms; argument structure; semantic feature analysis training; speech therapy; rehabilitation

人类通过不同的语句进行信息传递及交流,而语句是通过不同性质的词语(如名词、动词、形容词等)在句中承担不同的语法成分而合理的组合而成。其中动词在句中的作用尤为关键,因其不仅携带重要信息,而且在句子结构形成中有重要作用[1]。也就是说,动词通常携带论元,不仅表现出语义特征

DOI:10.3969/j.issn.1001-1242.2016.04.002

*基金项目:江苏省科技支撑计划(BE2012675);国家自然科学基金资助项目(81171854)

1 南京医科大学附属第一医院康复医学科,南京,210029; 2 通讯作者

作者简介:高婧,女,硕士研究生; 收稿日期:2016-01-12

也表现出语法特点。论元是指句子中的名词短语[2]，动词根据其携带的论元分为一论元动词（笑）、二论元动词（看）和三论元动词（给）。对于失语症患者，有研究表明越复杂的论元结构越难以产出动词[3]，这提示对患者进行言语治疗时应从简单论元开始。

Wambaugh等[1]首次将语义特征分析技术（semantic feature analysis,SFA）应用于动作命名，采用和名词语义特征分析一致的6个方面的特征进行训练，并对不适于动词的语义特征进行修改，改善了患者的动词命名能力。动词语义特征与其论元存在对应关系，是否可以通过动词论元语义类别了解所含征型的分布呢？Bird等[4]认为与名词相比，动词的感觉信息特征较少但是功能用途信息较多。Mätzig等[5]研究表明当仅动作命名损伤时，物品命名保留较好；而当物品命名损伤时，动作命名也受到损伤。动词的论元产出障碍和合规的论元搭配受损是失语症患者言语交流障碍的重要因素，论元动词的语义特征库有助于论元结构的组织，并可用于指导临床言语治疗。目前缺乏汉语动词的语义特征库，为了促进失语症患者正常语句的产出，亟需建立汉语动词语义特征库。

目前世界上多采用提名法在正常人群众采集语义特征[6]，然后对其进行分类进而分析不同类型的语义特征在概念中的分布及其关系。仅刘烨等[7]采用人群调查的方式对三类名词（动物、自然事物、人造物）的语义特征进行心理学研究，该研究只将3类名词的语义特征粗略分为外部特征、内部特征和功能特征，且没有语义特征的标注方案。本研究小组在先前的研究中采用提名法建立了汉语词汇语义特征质性分类方案[8]，探讨汉语名词语义特征分布特点[8—9]，为动词语义特征研究提供汉语方面的新依据。

本研究旨在通过采集正常人群的动词概念语义特征，对其加以分类，建立汉语动词语义特征库，探讨不同类型的语义特征在不同论元结构动词中的分布特点。

1 资料与方法

1.1 受试者

41名母语为汉语的健康人纳入本次研究。其中男性22名，年龄（25.62±2.77）岁，女性19名，年龄（25.39±1.65）岁，男性和女性之间年龄差异无显著性意义（t=0.309，P=0.759）。受教育年数男性（18±1.17）年，女性（17.84±1.21）年，两组间受教育年数差异无显著性意义（t=0.414，P=0.158）。所有受试者中除1名男性为左利手外，其余均为右利手。所有受试者均自愿参加本次实验。因有4名受试者基本资料填写不全，剔除本次研究中。

1.2 方法

1.2.1 刺激词：根据Snodgrass动词图按照词频选取30个常用动词，按照论元结构理论将动词分为一论元动词（6个），二论元动词（23个）和三论元动词（1个），并将30个动词按照经验分为5类：表达类、定位类、截获类、移位类和状态类。所选动词词频参考教育部语言文字应用研究所计算语言学研究室的语料库（http://www.cncorpus.org/）。

1.2.2 语义特征采集方式：通过语义特征提名完成动词的语义特征的采集。30个概念词随机制成2个手册，每个手册包括15个动词。手册包括2个部分，第一部分是填表说明，分别是指导语和一个范例，第二部分是15个刺激动词。要求受试者先对该刺激词进行熟悉度判定，之后写出该词的特征。不限制受试者填写的时间，但一般每个概念在1min内完成。每一册各由15名男性，15名女性完成，受试者可以完成2册，但同一名受试者不会重复做同一册。从而确保每一册都有男女各15名完成。

1.2.3 特征分类标注方案：动词语义特征标注方案参考林枫[8]等提出的汉语语义特征标注方案，对不适用于动词的部分进行修改。如“原始反应【用双手向前用力】和【脚的基本功能之一】，都涉及肢体，标注时明确是人的身体部位，则标注为人为配用物；不确定的情况下，标注为事件关系”。另外需注意的是：事件关系与所在事件的区别在于所在事件有人参与，而事件关系是无人参与的。

1.2.4 语义特征标注：动词语义特征采集结束后，由1名专业人员对所收集的原始反应条目进行标注，另外2名专业人员参与核对。例如，“洗的原始反应有【使物体清洁】、【把东西弄干净】、【清除污渍】等，统一标准化为〈用来清洁〉。推的原始反应是【需使用手】，因其包含受试者对推这个动作的一种评判及这个动作涉及的身体部位两个特征，则标注为〈带来

评判→需要使用手〉[评判]@[声称]和〈常共同出现→手〉[共现][人为配用物]。哭的原始反应是【眼睛,泪水】,标注为〈常共同出现→眼睛〉[共现][人为配用物]和〈常共同出现→泪水〉[共现][自然同在物]”。

1.3 指标

因本研究主要面向临床康复评估及治疗,故选取秩次作为语义特征的评价标准。秩次是指受试者对刺激词产生的原始条目的顺序,可以反映受试者对特征通达难易程度及特征对刺激词的重要性。秩次越靠前则越易通达,对刺激词也就越重要。

1.4 统计学分析

采用R软件3.2.2进行数据可视化和聚类分析[10]。

2 结果

2.1 基线数据

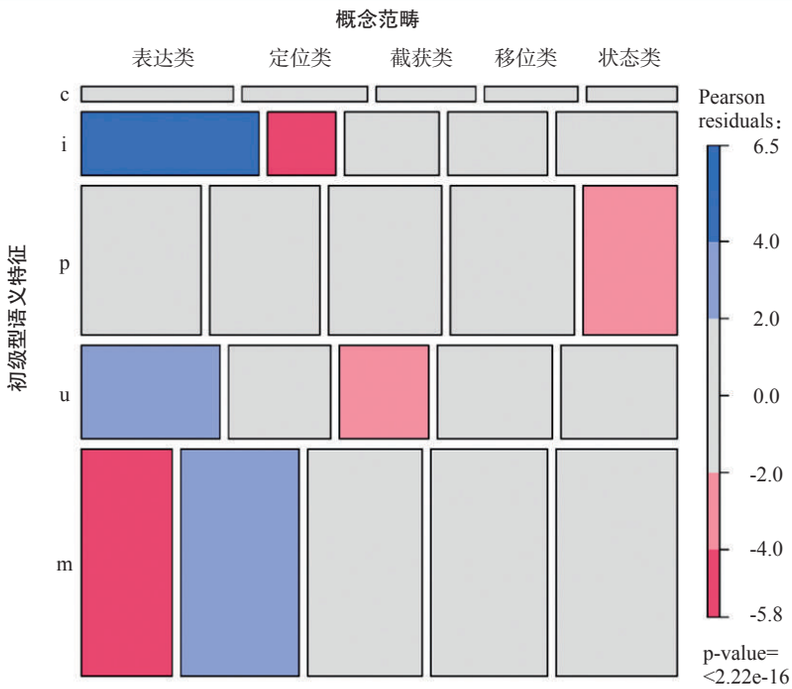
30个动词分为6个一论元动词,23个二论元动词和1个三论元动词,其熟悉度分别是8.36±1.63,7.64±1.98和6.34±2.25,三组间没有显著差异(F=3.316,P=0.0516)。性别对各概念的熟悉度分别是:男性7.82±0.98;女性7.57±0.86,性别对概念熟悉度无明显差异(t=0.91,P=0.3661)。对受试者采集的30个动词的原始反应条目进行标注,得到6963条语义特征。

2.2 不同概念范畴的语义特征分布

本研究用马赛克图显示各概念范畴的初级语义特征分布情况。近几年来欧美国家已经将马赛克图应用于语义特征研究[11],通过马赛克图可以反映出独立性检验、征型配比和皮尔逊残差分布,图中P值反映各方块取值是否独立,各方块的面积反映实际取值在不同维度及总体中的现实比例,方块的不同颜色反映实际取值与预期值的差[8],实现数据可视化。先前本研究组林枫、向伟华也已采用这种方法研究名词的语义特征分布情况[8—9]。

图1左侧的英文字母分别表示5个初级型语义特征:c表示分类范畴、i表示内省特征、p表示感觉信息、u表示功能用途、m表示杂类。图上方是5个概念范畴。矩形的面积表示语义特征被提名的次数,面积越大,被提名的次数越多。右下角的p值表示对每个矩形进行χ^2检验的结果,用于表示语义特征与范畴相关。图中各矩形的不同颜色代表皮尔逊残差检验的结果,即实际频数与预期频数之间比较的结果。蓝色表示显著高于预期,红色则表示显著低于预期,灰色表示无明显差异。颜色越深代表皮尔残差越大,颜色越浅代表皮尔逊残差越小。

图1 各概念范畴的初级型语义特征分布



从图1可以看出，5种初级型语义特征在各概念范畴中以杂类提名最多，次之是感觉信息。而分析各初级义征在各范畴中的分布发现：第一，表达类概念的内省特征和功能用途显著高于预期，而杂类则显著低于预期；第二，定位类概念与表达类相反，内省特征显著低于预期，杂类显著高于预期；第三，截获类概念功能用途显著低于预期；第四，状态类概念只有感觉信息显著低于预期。

2.3 不同概念范畴首位秩次语义特征分布

因本研究旨在探讨优先选择什么论元来进行训练，故我们选择首位秩次作为衡量指标。语义特征秩次是指受试者对概念产生的先后顺序，能反映出语义通达程度，首位秩次语义特征即为受试者产生的第一个特征。图2反映了不同范畴秩次语义特征分布情况。由此可以看出：①5个概念范畴的功能用途的首位秩次显著高于预期，而杂类的首位秩次显著低于预期；②内省特征的无论是首位秩次还是其他位秩次在表达类概念中是显著高于预期，而在定位类概念中显著低于预期；③功能用途特征的其他位秩次在除表达类的余概念范畴的中显著低于预期。

图2 各概念范畴首位秩次语义特征分布图

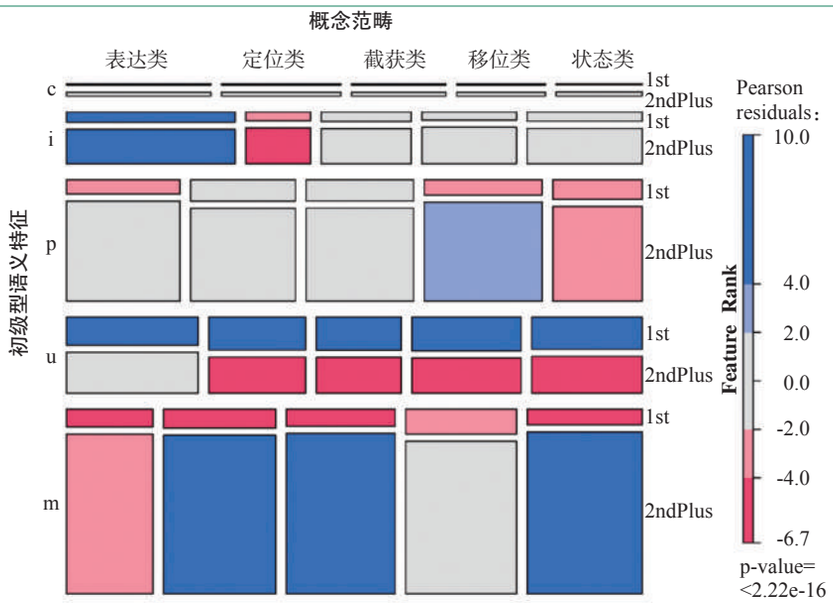


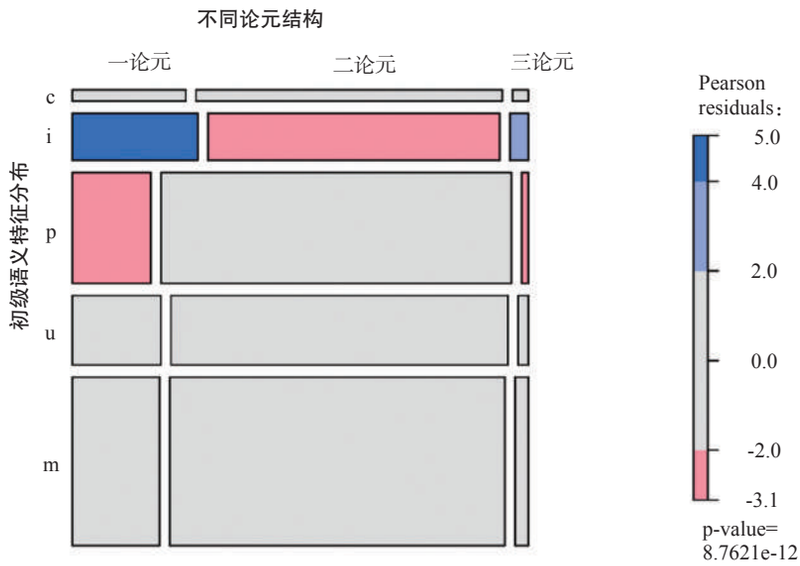
图2左侧英文字母为5个初级型语义特征，图上方是5个概念范畴，右侧标签1st表示首位秩次，2ndPlus表示其他位秩次。各矩形的面积表示语义特征提名次数，面积越大表示该语义特征被提名次数越多。右下角P值表示对各矩形进行三因素（语义特征类型、概念范畴和秩次）χ^2检验的结果。马赛克图的颜色表示对各矩形进行皮尔逊残差检验的结果，即进行实际频次与预期频次比较。蓝色表示显著高于预期，红色表示显著低于预期，灰色表示无明显差异。颜色的深浅表示皮尔逊残差值的大小，颜色越深表示残差越大，颜色越浅表示残差越小。

2.4 不同论元结构动词的语义特征分布

本研究意在了解不同论元的语义特征分布情况。从图3可以看出：①一论元结构动词和三论元结构动词的内省特征显著高于预期，而感觉信息显著低于预期；②二论元结构动词的内省特征显著低于预期。

图3左侧的英文字母分别表示5个初级型语义特征：c表示分类范畴、i表示内省特征、p表示感觉信息、u表示功能用途、m表示杂类。图上方是3个论元领域范畴，依次代表一论元动词、二论元动词和三论元动词。矩形的面积表示语义特征被提名的次数，面积越大，被提名的次数越多。右下角的P值表示对每个矩形进行χ^2检验的结果，用于表示语义特征与范畴相关。图中各矩形的不同颜色代表皮尔逊残差检验的结果，即实际频数与预期频数之间比较的结果。蓝色表示显著高于预期，红色则表示显著低于预期，灰色表示无明显差异。颜色越深代表皮尔逊残差越大，颜色越浅代表皮尔逊残差越小。

图3 各领域范畴的初级型语义特征分布



3 讨论

3.1 不同概念范畴的语义特征分布

通过比较不同范畴的初级特征实际频数发现,杂类所占实际频数最多,次之是感觉信息。因本文主要考察语义特征的初级型分布,而事实上本研究的杂类是指其一级类的百科知识,其中包含事件关系、空间关系、来源关系和联想关系。这与动作产生的时间(睡觉〈在晚上〉)、地点(吃〈在餐厅〉)、对该动作的解释(洗〈双手搓〉)及由该动作引发的联想(推〈出汗〉)等信息相关。感觉信息包括视觉、听觉、触觉、味觉、嗅觉、实体数和共现,共现则主要涉及产生动作的肢体(看〈眼睛〉)、动作常出现的情景(吃〈吃水果〉)以及配合该动作一起发生的事物(吹〈气流〉)。其中共现占感觉信息的主要,而余信息所占比例很小。Catricalà等[12]也有类似发现。

分析各范畴的语义特征分布发现,表达类的内省特征和功能用途显著高于预期。这是因为功能用途涉及动作的发出者,表达类的动作发出者多为人,而内省特征主要与受试者个人的主观感受相关,包括对目标词的情感、评价等。而表达类多与此相关,如:笑<开心>、唱〈愉悦〉。Spadacenta等[13]认为运动系统在理解动作词意义当中起重要作用,而动词的情绪极性(积极或消极)影响动作的执行和面部表情产生[14],哭〈难过〉。也有研究显示情绪在语言习得过程中起重要作用[15]。提示在进行言语训练时,可以先通过调节患者的情绪,改善其情绪状态,提高其治疗配合度;或模拟动作环境,引导情绪体验,提高治疗效果。

3.2 不同范畴概念首位秩次语义特征的分布

从不同范畴概念首位秩次语义特征分布情况来看,功能用途型的首位秩次在五种概念范畴中明显高,这提示动词与非生命性名词类似,功能用途特征较为显著。Marshall等[16]报道了一位伴杂乱语的患者RG在处理依赖外表的事物上表现非常差,如动物和自然现象,而处理功能性事物时表现较好。检查该患者的名词命名和动词命名时发现患者动词命名能力较名词好,且患者保留了动词的论元语义角色的能力,后者属于动词的功能特征之一[4]。

Gainotti等[17]认为不同范畴的知识分布在与大脑有关的知觉与运动区域,Bird[4]认为动作命名的语义表征必须依赖运动区,如果角回受损较多的话,会出现动词命名困难。神经影像学方面的研究也有类似发现,Boylan等[18]采用fMRI研究角回在动词论元,发现角回激活动词特定信息,如动词介导的语义角色关系和事件结构。首位秩次特征分布结果提示在进行临床言语治疗时,针对目标动词,可以首先提

示该词的功能用途类特征来促进其命名，如“笑”，可以先提示患者〈表达高兴的情绪〉。

词汇映射理论认为动词语义角色是按照阶层的形式排列的，施事>处所>客体[19]。简单来说，施事是指动作主体，比处所更重要，处所比客体更重要[20]。而上图中可以看出杂类在5个概念范畴的首位秩次均显著低于预期，杂类主要涉及处所、时间等信息，提示在给患者进行提示时，处所类等信息可以在后面进行提示，拓展语义的理解能力。

3.3 不同论元动词的语义特征分布

本文的研究目的是了解不同论元动词的语义特征分布情况，从图3种可以看出一论元动词和三论元动词的内省特征显著高于预期，而二论元动词的内省特征显著低于预期。造成这种差异的原因可能是本研究的三论元动词只有一个，三者中，一论元动词的语义丰富性最少，三论元的语义丰富性最多。一论元动词通常只涉及动作的发出者或者执行者，如：跑〈由谁用→运动员〉[用者]。二论元动词包括动作主体和客体。Javier等[21]研究发现处理及物动词比处理不及物动词的能力更强。对于失语症患者尤其是Broca失语患者而言，越是复杂的论元结构，动词越难产出。因为Broca区受损表达障碍显著，论元的增加会导致Broca失语患者动词产出能力下降，而Wernicke失语患者能够表达论元，但常无法察觉论元结构搭配不合规则[22]。提示在Broca失语症患者应从一论元动词训练开始，逐步增加论元动词的复杂性，另外在训练一论元动词时可以给予内省特征方面的提示，飞<速度快>；Wernicke失语症患者则强调复杂论元结构的搭配规则训练。另外，在实施不同论元动词训练时，可以借助语义特征分布提供的训练线索。

3.4 研究局限性和展望

本文采用语义特征提名测试来探究动词语义特征分布，将林枫等提出的汉语语义特征分类方案应用于汉语动词研究，建立汉语动词语义特征库，扩增了汉语康复分类法的表层临床衔接方案。尽管如此，本文尚存在一定局限性：第一，本研究采用的受试者是青年人，且受教育程度较高，而根据周亮等[23]的青年人与老年人词汇联想反应的研究，青年人与老年人的联想词汇特征和联想反应类型存在显著差异，并且林枫等[24]研究青年人与老年人认知联系网络的整体结构分析表明老年人的认知联系结构趋于精炼和内敛且存在更独立的概念群，而青年人则趋于联系更丰富和更富于拓展性。这些研究均提示还应该采集老年人的汉语动词语义特征进行分析。第二，本研究的样本量较小，还需要进一步的研究扩大样本量，探讨汉语动词语义特征分布情况是否与本研究相符。第三，本研究三论元动词取自于Snodgrass动词图的词频，只能选出一个，这对于研究动词复杂论元的语义是不够的，有研究表明，语义表征与轻松处理语义丰富性相关[21]，动词中三论元的语义丰富性最多。尽管本文存在一定局限性，但本文的研究有助于了解不同范畴概念的特点。对于进一步研究名词和动词差异，失语症患者表现出特定的名词缺失或者动词缺失有一定的指导意义，同时为临床实践提供帮助。

4 结论

本文将研究汉语名词语义特征分布的方法拓展至动词领域，并初步建立小型汉语动词语义特征常模。该库的建立进一步为临床言语认知功能训练实践提供量化、可视化和可扩增的素材选取策略。

参考文献

[1] Wambaugh JL, Ferguson M. Application of semantic feature analysis to retrieval of action names in aphasia [J]. The Journal of Rehabilitation Research and Development, 2007, 44(3): 381—394.

[2] Kim M, Adingono MF, Revoir JS. Argument structure enhanced verb naming treatment: two case studies. Contemporary Issues in Communication Science&Disorders[J]. 2007, 34:24—36.

[3] Thompson CK, Riley EA, Den Ouden DB, et al. Training verb argument structure production in agrammatic aphasia: Behavioral and neural recovery patterns [J]. Cortex, 2013, 49 (9): 2358—2376.

[4] Bird H, Howard D, Ffranklin S. Why Is a verb like an inanimate object? Grammatical category and semantic category deficits [J]. Brain and Language, 2000, 72(3): 246—309.

[5] Mätzig S, Druks J, Masterson J, et al. Noun and verb differences in picture naming: Past studies and new evidence [J]. Cortex, 2009, 45(6): 738—758.

[6] Mcrae K, Cree GS, Seidenberg MS, et al. Semantic feature

production norms for a large set of living and nonliving things [J]. Behavior Research Methods, 2005, 37(4): 547—559.

[7] 刘烨, 傅小兰. 自然概念语义特征提取的范畴效应 [J]. 心理科学, 2006, (2): 286—289.

[8] 林枫. 面向言语治疗的汉语词汇语义表征分析研究[D].南京;南京师范大学, 2015.

[9] 向伟华, 林枫, 江钟立. 汉语名词语义特征建模与分析 [J]. 中国康复医学杂志, 2015, (11): 1118—1124.

[10] R-Core-Team. R: A Language and Environment for statistical computing.Vienna,Austria.:RFoundation for statistical computing[Z].2015.Retrieved from http://www.R-project.org/.

[11] Lenci A, Baroni M, Cazzolli G, et al. BLIND: a set of semantic feature norms from the congenitally blind [J]. Behav Res Methods, 2013, 45(4): 1218—1233.

[12] Catricala E, Della Rosa PA, Plebani V, et al. Abstract and concrete categories? Evidences from neurodegenerative diseases [J]. Neuropsychologia, 2014, 64:271—281.

[13] Spadacenta S, Gallese V, Fragola M, et al. Modulation of arm reaching movements during processing of arm/hand-related action verbs with and without emotional connotation [J]. PloS one, 2014, 9(8): 1-15.

[14] Foroni F, Semin GR. Language that puts you in touch with your bodily feelings: The multimodal responsiveness of affective expressions [J]. Psychological Science, 2009, 20 (8): 974—980.

[15] Vigliocco G, Kousta ST, Rosa PAD, et al. The neural representation of abstract words: The role of emotion [J]. Cerebral Cortex, 2014, (7): 1767—1777.

[16] Marshall J, Chiat S, Robson J, et al. Calling a salad a federation: An investigation of semantic jargon. Part 2—verbs [J]. Journal of Neurolinguistics, 1996, 9(4): 251—260.

[17] Gainotti G, Silveri, Maria C, et al. Neuroanatomical correlates of category-specific semantic disorders: a critical suney [J]. Memory, 1995, 3(3/4):247—264.

[18] Boylan C, Trueswell JC, Thompson-Schill SL. Compositionality and the angular gyrus: A multi-voxel similarity analysis of the semantic composition of nouns and verbs [J]. Neuropsychologia, 2015, 78:130—141.

[19] 袁毓林.论元角色的层级关系和语义特征[J]. 世界汉语教学, 2002, 03): 10—22+2.

[20] 潘海华. 词汇映射理论在汉语句法研究中的应用 [J]. 现代外语, 1997, 4:5—18.

[21] Rodriguez-Ferreiro J, Andreu L, Sanz-Torrent M. Argument structure and the representation of abstract semantics [J]. PloS one, 2014, 9(8): 1—7.

[22] Den Ouden DB, Fix S, Parrish TB, et al. Argument structure effects in action verb naming in static and dynamic conditions [J]. J Neurolinguistics, 2009, 22(2): 196—215.

[23] 周亮, 江钟立, 林枫, 等. 青年人和老年人词汇联想反应的研究 [J]. 中国康复医学杂志, 2008, 23(4): 297—300.

[24] 林枫, 江钟立, 周亮, 等. 青年人和老年人认知联系网络的整体结构分析[J]. 中国康复医学杂志, 2008, 23(4): 291—296.