

基于语料库的多义动词“开”行为特征分析

王 婷

西南大学外国语学院, 重庆

收稿日期: 2023年5月18日; 录用日期: 2023年6月22日; 发布日期: 2023年6月30日

摘 要

二十一世纪以来, 认知语言学界出现了量化转向趋势。语料库与统计手段的有机结合为认知词汇语义学研究提供了新的研究方法。本文运用基于语料库的行为特征分析法, 从北京语言大学BCC语料库报刊子语料库中析出1000条语例, 对汉语动词“开”的多义性进行了研究。通过对动词“开”进行语义、形态和句法等语言学特征的标注, 并使用层次聚类分析法进行量化研究, 本文发现: 动词“开”的原型义为“打开”, 其他多个义项围绕该原型义向不同方向拓展延伸, 可以分为“位移义”“变化义”和“创建义”三大类。每一类义项既有共性又有个性, 共同构成了动词“开”的多义性网络。

关键词

一词多义, 行为特征分析法, 动词“开”

The Polysemy of the Chinese Verb “*Kai*”: A Corpus-Based Behavioral Profile Analysis

Ting Wang

College of International Studies, Southwest University, Chongqing

Received: May 18th, 2023; accepted: Jun. 22nd, 2023; published: Jun. 30th, 2023

Abstract

Cognitive Linguistics has gone through the quantitative turn since the 21st century. The combination of corpus and statistics provides a new research method for cognitive lexical semantics. Based on the corpus-based behavioral profile analysis, the present study collected 1,000 instances from BCC corpus to investigate the polysemy of the Chinese verb “*Kai*”. After manual annotation on semantic, morphological and syntactic characteristics of this verb, hierarchical agglomerative cluster analysis is employed to do quantitative studies. It turned out to be that the prototypical meaning of “*Kai*” is to open with the other senses divided into three classes, *i.e.*, to move, to change and

to create. All of these senses both share common characteristics and features in their own way.

Keywords

Polylemy, Corpus-Based Behavioral Profile Analysis, Chinese Verb “Kai”

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

词汇在数量上的有限性催生了一词多义现象。关于多义词的研究在语言学界尤其是认知语言学届颇受重视。历年来, 认知语言学家们纷纷运用原型理论[1]、辐射范畴理论[2]、认知隐喻理论和认知转喻理论[3]等来描述和解释语言的一词多义现象, 并取得了丰硕的成果。随着认知语言学的量化转向(the quantitative turn), 语料库、脑电位、功能性核磁共振等量化研究方法开始进入研究视野, 在进一步提高语言学作为一门科学的科学性、客观性、严谨性的同时, 也大大拓宽了认知语言学的研究视角, 许多具有创新性的实证研究方法应运而生。

基于语料库的行为特征分析法(Corpus-based Behavioral Profile Analysis, 以下简称“BP 分析法”)由 Gries 和 Divjak [4]在 Hanks [5]的研究基础上提出, 充分结合了语料库法和多元统计法。其理论假设是: 分布的相似性反映了功能的相似性, 其中功能囊括了语义、语用等方面。具体而言, 相似义项之间是相互连接的, 连接的强度反映了它们在语音、句法、语义等方面显示出相似性的可能性[6]。在国外, 该方法被广泛应用于认知词汇语义学, 研究不同语种的同义关系、多义关系、反义关系等词汇关系。在国内, 自吴淑琼、刘迪麟[7]对 BP 分析法作了系统性阐释以来, 渐有学者将该方法运用到对现代汉语词汇关系的研究(江艳艳[8]; 吴淑琼等[9]; 罗思明、王佳敏[10])。

动词“开”是现代汉语中一个常见的多义词, 在前人研究中已得到较多关注。王冬芝、萧国政[11]根据事件语义理论, 将动词“开”的语义划分为位移、开始、出现三大类, 并进一步根据致因、动体、方式、背景等要素差异分解出不同子类。其中, 位移是人们建构和理解“开”不同义项的概念起点。孙安琪[12]通过自建语料库考察了动词“开”不同义项在真实语境中的运用情况, 归纳出其基本义——打开, 即使关闭着的东西不再关闭。围绕这一基本义, 通过发生位移、改变“打开”的方式、分离物体又可以引申出其他义项。张琪[13]对比分析了汉语动词“开”和英语动词“open”, 并借助概念隐喻和概念转喻理论解释了多义性的产生。即是说, 从原型义“将封闭的空间敞开”到非原型义的延伸过程行为实体和空间域、空间域和时间域、行为实体和思维域之间映射的结果。

综上所述, 有关动词“开”的多义性研究已经取得了一定成果, 但以往的研究多以定性分析为主, 对真实语料的运用和分析缺乏深入性, 且未见有统计方法的使用。有鉴于此, 本文拟运用 BP 分析法考察动词“开”的多义性, 探索其原型义和多义性网络, 总结归纳不同义项之间的区别与联系, 以期提高对该多义词的认识。

2. 研究方法

Gries & Otani [14]对 BP 分析法在同义关系和多义关系研究上的应用做了规范, 总结出以下四个操纵步骤。

第一步,通过索引程序、编程语言、语料库接口等手段检索出研究对象的所有语例。本研究选取的是北京语言大学 BCC 语料库的报刊子语料库,以动词“开”为关键词,检索出共 664,628 条语例。通过随机抽样的方法,人工剔除动词“开”加名词生成专有名词(如开罗)、“开”做趋向动词等不在研究范围内的情况,最后得到 1000 条语例以供分析。

第二步,人工分析每条语例的各项特征,并标注标识码。标识码(ID tag)这一概念由 Atkins [15]率先提出,指的是词汇在形态、句法、语义、搭配等方面的特征。结合汉语实际情况,本研究共选取了三种类型的标识码进行标注:形态、句法、语义,每个大类下又分多个小类,具体标注特征如下。

2.1.1. 形态特征

汉语动词的形态变化共有五类[16],在本研究所选取的语例中未出现重叠式(即“开开”)这一形态。因此,在形态层面共有以下 4 个标识码:1) 复合式,如“开剥”“开建”“开立”;2) 词尾,如“开了”“开着”“开到”“开过”;3) 词嵌,如“开不到一块”“开得过长”“开的太猛”;4) 零形态,即“开”。其中,7 个标识码水平见表 1:

Table 1. Morphological ID tags

表 1. 形态标识码

标识码类型	标识码	标识码水平
形态	复合式	V1 + V2
	词尾	“了”类、“着”类、“过”类
	词嵌	“不”类、“得”类
	零形态	开

2.1.2. 句法特征

与 BP 分析法的基本假设契合的是,现代汉语研究亦提出,确定一个语言结构里某个组成成分的性质时,不仅要考虑该成分本身的性质以及跟它发生关系的其它成分的性质,还要考虑整个结构的性质[17]。因此,在句法特征的选取上,本研究主要考虑了三个因素。首先,作为动词,“开”的基本特征体现在是否及物,即及物性。其次,在现代汉语中,动词主要充当谓语成分,动词“开”的同现成分包括前成分(主语)、后成分(宾语、补语)和可前可后成分(状语)。这些句子成分对动词“开”不同义项的性质特征也造成了一定的影响。最后,在真实语料中,每个动词都存在于句子这一更大的语言单位中。因此,句子结构和句子类型也构成了动词“开”的两大句法特征。总之,在句法层面共有 7 个标识码,其中包含 30 个标识码水平,如表 2 所示:

Table 2. Syntactic ID tags

表 2. 句法标识码

标识码类型	标识码	标识码水平
句法	主语	无生命、有生命、零主语
	宾语	结果、对象、工具、施事、零宾语
	补语	趋向补语、程度补语、结果补语、零补语
	状语	时间状语、范围状语、方式状语、处所状语、否定状语、关联状语、无
	句子结构	动宾结构、主谓结构、联合结构、偏正结构
	句子类型	直陈句、祈使句、反问句、感叹句、疑问句
	及物性	及物、不及物

2.1.3. 语义特征

《汉语动词用法词典》《现代汉语词典(第7版)》《现代汉语八百词》等权威词典分别列出了16、18、14个动词“开”的义项。结合前人研究和所选语例的实际情况,本研究共确定了9个动词“开”的义项,如表3所示:

Table 3. Semantic ID tags

表 3. 语义标识码

标识码类型	序号	标识码	语例
语义	1	建立	山东省济南市郊段店大队近几年种蔬菜、办工厂、开商店、搞运输,收入逐年增加。
	2	打开	人走灯灭,人离扇停,光线充足时,不要开灯。
	3	展开	我们号召全国人民都来学习科学精神,使它开出更加绚丽的社会主义的精神文明之花。
	4	打通	接着又一锹一镐,除掉杂草、卵石,开出了七亩荒地,种下了蔬菜。
	5	操纵	幸好司机见到这个情况,立即把车子开回,把服毒的父女送进基隆医院急救脱险。
	6	举行	现在开完会议,人们都赶快下去,没有一个人愿意留在机关。
	7	开始	本届比赛吸引了2万多人参加,其中4000多人还没有开跑,整个比赛就因为爆炸而终止。
	8	写出	商店雇请临时工装运货物,付费只开白条。
	9	出发	一月三十一日,解放军正式开入北平城,这个世界著名的古都从此解放。

第三步,将标注结果转换为共现表(见表4),以清晰地展示研究对象与每个标识码共同出现的相对频率。行为特征指的就是一个词汇同现百分比的向量(vector)。

Table 4. Relative frequency of co-occurrence (sample)

表 4. 相对频率共现表示例

标识码	标识码水平	建立	打开	展开	打通	操纵	举行	开始	写出	出发
复合式	V1 + V2	0.0149	0.0455	0	0	0.0105	0	0.6422	0	0.0233
词尾	“了”类	0.1940	0.0065	0.3519	0.1630	0.3000	0.0052	0.0092	0.1979	0.6047
	“着”类	0	0	0	0	0.0053	0	0	0	0
	“过”类	0	0	0	0	0	0.0052	0	0	0
词嵌	“不”类	0	0	0	0	0	0.0052	0	0	0
	“得”类	0	0.0065	0	0	0.0053	0.0052	0	0	0
零形态	开	0.7910	0.9416	0.6481	0.8370	0.6789	0.9794	0.3486	0.8021	0.3721

第四步,通过探索技术和其他统计方法对相对频率共现表进行评估与分析。本研究选取的统计方法是层次聚类分析法(hierarchical agglomerative cluster analysis, HAC),即通过建立一个集群层次结构来聚类分析。在进行聚类分析时,参照前人研究,距离度量选择欧式距离(Euclidean distance),聚类方式则选择Ward最小方差法。

3. 研究结果与讨论

通过在 R studio 进行聚类分析, 得到树形图 1。整体而言, 动词“开”的九个义项形成了三大聚类群: 聚类群一包含“建立”和“举行”两个义项; 聚类群二包含“操纵”“开始”“打通”和“写出”四个义项; 聚类群三包含“展开”“出发”和“打开”三个义项。从纵向来看, 不同义项之间形成聚类的时间有先后之分。Divjak & Fieller [18] 提出聚类分析的总原则: 义项之间聚集的时间早晚与相似度大小成正比。即是说, 越早形成聚类的义项之间越相似, 相反越晚形成聚类的义项之间差异越大。从图上可知, 义项“操纵”和“开始”最先形成聚类, 随后和义项“打通”及“写出”聚集成一个聚类群, 说明“操纵”和“开始”之间的相似度最高。从横向来看, 不同义项之间形成聚类的聚类有远近之分, 距离越近义项之间相似度越高, 反之距离越远的义项之间相似度越低。从图 1 可知, 聚类群一与聚类群三之间相隔较远, 因此两个聚类群之间的语义相似度也较低。

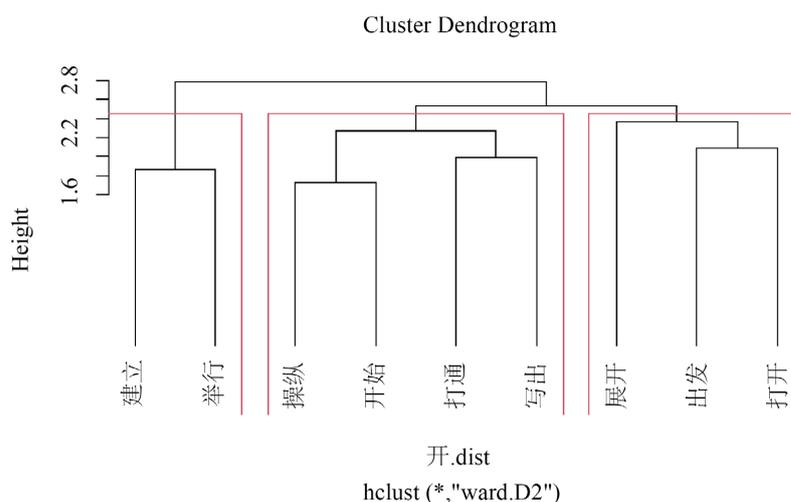


Figure 1. Tree diagram of the clustering of senses of “Kai”

图 1. “开”的义项聚类分析树形图

3.1. 动词“开”的原型义

自提出以来, 原型范畴理论在认知词汇语义学得到广泛应用的同时, 也引发了学者们关于如何确定原型义的讨论。有学者认为, 义项的原型性与使用频率成正比, 因此使用频率越高的义项越可能被识别成原型义[19]。但该看法因忽略了词汇在实际应用中所涉及的语境等因素而遭致质疑。Gilquin [20] 提出, 一个词汇的义项之所以能成为原型义, 不仅是因为其使用频繁, 还因为它与该范畴的其他成员有很高的家族相似性。因此, 频率和原型性之间的关系并不总是直接的。在 BP 分析法中, 行为特征被当做识别原型意义的证据[21]。具体而言, 拥有标识码水平数最多的义项在使用时受到的限制最小, 因此具有原型性。基于此, 本研究统计了动词“开”九个义项所拥有的标识码水平总数:

从表 5 可知, 在动词“开”的九个义项中, “打开”拥有的标识码水平总数最多, 而“建立”拥有的标识码水平总数最少。据此推断, 义项“打开”是动词“开”的原型义。进一步查看语例发现, 当表示“打开”义时, 动词“开”的形态变化较为丰富, 囊括了复合式、词尾、词嵌、零形态四大类。句法上的使用也较为灵活, 前可接无生命主语、有生命主语或零主语, 后既可跟对象或结果作宾语, 也可以跟零宾语, 补语则囊括了处所、方式、否定、关联、时间等多种形式。动词“开”所在的句子, 结构上可以是动宾结构、主谓结构、联合结构、偏正结构, 类型上可以是直陈句、祈使句、反问句、感叹句、

疑问句。因此，通过运用行为特征，结合在真实语境中对语例的考察，可以识别出动词“开”的原型义——“打开”。

Table 5. Total of ID tag levels

表 5. 标识码水平总数

义项	建立	打开	展开	打通	操纵	举行	开始	写出	出发
标识码水平数	22	29	23	24	28	28	27	26	23

3.2. 动词“开”的多义网络

根据层次聚类分析结果，动词“开”的九个义项聚集成三个聚类群，以原型义“打开”为中心拓展和延伸出不同义项，从而形成如图 2 所示的下多义网络：

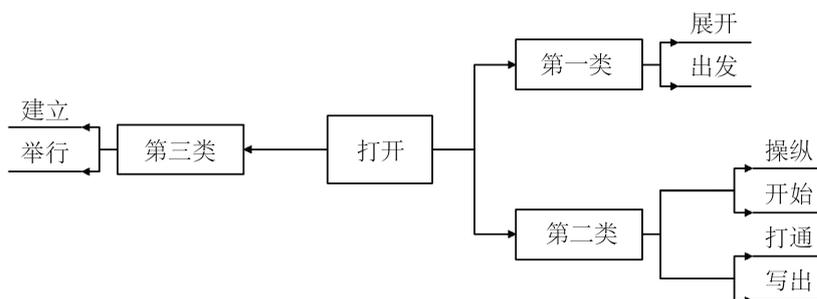


Figure 2. Diagram of the polysemy network of “Kai”

图 2. “开”的多义网络

原型义“打开”的拓展分为两个方向。从语义上看，左边的拓展路径揭示了新事物的产生，而右边的拓展路径展现了事物本身所发生的变化。第一类包括“展开”和“出发”两个义项，动词“开”从“打开”义拓展到“位移”义，“位移”义又可细分为动态的位移（“出发”）和静态的位移（“展开”）。前者指事物从一个地方移动到另一个地方，发生了位置上的变化；后者指事物自身的物理性质发生改变，即物理状态从小变大。如下例：

- (1) 一连英国军队还曾经开到离西柏林边界仅五百码的地方。
- (2) 但是你方武装部队却开入该病院，向我方被俘人员投掷毒气弹。
- (3) 昔日烧焦过的土地，却仍能为我们开出灿黄灿黄的菜花。

第二类包括“操纵”“开始”“打通”和“写出”四个义项，动词“开”从“打开”义拓展到“变化”义，即事物的状态发生了变化，如受到其他人或物的控制（“操纵”）、着手进行某一动作（“开始”）、打开通路（“打通”）、填满单据信件（“写出”）。相较于第一类在物理距离或性质上发生变化，第二类的“变化”义着重体现在事物本身所具有的性质和状态，与外界联系较小。如下例：

- (4) 他们还到铁道建筑工程队和门头沟等地教人家开机器呢！
- (5) 干净利落的开打，造型精美的椅子功，使人眼花缭乱。
- (6) 解放初期，东乡县基本上是采取挖山塘、开平塘的办法。
- (7) 先到单位开证明并拿着住房产权证，到物业管理部门交钱开单子。

第三类包括“建立”和“举行”两个义项，动词“开”从“打开”义拓展到“创建”义，即动作的结果是新事物的产生。在语例中，“建立”和“举行”带的宾语多为由这种动作或行为产生的结果，如公司、商店、会议等。第一类和第二类涉及的对象往往是已存在的事物，而第三类涉及新事物的产生，因此两者的拓展路径分处不同方向。如下例：

- (8) 现在，我们夫妻俩开饭店，每年收入几万元。
 (9) 旧历五月下旬，全家又召开了家庭会，订出三天一开会，十天一总结的制度。
 (10) 身后不举办任何丧事礼仪，不开追悼会，不搞遗体告别，不设灵堂。

3.3. 动词“开”各义项的区分度

区分度指各义项之间的异同。基于语料库的研究对区分度的研究基于以下假设：句法结构的重叠性越大，在语义结构上就越接近。将该假设应用到 BP 分析法，表现在义项不同，相对共现频率最高的标识码水平也有所差异。在上节讨论的基础上，本研究统计了三类义项在形态和句法层面共现频次最大的前 10 个标识码水平，如表 6 所示。

Table 6. ID tag levels with highest frequency

表 6. 共现频次最大的标识码水平

第一类		第二类		第三类	
标识码水平	频次	标识码水平	频次	标识码水平	频次
直陈句	228	直陈句	474	结果宾语	260
及物	205	零形态	321	及物	260
零形态	196	及物	320	直陈句	254
对象宾语	190	零补语	271	零形态	243
零补语	158	无状语	269	零补语	184
零主语	110	零主语	258	动宾结构	166
时间状语	104	主谓结构	213	零主语	155
无状语	84	结果宾语	197	无状语	118
无生命主语	76	动宾结构	183	主谓结构	90
趋向补语	69	趋向补语	169	有生命主语	81

通过对比分析发现，动词“开”各义项共有部分形态句法特征，如多出现在直陈句中，主要作为及物动词使用，形态变化上以零形态为主，等等。与此同时，不同类的义项又各自享有一些独有的形态句法特征。第一类中，动词“开”表示使关闭着的东西不再关闭、紧缩的东西展开、静止的东西运动起来，因此宾语多为该动作或行为的对象。尽管动词“开”表示“打开”义时主语较为灵活，但在该类中，“展开”义和“出发”义的主语往往分别对应植物和部队，因此无生命主语出现的频次较大。例如：

- (11) 有的同学晚上睡觉时，宁肯摸黑也不开灯。
 (12) 当苏联装甲部队开到出事地点时，游击队已安全撤退。

第二类中，动词“开”表示事物的变化，动作性较强，句子结构偏向于动宾结构和主谓结构。其中，“操纵”“打通”两个义项从语义上与表示事物运动方向的趋向补语联系较为紧密，因此该类与趋向补

语共现的频率也较高。例如：

(13) 几分钟以后，一列军用火车满载着士兵轰隆隆开过来。

(14) 过去吃水都困难的千山坡，而今已开出八十多亩水稻田。

第三类中，动词“开”表示人的行为动作导致了工厂、商店、医院等的开办或会议的举行。因此，往往由有生命的事物作主语，如人称代词、专有名词、集体名词等。由于涉及新事物的产生，宾语多为这种行为动作产生的结果。例如：

(15) 临走那一天，警予同志刚开完会匆匆地赶来送我们。

(16) 为此，父亲带他退学、变卖家产、开台球城、背井离乡……

4. 结语

受到认知语言学量化转向的启发，本文运用 BP 分析法对汉语多义动词“开”进行了量化研究。在对 1000 条发生在真实语境的语例进行分析后，研究发现动词“开”的原型义为“打开”，以该原型义为核心延伸出三大类语义，分别是“位移义”“变化义”和“创建义”，从而构建出动词“开”的多义网络。在以上三大类中又存在着九个义项，义项与义项之间既有共性又有个性。第一类的主要特征体现在前接无生命主语、后跟对象宾语；第二类常与趋向补语共同出现；第三类的主语多为有生命主语，宾语多为结果宾语。

BP 分析法对统计方法的运用在一定程度上克服了内省法的弊端，尽管它自诞生到引入国内所经历的时间尚短，但已在词汇关系的研究上做出了较大贡献。目前，国内运用 BP 分析法讨论多义词的研究还不遑多见，本文对动词“开”的探索既借助了语料库、编程语言等量化工具，又融合了对真实语料的定性分析，充分考察了动词“开”在语义、句法、形态三方面的行为特征，以期形成对该词的全面了解，并对借助 BP 分析法研究现代汉语中的多义词提供一定的借鉴意义。

参考文献

- [1] Lakoff, G. and Johnson, M. (1999) *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books, New York.
- [2] Lakoff, G. (1987) *Women, Fire and Dangerous Things*. The University of Chicago Press, Chicago. <https://doi.org/10.7208/chicago/9780226471013.001.0001>
- [3] Leech, G. (1974) *Semantics*. Penguin, London.
- [4] Gries, S. and Divjak, D. (2009) Behavioral Profiles: A Corpus-Based Approach towards Cognitive Semantic Analysis. In: Vyvyan, E. and Pourcel, S., Eds., *New Directions in Cognitive Linguistics*, John Benjamins, Amsterdam, 57-75. <https://doi.org/10.1075/hcp.24.07gri>
- [5] Hanks, P. (1996) Contextual Dependency and Lexical Sets. *International Journal of Corpus Linguistics*, 1, 75-98. <https://doi.org/10.1075/ijcl.1.1.06han>
- [6] Gries, S. (2010) Behavioral Profiles: A Fine-Grained and Quantitative Approach in Corpus-Based Lexical Semantics. *The Mental Lexicon*, 3, 323-346. <https://doi.org/10.1075/ml.5.3.04gri>
- [7] 吴淑琼, 刘迪麟. 词汇语义的语料库量化研究: 行为特征分析法[J]. 英语研究, 2020(1): 153-164.
- [8] 江艳艳. 温度词“冷”的语义演变研究: 基于语料库的行为特征分析[J]. 外语研究, 2022, 39(6): 33-41.
- [9] 吴淑琼, 刘迪麟, 刘青. 基于语料库的“确认”类同义副词的行为特征研究——以“的确、确实、实在、着实”为例[J]. 外语教学, 2021, 42(5): 19-25.
- [10] 罗思明, 王佳敏. 汉英时间词轴线类型特质及其语言蕴含共性[J]. 外语与外语教学, 2021, 319(4): 13-22+147.
- [11] 王冬芝, 萧国政. 多义词的事件语义及演化动力——以动词“开”为例[J]. 湖北社会科学, 2017(10): 125-130.
- [12] 孙安琪. 单音多义动词“开”的语义分析及汉语二语习得建议[J]. 现代语文, 2018(3): 145-151.

-
- [13] 张琪. 基于语料库的汉语“动词开-名词”与英语“动词 open-n”对比研究[J]. 海外英语, 2019(10): 73-75.
- [14] Gries, S. and Otani, N. (2010) Behavioral Profiles: A Corpus-Based Perspective on Synonymy and Antonymy. *ICAME Journal*, **34**, 121-150.
- [15] Atkins, B. (1987) Semantic ID Tags: Corpus Evidence for Dictionary Senses. *Proceedings of the Third Annual Conference of the UW Centre for the New Oxford English Dictionary*, Waterloo, 9-10 November 1987, 17-36.
- [16] 俞敏. 汉语动词的形态[J]. 语文学学习, 1954(4): 43-51, 42.
- [17] 朱德熙. 朱德熙文集[M]. 北京: 商务印书馆, 1991: 64.
- [18] Divjak, D. and Fieller, N. (2014) Cluster Analysis: Finding Structure in Linguistic Data. In: Glynn, D. and Robinson, J., Eds., *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*. John Benjamins, Amsterdam, 405-442. <https://doi.org/10.1075/hcp.43.16div>
- [19] Schmid, H. (2000) English Abstract Nouns as Conceptual Shells: From Corpus to Cognition. Mouton de Gruyter, Berlin. <https://doi.org/10.1515/9783110808704>
- [20] Gilquin, G. (2008) What You Think ain't What You Get: Highly polysemous Verbs in Mind and Language. In: Lapaire, J., Desagulier, G. and Guignard, J., Eds., *From Gram to Mind: Grammar as Cognition*, Presses Universitaires de Bordeaux, Pessac, 235-255.
- [21] Divjak, D. (2010) Structuring the Lexicon: A Clustered Model for Near-Synonymy. Mouton de Gruyter, Berlin. <https://doi.org/10.1515/9783110220599>