



汉语近义词搭配知识库构建及其可视化研究

李娟

2021年1月7日

目 录

- 1 引言
- 2 搭配资源研究现状
- 3 汉语近义词搭配自动抽取
- 4 汉语近义词搭配知识库构建研究
- 5 汉语近义词搭配结果可视化
- 6 总结

1. 引言

- 搭配信息是单词习得的重要语言信息，对于二语学习者来说，搭配知识不但有助于他们提高语言表达的流利程度，还能帮助学习者克服母语负迁移，输出更地道的目的语，是语言教学的理想单位 (Nattinger, 1992)。
- 林杏光(1994)指出:对外汉语的教学实践告诉我们，外国人学习汉语，往往难于掌握汉语词语的搭配规律，因而词语搭配的问题在对外汉语中至关重要。
- 然而，从词汇习得效果来看，搭配不当往往是词语使用过程中最突出的问题之一(邢红兵，2013)。

1. 引言

全球汉语中介语语料库（V1.0）中词错误类型	数量	占比
[Cb]词语误用	37735	30.31%
[Cq]缺词	32825	26.37%
[Cd]多词	30072	24.16%
[C?]偏误存疑	6241	5.01%
[Cz]生造词	4679	3.76%
[Cw]外文词	2999	2.41%
[Cx]错序词	1284	1.03%
[Cdj]同义叠加	455	0.37%
[CI]离合词	438	0.35%

我也【回去】沈阳。
我也[回]沈阳。

吃这种东西【可以】【得到】病。
吃这种东西[会][得]病。

其实中国人【太】勤劳的，一天从很早开始很晚才结束，日程很满，从早到晚一整天【充满】忙碌。
其实中国人[很]勤劳的，一天从很早开始很晚才结束，日程很满，从早到晚一整天[都在]忙碌。

在全球汉语中介语语料库（V1.0，网址：<http://qqk.blcu.edu.cn/#/search/searchModuleIndex>）中检索词错误的中介语料，被标注为词错误的语料共有124494句，其中被标注为[Cb]的“词语误用”语料共有37735句，是词错误占比最高的类型，占到了词错误总量的30.31%。

➤针对这一问题，邢红兵(2013)等学者指出，要在词汇本体知识研究的基础上，建立目的语词汇知识体系，特别是搭配知识体系，重视从语料库等资源中提取构建词语搭配数据库，包含词语的搭配信息及其频次。

2. 搭配资源研究现状

2.1 传统的搭配词典

➤ 如《现代汉语搭配词典》

- 搭配准确，具有拼音、词义、搭配类型等信息
- 搭配词数量较丰富
- 缺少词性、例句
- 缺少词频、互信息等计量信息



(梅家驹主编，上海：汉语大词典出版社，1999.12)

8 ān 安

…得无法安宁：吵~/闹~

【安排】 ānpái 有条理地处置。

语 安排…：~劳动力/~复员军人/~外来民工/~待业青年/~下岗人员//~工作/~生产/~劳动/~农事/~生活/~学习/~时间/~人事/~后事/~膳宿/~吃住/~议事日程/~工作单位//~两个房间/~吃食堂//~好/~妥当/~周到/~停当/~就绪

…安排：统一~/集中~/分头~/单独~/妥善~/精心~/细心~/巧妙~/临时~/暂且~/作出~/另作~/进行//设法~/千方百计~

…的安排：妥帖~/适当~/合情合理~/实事求是~

安排得…：~妥妥当当/~清清楚楚/~井井有条/~万无一失/~使大家满意/~莫明其妙

安排到…：~基层/~招待所/~单间房

为…作出安排：~学生~/~客人~/~旅客~

【安全】 ānquán 不危险。

词 安全…：~灯/~帽/~阀/~感

语 安全…：~第一//~抵达/~到达/~撤离/~生产/~操作/~驾驶/~行车/~飞行/~起飞/~返航/~降落/~启航/~着陆/~用电//~问题/~措施/~教育/~条约/~地带/~系数/~炸药/~火柴/~玻璃

心/~可以高枕无忧

【安适】 ānshì 安定舒适。

语 安适…：~如常

…安适：实在~/确实~/感到~/觉得~

安适的…：~生活/~环境/~房间/~疗养院/~生活环境/~沙发/~太师椅/~席梦思床

安安适适地…：~睡了一晚/~过了一天/~洗了一个痛快澡/~吃了一顿太平饭

…得安安适适：睡~/生活~/日子过~

【安慰】 ānwèi ①宽心。

语 …安慰：寻求~/得到~/感到~

…的安慰：很大~/莫大~/知人知心~

心中安慰了…：~不少/~许多

②使人宽心。

词 ~奖/~赛/~话

语 安慰…：~病人/~伤员/~老人/~残疾人//~几句

…地安慰：百般~/耐心~/细言细语~/细声细气~/好言好语~/体贴入微~

老当益壮的~赛/精神可嘉的~奖

【安稳】 ānwěn 平安稳当。

语 安稳的…：~生活/~工作/~职业/~情绪/~举止

安稳地…：~过几天太平日子/~过

2. 搭配资源研究现状

2.1 传统的搭配词典

➤ 如《现代汉语实词搭配词典》

- 搭配准确，具有词义、词性、短语结构等信息
- 搭配词数量丰富
- 缺少例句
- 缺少词频、互信息等计量信息

安排 ānpái (动)有条理、有次序地处理、安置。

【主】● ~+动： a. ~实现 ~落空 ~落实 ~取消 ~批准 ~通过 b. ~推迟 ~提前 ~改变 ~恢复 ~放弃 ~拖后 ~延期 c. ~开始 ~继续 ~停止 ~结束 ● ~+形： a. ~容易 ~困难 ~顺利 b. ~有效 ~无效 c. ~妥当 ~巧妙 ~正确 ~失策

【谓】● 名+~： a. 上级~ 国家~ 公司~ 组织~ 单位~ 厂长~ 主席团~ 医院~ 大夫~ 上帝~ 命运~ 老天爷~ b. 参见[谓]●~+名, a: 工作~等 c. 参见[谓]●~+名, b: 活动~等 d. 参见[谓]●~+名, c: 事情~等 e. 参见[谓]●~+名, d: 人员~

【宾】动+~： a. 开始~ 继续~ 停止~ 恢复~ 进行~ 着手~ 结束~ 完成~ b. 想~ 喜欢~ 盼望~ 希望~ 决定~ 赞成~ 同意~ 要求~ 需要~ 主张~ 允许~ 禁止~ 拒绝~ 听从~ 服从~ 听凭~ 接受~ 反对~ 抵制~ 支持~ 给予~ 予以~ c. 负责~ 管~ 参加~ 参与~ 改变~ 变更~ 落实~ 实现~ 取消~ 斟酌~ 实施~ d. 想法~ 设法~ 计划~ 打算~ e. 得到~ 获得~ 受到~ 争取~ 等待~

【中】形+~： a. 容易~ 难~ 好~ b. 重新~ 集中~ 分别~ 个别~ 具体~ 统一~ 公开~ 私下~ 秘密~ 正式~ 共同~ 长期~ c. 妥善~



(张寿康、林杏光主编，商务印书馆，2002)

2. 搭配资源研究现状

2.2 基于语料库的搭配资源库

➤ 如中文搭配助手(胡韧奋、肖航, 2019)

➤ 自动化获取搭配, 搭配库规模较大

➤ 具有频率、互信息、例句等信息

➤ 搭配数量非常少(每类最多展示10条)

➤ 存在一定的错误搭配

➤ 展示的词语语法信息较少

➤ 缺乏词语的语义信息

ID	Verb-Object Collocation	Frequency	Mutual Information	Typical Variants
1	安排 时间	17	2.85	安排好时间 安排一下时间 安排出时间 安排了时间 安排得时间
2	安排 日程	10	5.44	安排好了日程 安排满了日程 安排好日程 安排一下日程 安排得日程
3	安排 座位	3	4.19	
4	安排 桥梁	3	5.37	

ID	Subject-Predicate Collocation	Frequency	Mutual Information	Typical Variants
1	日程 安排	7	6.31	日程安排得 日程安排好了

ID	Adverb-Predicate Collocation	Frequency	Mutual Information	Typical Variants
1	都 安排	17	0.50	
2	就 安排	12	-0.19	
3	已经 安排	6	1.29	
4	合理 安排	4	5.36	
5	特意 安排	4	4.57	
6	没有 安排	3	0.32	
7	只好 安排	3	2.45	

ID	Classifier-Noun Collocation	Frequency	Mutual Information	Typical Variants
1	里地	6	4.46	
2	块地	6	2.19	
3	圈地	3	5.75	

ID	Verb-Object Collocation	Frequency	Mutual Information	Typical Variants
1	落了地	4	4.91	
2	买地	3	0.62	
3	去地	3	-3.47	去过地 去了地

ID	Subject-Predicate Collocation	Frequency	Mutual Information	Typical Variants
1	地有	4	0.54	

ID	Adjective-Noun Collocation	Frequency	Mutual Information	Typical Variants
1	满地	6	3.12	
2	背地	3	5.87	

量名搭配是不是处理成数量名搭配更好一些
留学生要去绍兴、杭州等地进行语言实践活动。
防治“网瘾”三地各有招数。

2. 搭配资源研究现状

2.2 基于语料库的搭配资源库

Word Sketch Engine(Kilgarriff et al., 2004年)根据自动提取词语的搭配和语法信息，并将其按照频率和显著性进行统计排列。

Word Sketch Difference（差异素描）：对比两个词的搭配信息并根据语法关系将搭配词划分为不同的类别。

WORD SKETCH DIFFERENCE Chinese Web 2011 (zhTenTen11, Stanford tagger)

吃 433,702x 喝 130,848x

Object	Object_of	SentObject	SentObject_of
饭 11476 0 ...	没事 15 0 ...	消 1058 0 ...	爱 4570 494 ...
肉 5673 0 ...	挑 18 0 ...	透 1058 0 ...	想 6216 1080 ...
水果 3685 0 ...	包 13 0 ...	患上 1425 98 ...	及 870 121 ...
晚饭 3358 8 ...	谈 9 0 ...	饱 1667 156 ...	喜欢 4785 1287 ...
食物 6262 42 ...	确保 45 43 ...	吃 3494 679 ...	饿 592 80 ...
药 4316 256 ...	煮 14 23 ...	完 2756 600 ...	宜 1104 272 ...
茶 1105 8482 ...	吃 40 117 ...	喝 3040 1438 ...	吃 3494 3040 ...
牛奶 262 2478 ...	熬 0 10 ...	下去 1079 663 ...	喝 679 1438 ...
汤 275 3048 ...	泡 0 39 ...	惯 266 107 ...	吃饭 152 308 ...
酒 883 13736 ...	吃饱 0 17 ...	凉 130 174 ...	空腹 82 138 ...
啤酒 10 1408 ...	渴 0 10 ...	聊天 74 226 ...	泡 87 352 ...
咖啡 22 3866 ...	咕咚 0 8 ...	醉 59 1427 ...	渴 15 256 ...

Indirect-Object	Direct-Object	Direct-SentObject	Subject
饭 1522 0 ...	定心丸 35 0 ...	打针 88 0 ...	饭 6128 46 ...
晚饭 354 0 ...	饭 48 0 ...	吐 153 22 ...	晚餐 1034 42 ...
肉 902 0 ...	豆腐 12 0 ...	瘦 265 56 ...	晚饭 966 39 ...
饭时 196 0 ...	肉 32 0 ...	吃 2144 666 ...	中午 996 161 ...
苹果 577 0 ...	苹果 12 0 ...	减肥 1214 505 ...	晚上 2305 588 ...
啥子 427 12 ...	答案 20 0 ...	喝 1066 618 ...	天天

2. 搭配资源研究现状

2.3 现有研究的不足

- 缺乏专用的近义词辨析搭配资源库
 - 只展示词语和搭配词，不能对比近义词的异同
 - 规模太小，信息有限，难以应用于搭配教学和相关研究
- 缺乏可视化展示
 - 可视化的词汇网络有助于学习者直观地观察词语的组合和聚合
 - 只展示词语和搭配词之间的线性搭配，缺乏搭配网络的展示
- 缺乏系统性的描述框架
 - 没有整体的语言知识描写框架，对词语的特征描写不够系统，
 - 只展示词语的搭配词和搭配类型，不标注其他语法信息
 - 没有语义信息

3. 汉语近义词搭配自动抽取

- 西方学界对搭配的定义强调词语共现的统计特征，传统的搭配自动抽取普遍采用词语共现频次、互信息等统计方法（Church, 1990; Smadja, 1993）。
- Word Sketch Engine(Kilgarriff et al., 2004年)利用词性标签构建搭配抽取规则
 - 例如规则“动词 + (冠词 / 数词 / 形容词 / 副词 / 名词) * + 名词”表示当动词和名词之间有冠词、数词、形容词、副词、名词时，可抽取二者作为一条动宾搭配。
- Huang 等(2005) 将这种方法应用到汉语上，发现这种基于词性的正则表达式规则的确可以帮助识别语法关系，但识别准确率、召回率仍然较低，尤其是远距离的组合抽取。
- 要做到高精度的抽取，需要融合统计特征和语言学特征
 - 本文使用Stanford Parser获取词与词的语法关系，融合频次和互信息等统计特征，自动过滤低频、互信息值低的搭配对，经过人工筛选后可以较为方便的获取搭配资源库。

3. 汉语近义词搭配自动抽取

➤ 搭配种类

- 固定搭配：在满足一定的语法关系的基础上，词与词在组合关系上的反复共现（濮建忠，2003），如 < 因为，所以 >
- 自由搭配：满足一定语法关系但联系松散的，如 < 买，书 >

3. 汉语近义词搭配自动抽取

抽取流程

➤ 1、收集含有节点词的句子构建待标注语料库，语料来自于CCL语料库、BCC语料库

1 他不住气地干咳着，东瞅瞅，西瞧瞧。 23 ^ v

2 她捡起草帽，甩了几甩，但她并没把草帽扣在头上，而是挂在东厢房墙壁的一根钉子上。

3 见对面人家屋脊上，受东起的太阳斜照着，抹上一片殷红的阳光。

4 两只长毛大狗，那种又像灵提，又像狼犬的东西，已站在那里，准备等门一开便扑向旅人。

5 若不是两方主子赌得有东道，那么打赢了都领饷，这饷就不晓得打哪儿来了。

6 到了东厢房的门前，她蓦地站住，侧着耳静听，然后，把脸儿轻轻贴在门上，从板缝中向内窥探。

7 卡兰希尔的人民住到越过吉理安河上游的最东边，位在瑞莱山山下海伦佛恩湖的南边地区；他们也曾爬上高耸的林

8 他转身跑到东厢房里，拿出了豁边的青瓷大碗和那幅鸟画，对着我们炫耀。

9 水獭似的脑袋从门外探进来，似笑非笑，一脸害羞的样子，眼睛东瞅西看。

10 就在镇东头那座红大门里。

11 月亮把东楼的阴影缩小，再把西楼的阴影放大，夜夜如此。

12 这一封写得比较简略，主要是说，自从周延儒重新进入内阁，当上首辅之后，颇思振作有为，举措处事，能够顺从

13 受活庄街长有二里，断断续续，脚下是河，背里是山，靠西瞎盲人多的地方叫瞎地儿，靠东聋哑人多的地方叫聋哑

14 可是在他离开这座都市向东去的时候，他还是朝着她住的那个地区望去，希望重温一下跟她一块儿消磨的那些可爱

15 当火车在斯布丁杜佛尔向东转弯，沿着哈莱姆河东岸行驶时，赫斯渥紧张地提醒她，他们已经到了纽约城边。

16 东水撤了职以后，一直没有补选，五个支委就只有四个了。

17 他在乌姆巴拉下车朝东去踏过水田要到老军人住的村庄去。

18 他是从来不到东厢房里来的，但是赛马前头天晚上他却钻到厢房里来了。

19 格雷罗算出，从现在起再过两个半小时，将飞越纽芬兰的东海岸。但是，在这以前，机长可能还要作一次关于飞机

20 如今图尔已经十六岁，精灵们打算离开长久以来所居住的山洞安卓斯，秘密前往位在遥远南方西瑞安河口的海港；

21 是从东偏院来吗？

22 原来近日也是澳眺，来了几个西洋和东洋的茶商，又由几个李大一类的人陪着，在候潮路各个茶行，东转转，西转

23 俩人正说着话，就看见东面老二房的那条街有一片红光冲起，描出一排参差的脊顶；人群里有人大喊说：“东大街走

24 走到北新桥，往东看黑洞洞的城楼一声不发的好象一个活腻了的老看护妇，半打着盹儿看着这群吃多了闹肚子的病

词语	来源	数量
东	BCC语料库	3478条
东方	BCC语料库	2813条
东边	BCC语料库	937条
东面	BCC语料库	364条

3. 汉语近义词搭配自动抽取

抽取流程:

➤ 2、用Stanford Parser对语料库进行句法分析（分析准确率约80%）

id: 1 word: 我们 head id: 10 head: 爱护 deprel: nsubj

id: 2 word: 应该 head id: 5 head: 开始 deprel: aux

id: 3 word: 从 head id: 4 head: 源头 deprel: case

id: 4 word: 源头 head id: 5 head: 开始 deprel: obl

id: 5 word: 开始 head id: 10 head: 爱护 deprel: advcl

id: 6 word: 保护 head id: 5 head: 开始 deprel: xcomp

id: 7 word: 野生 head id: 8 head: 动物 deprel: amod

id: 8 word: 动物 head id: 6 head: 保护 deprel: obj

id: 9 word: ， head id: 10 head: 爱护 deprel: punct

id: 10 word: 爱护 head id: 0 head: root deprel: root

id: 11 word: 万物 head id: 10 head: 爱护 deprel: obj

id: 12 word: 。 head id: 10 head: 爱护 deprel: punct

搭配类型：目前暂时使用stanford parser中的依存关系表示搭配类型，之后人工优化搭配结果时再考虑搭配类型

3. 汉语近义词搭配自动抽取

抽取流程:

- 3、根据搭配类型抽取搭配对、计算搭配频次
- 获得745对搭配
 - 后续需要根据搭配类型、频次、互信息筛选

近义词	搭配词	搭配类型	频次
东方	的	case:dec	338
东边	的	case:dec	221
东方	从	case	111
东	向	case	107
东面	的	case:dec	85
东边	从	case	54
东方	西方	conj	51
东方	向	case	49
东方	在	case	28
东面	从	case	26
东方	之	case:dec	25
东方	和	cc	24
东边	在	case	22
东边	向	case	18
东	从	case	17
东方	个	clf	17
东边	山脉	nmod	15
东方	是	cop	13
东	往	case	12
东面	和	cc	12
东方	遥远	amod	12

3. 汉语近义词搭配自动抽取

抽取流程:

➤ 4、计算搭配互信息值

在	东边	6.0895938
在	东面	5.9819361
在	东方	5.7242043
在	东	5.2820863
头上	东方	5.5152952
而是	东边	5.9092409
而是	东方	5.8694774
挂	东边	7.1184879
挂	东面	6.5206756
挂	东方	5.5152952
挂	东	3.2164979
房	东	6.8350912
房	东边	4.3423123
房	东方	-0.142236
墙壁	东面	7.6841743
墙壁	东边	7.5259123

3. 汉语近义词搭配自动抽取

抽取流程:

➤ 5、根据频率及互信息筛选搭配对，构建高频搭配词矩阵

	东	东边	东方	东面
把	0	3	1	0
从	17	54	111	26
的	2	225	348	85
对	0	0	8	0
对于	0	0	3	0
房子	0	2	0	0
高地	0	2	0	0
更	0	2	0	0
光明	0	0	2	0
和	0	4	24	12
黑暗	0	0	5	0
红	0	0	2	0
花园	0	3	0	0
家	0	3	0	0
就是	0	3	7	0
靠	0	2	0	0
人	0	0	2	0
山	0	3	0	4
山丘	0	0	0	4
山脉	0	15	0	0
天空	0	0	2	0
往	12	3	2	0
我们	0	2	3	3
屋后	0	0	0	2
屋内	0	2	0	0
向	107	18	49	8
由	2	0	0	0
院子	0	6	0	0
在	3	24	32	2
最	2	4	3	0

3. 汉语近义词搭配自动抽取

- 不合理的搭配
 - 切词错误
 - <东边费诺, 了>, <在, 受东方>
 - 搭配错误
 - <东方, 头上>

3. 汉语近义词搭配自动抽取

- 优点

- 快捷方便地获取大量搭配词条

- 缺点

- 准确率并不高，有大量错误搭配

服务于教学的资源库后期需要人工筛选搭配结果，确保呈现正确的搭配

4. 汉语近义词搭配知识库构建研究



子库	属性
1. 基本信息库	
1.1 近义词词库	词语ID、词音、词形、词性、词语HSK等级
1.2 含有近义词的语料库	句子ID、句子
2. 语法特征库	
2.1 搭配词库	搭配词ID、搭配词、搭配词频次、搭配类型、搭配频次、搭配互信息
2.2 其他语法信息	
3. 语义特征库	词义、语义信息
4. 语境特征库	语义韵信息、语体风格

4. 汉语近义词搭配知识库构建研究

近义词的选择

《对外汉语常用词语对比例释》(卢福波, 2000)把汉语作为外语或第二语言的学习者在学习汉语中最常用的、意义或用法易混淆的词语为对比对象和选词范围, 包括对比的词语组例254例, 涉及630多个词。

《1700对近义词语用法对比》选择了汉语学习者在使用中容易出错的1700对同义词或近义词(包括关系密切的词或词语结构), 从语义、语用等方面对其进行了分析对比, 每个词条下面分为词义说明、词语搭配和用法对比三部分。

这两本词典选词丰富, 涵盖了大部分易错近义词, 我们取两本书的并集, 从中筛选了1558组常用近义词作为搭配库的词语; 词典描述了较为丰富的用法信息, 与其对比, 便于评价我们构建的知识库的质量。

4. 汉语近义词搭配知识库构建研究

语料的选择

从大规模语料库中下载含有近义词的所有语料。

4. 汉语近义词搭配知识库构建研究

4.1 基本信息库

词语	搭配词	ID	词性	拼音	HSK等级	词义
安排		56	v+n	an1pai2	4	有条理有次序地处理（事物）、安置（人员）
	工作	128	n+v	gong1zuo4	1	1.从事体力或脑力劳动，也泛指机器、工具受人操纵而发挥生产作用 2.职业 3.业务
	容易	654	a	rong2yi4	3	1.做起来不费事的2.发生某种变化的可能性大

HSK 等级将词语分为不同难度等级，方便检索者自行检索符合自己学习水平的词语

4. 汉语近义词搭配知识库构建研究

4.2 语法特征库

词语	搭配词	ID	搭配关系	搭配频次	搭配互信息
安排		56			
	工作	128	动宾搭配	46	10.67
	容易	654	状中搭配	4	5.46

4. 汉语近义词搭配知识库构建研究

4.3 语义特征库（以形容词为例）

词语	搭配词	语义角色
麦穗儿	金灿灿	1.主事(theme, 以下简称 TH): 形容词所表示的性质、状态等事态的非感知性的主体。
我们	着急	2. 感事(sentient, 以下简称 SE): 形容词所表示的心理感觉、精神体验等事态的感知性的主体。
面积	小	3. 范围(range, 以下简称 RA): 形容词表示的性状所涉及的主体的具体方面(即 维度), 或者形容词表示的性状所描述的部件、属性等所依附的相关主体(即领主)。
庐山	险	4.与事(dative, 以下简称 D): 在形容词所表示的性状方面进行比较的参照者, 一般是 与该形容词的主事、感事、范围等伴随成分相比较的对象。
大	三岁	5.量幅(extent, 以下简称 EXT): 形容词表示的性状等事态所涉及的数量、幅度等事 项, 一般是对该形容词的主事、感事、范围等和与事相比较的结果的某种度量。
婚事	烦恼	6.对象(target, 以下简称 TA): 形容词所表示的心理感觉、精神体验等感知性行为的 对象和目标, 一般用介词“为、为了、对、对于”等引导。
吃着	咸	7. 系事(relative, 以下简称 RE): 主事呈现出形容词表示的某种性状时所处的状态或 所进行的活动, 一般做形容词的状语, 或者做中心语受形容词修饰或补充。
缺水	(而) 枯黄	8.原因(reason, 以下简称 RN): 主事或感事产生形容词所表示的某种性状或心理状 态的原因, 一般用介词“因、因为、由于、为、为了”等引导。
	为了孩子的健康, 她一直很谨慎	9. 目的(aim, 以下简称 AI): 主事表现出形容词所表示的某种性状或心理状态的目 的, 一般用介词“为、为了”等引导。
	瑞士在11月就非常寒冷	10. 时间(time, 以下简称 T): 主事或感事产生形容词所表示的某种性状或心理状 态的时间, 可以用介词“在、于”等引 导, 也可以不用介词。
	在教室里我跟你说话不方便	11. 处所(location, 以下简称 L): 主事或感事产生形容词所表示的某种性状或心理状 态的处所, 可以用介词“在、于”等 引导, 也可以不用介词。
左 右	偏 斜	12. 方向(direction, 以下简称 DI): 主事产生形容词所表示的某种性状所朝的方向, 可 以用介词“向、朝”等引导, 也可 以不用介词。

4. 汉语近义词搭配知识库构建研究

4.4 语境特征库

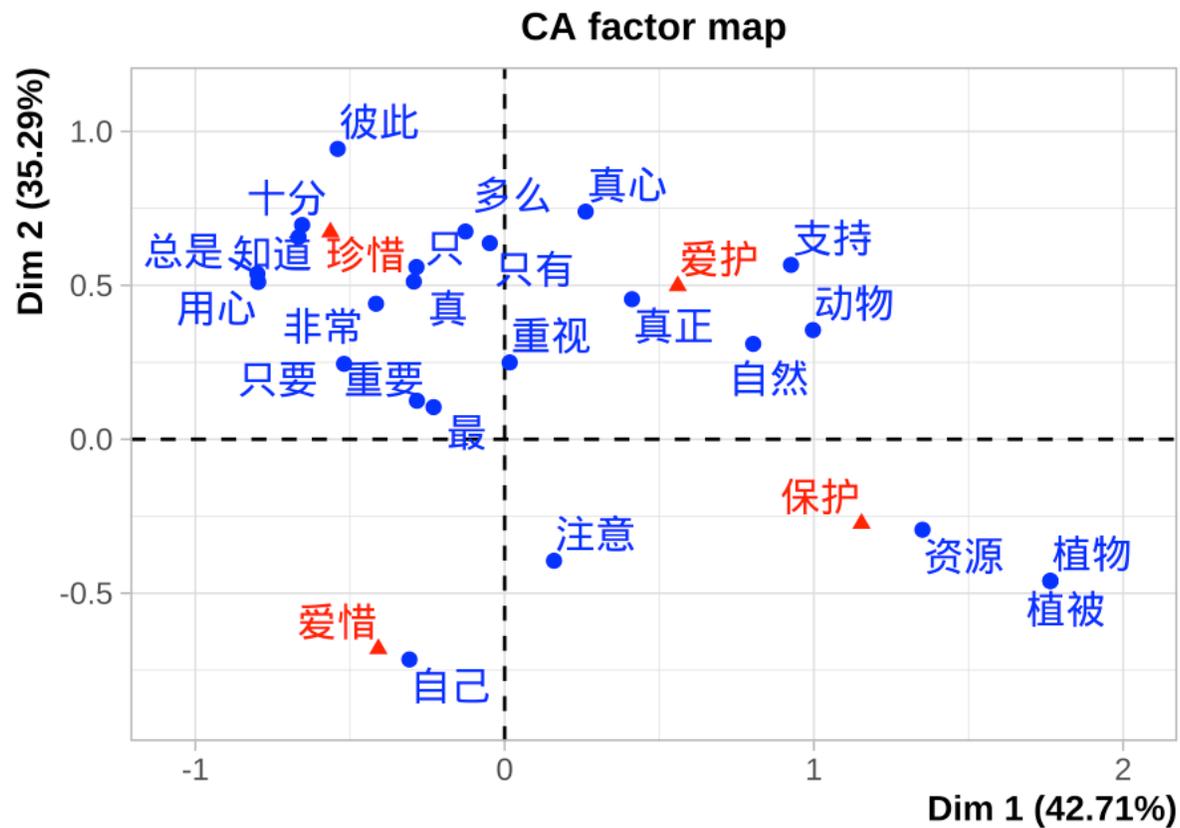
词语	搭配词	ID	语义韵	感情色彩	语体风格
安排		56			
	工作	128	中性	中性+贬	书+口
	容易	654	积极	中性	书+口

语义韵：词项的搭配行为显示一定的语义趋向：一定的词项会习惯性地吸引某一类具有相同语义特点的词项，构成搭配。由于这些具有相同语义特点的词项与关键词项在文本中高频共现，后者就被“传染”上了有关的语义特点，整个语境内就弥漫了某种语义氛围。（卫乃兴）

语义韵大体可分为积极、中性和消极三类。

5. 汉语近义词搭配结果可视化

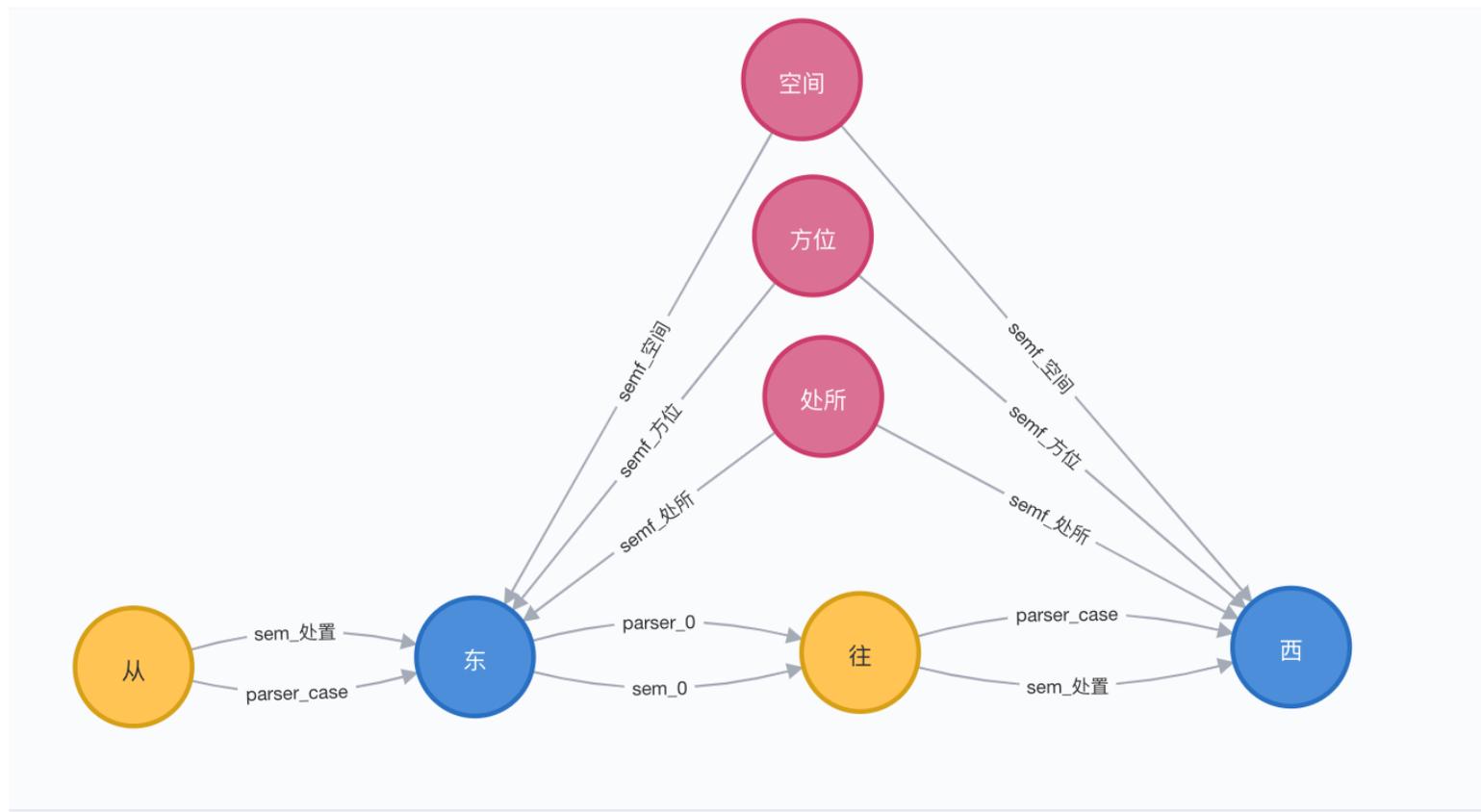
搭配偏好图



R语言包: CA
相关分析, 将高维搭配数据转为二维

5. 汉语近义词搭配结果可视化

句法语义特征可视化图



以结构为单位，描写近义词的句法语义信息。

用图数据库Neo4j储存数据，既便于学习者用可视化方式查看数据，又可作机读知识库。
Displaying 7 nodes, 12 relationships.

5. 汉语近义词搭配结果可视化

近义词例句推荐

搭配可视化只能呈现近义词的搭配偏好，不能为词典编纂者或语言教学提供现成的解决方案来分析词语的语义区别，我们可以给关键词加上超链接，将可视化图表上的关键词与含有关键词的例句链接起来，用户可以点击关键词获取语料库中的句子实例，结合语料观察节点词的语境，对词语进行深入的分析，从而自行归纳词语的使用模式。

6. 总结

- 本文尝试基于依存信息、频次信息、互信息值构建一个半自动化的词语搭配资源库。
- 本文的贡献如下
 - 1、设计了专用的近义词搭配资源库构建方案
 - 2、面向应用，提供大规模的搭配词，构建了词语的基本信息库、语法特征库、语义特征库、语境特征库
 - 3、具有可视化展示效果

参考文献

略



**敬请大家提出意见，
谢谢！**