

外向型汉语词典词语义项概率分布研究

宫新宇, 刘嘉雯, 徐睿姿, 徐艳华

鲁东大学文学院, 山东 烟台

收稿日期: 2024年4月22日; 录用日期: 2024年6月17日; 发布日期: 2024年6月27日

摘要

内向型词典义项的编排是根据词义出现时间的先后安排顺序的, 并没有考虑义项使用的频率, 有些外向型汉语词典义项的排列基本上参考了内向型词典, 这不利于二语学习者学习汉语词汇。因此, 本文依据BCC语料库, 对《国际中文教育中文水平等级标准》中初等和中等词汇的语义概率进行了统计, 在此基础上考察《汉语教与学词典》词义排列顺序, 对排序不合理的词语提出排序建议, 使词典达到“易查、易懂、易学”的目标, 以期为国际中文教育和汉语词汇教学提供参考。

关键词

外向型汉语词典, 语义项, 概率分布, 国际中文教育

Research on the Probability Distribution of Word Semantic Terms in Export-Oriented Chinese Dictionaries

Xinyu Gong, Jiawen Liu, Ruizi Xu, Yanhua Xu

School of Chinese, Ludong University, Yantai Shandong

Received: Apr. 22nd, 2024; accepted: Jun. 17th, 2024; published: Jun. 27th, 2024

Abstract

The arrangement of the semantic items in the introverted dictionary is arranged according to the order of the time when the meaning appears, and the frequency of the use of the semantic items is not considered, and the arrangement of the semantic items in some extroverted Chinese dictionaries basically refers to the introverted dictionary, which is not conducive to the learning of Chinese vocabulary by second language learners. Therefore, based on the BCC corpus, this paper counts the semantic probabilities of elementary and intermediate words in the International Chi-

nese Education Chinese Proficiency Grading Standards, and on this basis, the semantic order of the words in the Chinese Teaching and Learning Dictionary is investigated, and the sorting suggestions are put forward for the unreasonable words, so that the dictionary can achieve the goal of “easy to check, easy to understand, and easy to learn”, in order to provide reference for international Chinese education and Chinese vocabulary teaching.

Keywords

Export-Oriented Chinese Dictionary, Semantic Terms, Probability Distributions, International Chinese Education

Copyright © 2024 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

随着中国综合国力的提高和世界交流的拓展,汉语越来越受欢迎,已成为“应用力”排名仅次于英语的世界第二主要语种[1],在世界范围内得到普遍推广。目前外向型汉语词典义项的排列主要参考内向型词典,而内向型词典义项排序不是依据频率而是出现先后,这不利于二语学习者的词汇学习。如陆俭明教授(2006)所说的,外向型汉语学习词典基本上都是《新华字典》和《现代汉语词典》的删减本,没有真正从外国学生学习汉语的角度来考虑编写,包括收字、收词、释义、举例等。究其原因,我国外向型汉语词典编纂者受本民族理解思维的影响较大,且并未冲破内向型词典的传统编纂模式,词语义项的选择和排列虽按一定原则,但对于能否满足外国汉语学习者的需求,仍存在着一定的客观差距,使得外国汉语学习者准确理解和掌握词语的基本义仍存在困难。而对外向型汉语词典语义项概率分布进行专门研究的学者则更少,大部分学者对于语义项的研究也主要是集中在某一类词之上,没有对词汇语义项进行整体性研究,如刘钰静的硕士论文《两部近义词典共有单音节动词义项比较研究》仅研究单音节动词义项;李娟的《基于认知的外向型汉英词典多义词义项处理模式》一文也是仅局限于多义词义项研究;王安民、王健在《从认知隐喻学看外向型汉英学习词典中单字义项和复词条目的编排》一文仅局限于某一类型的词汇义项研究。

鉴于此,本文以北京语言大学语料库中心(以下简称 BCC 语料库)为基础,以《国际中文教育中文水平等级标准》为参考,对《汉语教与学词典》中收录的初等词汇 2245 个和中等词汇 3211 个的义项排列顺序进行考察,并根据义项分布概率对词语义项排序进行验证,对排序不合理的词语提出排序建议,使外向型汉语词典真正能够满足外国汉语学习者的需求,从而推进国际汉语教育和汉语国际化进程。

2. 初等词汇义项概率分布

词汇是语言三要素之一,也是语言的建筑材料,是学习一种语言必须要掌握的内容。对于将汉语作为第二语言的学习者来说,词汇是提高汉语理解和表达能力的关键,《国际中文教育中文水平等级标准》(以下简称《标准》)中初等词汇是汉语中最常用的词汇。

《标准》中初等词汇一共包括三级,总共 2245 个词,其中,一级词汇 500 个,二级词汇 772 个,三级词汇 973 个。为了能够更深入地了解初等词汇义项的概率分布,我们从 BCC 语料库中为每个初等词随机抽取 500 个例句,采用人工分析的方法对这些词的义项进行辨析,统计常用词汇义项的使用频次,并

与外向型汉语词典——《汉语教与学》词典中所收录的该词汇的义项排列顺序进行对比。根据统计结果，发现《汉语教与学》词典中所收录的词汇语义项排序与语料库中该词汇常用语义项排序存在较大差异，具体情况如下表 1 所示：

Table 1. Ranking statistics of elementary vocabulary semantic items
表 1. 初等词汇义项排序统计

初等词汇	总数量	问题数量	百分比
一级词汇	500	182	36.4%
二级词汇	772	181	23.4%
三级词汇	973	285	29.2%

初等词汇作为在日常生活中最常用的一类词语，义项排列顺序的恰当与否直接影响着将汉语作为第二语言学习者对于相同词汇在不同语境中的理解。而对于初等词汇义项概率分布的研究能够帮助第二语言学习者更好的理解词语在不同语境下所表达意义的不同。因此，笔者将一、二、三级词汇中义项排列顺序存在问题的词汇按照词性进行了以下汇总。

在一级词汇当中，共有 182 个词汇义项排列顺序存在问题，其中名词如班、包、北边、病、菜、茶、车、床、地方、弟弟、电话、东西、国、孩子、后、花等 59 个；动词如爱、帮、出、穿、到、等、动、工作、关、过、回、会、叫、进等 46 个；形容词如错、大、差、形、坏、老等 23 个；副词如别、不用、不、常、就、没有等 18 个；助词如吧、的、着等 5 个；连词如还是、和两个；介词如比、从、给、连等 5 个；量词如本、点、分、号、家、路、毛等 18 个；数词如半、二、几、两、零、三等 8 个；代词如多、几、哪、什么、那等 14 个。

在二级词汇当中，共有 181 个词汇义项排列顺序存在问题，其中名词如边、表、菜单、草地、成绩、出口、单位、蛋、道理、店、东北、东方、海、好处等 70 个；动词如变、不行、查、成、出发、吹、答应、带、分、过来、换、活动、讲话、接受、经过、算等 64 个；形容词如低、红、满、轻、清楚、全、热情、实际、舒服、熟等 26 个；副词如才、重新、多么、好、好像、老、实在等 19 个；助词如得 1 个；连词如不过、同样、同时 3 个；介词如对、让、通过、往、向 5 个；量词如遍、层、道、度、份、刚、够、句、刻等 21 个；数词如千 1 个；代词如那么、这么、这样 3 个。

在三级词汇当中，共有 285 个词汇义项排列顺序存在问题，其中名词如报告、背、比例、城、村、错误、代表、底下、地区、电台、动力、队员、发明、工具等 111 个；动词如把握、保、保证、变化、表现、表演、创作、从事、谓、订、定期、断、发表、防、访问、分配、付、挂等 129 个；形容词如不安、吵、富、根本、公开、光、好好、活、静等 60 个；副词如白、本来、比较、重、初、当然、到底、反复、反正、分别、较等 38 个；连词如并、不光、不仅、任等 8 个；介词如按、把、朝、叫、为等 8 个；量词如部、步、顿、架、类、期、团等 15 个；代词如另外、某、如何 3 个。

从以上内容所呈现的结果来看，义项排列顺序存在问题的词汇主要集中在名词、动词、形容词三类，而这三类词汇也恰恰在汉语学习中占据着核心地位，是语言交流中必不可少的存在。因此，笔者认为对于初等词汇义项排列顺序的重新定位，尤其是对初等词汇中名词、动词、形容词义项排列顺序的再统计，有利于外向型汉语词典的编纂者在词典编撰中对词汇的义项进行更加合理的排序，也能够帮助第二语言的学习者更好的理解在不同语境名词、动词、形容词所呈现出的不同含义，体会中国文字的魅力。

3. 中等词汇义项概率分布

《标准》中的初等词汇、中等词汇和高等词汇的划分主要是从词汇的常用性、覆盖率以及学习者语

言能力的角度进行的划分，便于教师更好的指导学生的学习内容和难度。中等词汇作为连接初等词汇和高等词汇之间的桥梁，在推动第二语言学习者更加熟练的掌握汉语上有着不可估量的作用，虽然中等没有初等词汇常见，但也在对话、交流与合作当中占据一席之地，是不能够缺少的存在，在教育部语言文字研究所发布的《现代汉语语料库词语分类频率表》中，中等词汇在其语料库中出现的频率基本上大于50次以上[2]，因此，对于中等词汇义项概率的统计能够帮助第二语言学习者完善其汉语学习的构架。

《标准》中，中等词汇一共 3211 个，其中四级词汇 1000 个、五级词汇 1071 个、六级词汇 1140 个。在对中等词汇的义项的概率统计中笔者采用了与初等词汇相同的统计方法，都是参照 BBC 语料库进行的人工分析，具体情况如下表 2 所示：

Table 2. Ranking statistics of medium vocabulary semantic items
表 2. 中等词汇义项排序统计

中等词汇	总数量	问题数量	百分比
四级词汇	1000	282	28.2%
五级词汇	1071	108	10.1%
六级词汇	1140	67	5.8%

中等词汇是第二外语学习者从对汉语的认识到的熟练必经的一个学习阶段，在外语学习过程中占据着重要地位，因此中等词汇义项排列顺序的恰当与否间接的影响着第二外语学习者对于词汇含义的影响。按照对初等词汇的统计方法，笔者将四、五、六级词汇中义项排列顺序存在问题的词汇按照词性进行了以下汇总。

在四级词汇当中，共有 282 个词汇义项排列顺序存在问题，其中名词如包裹、宝、宝贝、背景、标志、冰、成人、刺激、大会、大陆、待遇、导游、肚子、法、方案等 87 个；动词如安、安置、摆、保守、抱、避、编、别、冲、抽、定、堵、锻炼、对比、对付、发挥、翻、反映等 105 个；形容词如暗、沉、充分、次、粗、淡、恶心、烦、放、肥、分散、概括、个别等 55 个；副词如反、非、纷纷、共、怪、几乎等 17 个；量词如顶、户、伙等 7 个；介词如对于等 3 个；代词如此等 3 个；连词如不然、而、既等 5 个。

在五级词汇当中，共有 108 个词汇义项排列顺序存在问题，其中名词如包装、比方、层次、产业、肠、场面、处分、存款、答复、贷款、鼓、鬼等 34 个；动词如安慰、拨、编辑、变动、补贴、尝、裁判、叉、插、称、打击、大于、扶、规划、滚、恨、恢复等 44 个；形容词如不得了、丑、脆、呆、毒、高度、光荣、慌等 20 个；副词如差点儿、大致、干脆等 6 个；量词如册、堆等 3 个；介词如打 1 个。

在六级词汇当中，共有 67 个词汇义项排列顺序存在问题，其中名词如电器、队伍、高层、高峰、贡献、轨道、幻想等 17 个；动词如跌、斗争、渡、端、夺、夺取、服、负、攻击、昏等 24 个；形容词如巩固、光辉、横等 10 个；副词如仿佛、绝等 4 个；量词如餐、滴、股等 8 个；介词如除、顺 2 个；代词如本、诸位 2 个。

从以上统计所呈现的结果看，中等词汇义项排列顺序存在问题的词汇与初等词汇相同主要集中在名词、动词、形容词三类上，而量词、数词、助词等词性存在问题的概率较低，这也要求研究者在对词汇进行的过程中需要更加注重名词、动词、形容词的义项排序变化。

4. 初等词汇与中等词汇义项排序分析

《大纲》中初等词汇和中等词汇是第二外语学习者想要熟练地掌握汉语必须要学习的内容，而明确初等词汇与中等词汇之间义项排列顺序的变化，能够帮助第二外语学习者更好的突破语言学习中的壁垒，

体会汉语的魅力。

整体上,从表 1 和表 2 对于初等词汇和中等词汇义项排列顺序的问题统计中,可以看出中等词汇义项排列顺序存在问题的概率,相较于初等词汇语、义项顺序存在问题的概率要低,笔者推测这与中等词汇其适用范围相较于初等词汇来说有所缩减相关,初等词汇多应用于日常生活,在不同的语境中会呈现出不同的含义,中等词汇多数是用于书面表达或者是正式场合下的交流,应用的语境相对固定,义项排列顺序存在问题的概率也相对于初等词汇有所下降。

词性上,在语言运用当中,最常用的三种词性是名词、动词和形容词,这也是第二汉语学习者想要学好汉语必须要掌握的三类词性,笔者也将以这三类词性进行对比分析。初等词汇中一级词汇义项排列顺序存在问题的 182 个,其中名词存在问题有 59 个,动词有 46 个,形容词有 18 个;二级词汇中义项排列存在问题的一共 181 个,其中名词 70 个,动词 64 个,形容词 19 个;三级词汇中义项排列存在问题的一共 285 个,其中名词存在问题的 111 个,动词 129 个,形容词 60 个。中等词汇中四级词汇义项排列顺序存在问题一共有 282 个,其中名词存在问题的有 87 个,动词有 105 个,形容词有 55 个;五级词汇中义项排列顺序存在问题的一共有 108 个,其中名词有 34 个,动词有 44 个,形容词有 20 个;六级词汇中义项排列存在问题的一共有 67 个,其中名词有 17 个,动词有 24 个,形容词有 10 个。从总体趋势上来看初等词汇当中义项排列顺序存在最多的名词,笔者认为这主要和名词的性质相关,名词主要表示人、事物、时间、处所、方位的名称。意义一般是比较固定,但是在不同的语境中,名词引申义运用和使用的频率可能比本义要高,尤其是日常生活中。而中等词汇中三类词性,义项排列顺序存在问题最多是动词,笔者认为出现词中现象与中等词汇的难度的增加有一定的关联性。首先,动词通常具有多个含义和用法;其次,动词的义项往往涉及到动作、行为等抽象概念,在理解和界定这些抽象概念时,排序便不可避免的存在一定的主观性;所以涉及到中等词汇中对动词不同含义的界定,动词义项的排序就更容易出现问题。最后,随着语言发展和社会变迁的影响,很多词汇义项的使用频率也出现了一定的变化。

因此,很多词汇义项的排列顺序出现了问题,笔者在此建议词典的编撰中,重视词典编纂中频率原则,结合历时顺序、逻辑顺序原则,可以将高频义项排在前面,以便第二语言学习者更加便捷的进行查找,推进汉语走向国际化。

5. 总结

北京大学的宋立文在其论文《外向型汉语学习词典的出版概况、编写问题与未来之路》谈到:“释义是词典编撰最核心的部分,释义好坏是判断一部词典编写质量的第一标准。从用户视角来看,一部词典‘能不能用’‘好不好用’,其实主要也是对释义的判断。”[3]释义是基础,而对语义项顺序的正确选择则是做好词典释义编撰的前提,能够有效地缓解由“内向型”汉语词典向“外向型”汉语词典之间过渡的壁垒。

随着科技的进步,新媒体时代的到来,外向型汉语词典融媒化成为必然趋势,这也对词典的编纂者提出了更高的要求。但是目前已出版的外向型汉语词典对于第二语言学习者来说释义和配例用词过难,没有呈现词语语义项使用的特点和常用程度[4],不利于第二语言学习者理解词汇的意义,而造成此种状况的原因蔡永强指出:“目前已出版的对外汉语学习词典之所以被冷落,恐怕在很大程度上是由查找困难造成的。”[5]在外向型汉语词典与新媒体结合的状态下,对于词汇语义项的正确排序也越发显得重要,不仅能够帮助第二语言学习者认识到词汇义项中的重要程度,更能够推动国际中文教育的发展。

基金项目

省级大学生创新训练项目“外向型汉语词典词语义项概率分布研究”。

参考文献

- [1] 刘善涛. 系统研究对外汉语学习词典的结构特征[EB/OL]. 中国社会科学报, 2014.
<https://news.nankai.edu.cn/mtnk/system/2014/02/17/000166296.shtml>, 2024-03-05.
- [2] 路珺尧. 国际中文教育中等词汇研究——以《国际中文教育中文水平等级标准》为依据[D]: [硕士学位论文]. 哈尔滨: 黑龙江大学, 2022.
- [3] 宋立文. 外向型汉语学习词典的出版概况、编写问题与未来之路[J]. 北华大学学报(社会科学版), 2021, 22(5): 35-48+152.
- [4] 杨玉玲. 融媒时代外向型汉语学习词典编纂理念与实践[J]. 首都师范大学学报(社会科学版), 2022(2): 167-177.
- [5] 蔡永强. 从释义用词看中高级汉语学习词典的释义[J]. 国际汉语教学研究, 2021(1): 88-96.