

◎ 龚 兵 著

中国习得者

二语阅读词汇能力 发展研究

THE DEVELOPMENT OF CHINESE LEARNER'S
ENGLISH READING VOCABULARY:
AN EMPIRICAL STUDY



西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS

◎ 龚 兵 著

中国习得者

二语阅读词汇能力 发展研究

THE DEVELOPMENT OF CHINESE LEARNER'S
ENGLISH READING VOCABULARY:
AN EMPIRICAL STUDY



西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS

图书在版编目(CIP)数据

中国习得者二语阅读词汇能力发展研究/龚兵著.
—西安:西安交通大学出版社,2013.6
ISBN 978-7-5605-5330-6

I. ①中… II. ①龚… III. ①英语—词汇—研究
IV. ①H313

中国版本图书馆 CIP 数据核字(2013) 第 123119 号

书 名 中国习得者二语阅读词汇能力发展研究
著 者 龚 兵
责任编辑 董云梅

出版发行 西安交通大学出版社
(西安市兴庆南路 10 号 邮政编码 710049)

网 址 <http://www.xjtupress.com>
电 话 (029)82668357 82667874(发行中心)
(029)82668315 82669096(总编办)

传 真 (029)82668280
印 刷 陕西时代支点印务有限公司

开 本 880mm×1 230mm 1/32 印张 10.75 字数 341 千字
版次印次 2013 年 6 月第 1 版 2013 年 6 月第 1 次印刷
书 号 ISBN 978-7-5605-5330-6/H·1505
定 价 26.00 元

读者购书、书店添货、如发现印装质量问题,请与本社发行中心联系、调换。

订购热线:(029)82665248 (029)82665249

投稿热线:(029)82665370

读者信箱:cf_english@126.com

版权所有 侵权必究

foreword 前言

中国习得者
二语阅读词汇能力
发展研究

本项研究基本可以分为四部分。第一部分从词汇知识入手,详细介绍了各类词汇知识框架、词汇能力、词汇能力模型等相关概念和理论以及词汇知识和词汇能力的测试工具等。第二部分着重介绍二语词汇习得中的相关理论假设,包括注意假设、输入假设、词频效应假设、语境效应假设、累积习得假设、伴随性词汇习得假设以及信息处理理论等。第三部分为实证研究,主要包括词汇知识与阅读能力的相关性研究、词汇知识的自动性教学实验、注意假设与频率效应的实证研究、直接词汇学习与间接词汇学习的对比实验研究、阅读词汇水平与英语隐喻理解能力的相关性研究。第四部分为公共英语教学中的阅读词汇学习问题与对策,主要探讨公共英语教材的词汇量与阅读的关系、公共英语教学大纲词汇表的研制等问题。

词汇知识广度和深度调查所探讨的问题主要包括高职院校学生的词汇知识广度和深度、词汇知识与阅读能力的关系,词汇知识广度和深度的发展关系。直接词汇教学实验主要探讨新学得的词汇知识在词、句、篇三种理解水平上的自动化程度。注意假设研究主要通过阅读附带词汇习得探讨纯阅读活动中,词义、词形和搭配附带习得的顺序问题、边缘学习和聚焦学习对阅读附带词汇习得的影响。频率效应研究主要回答一次性阅读是否对词义、词形和搭配知识具有相同影响、不同接触频率是否对词义、词形和搭配知识的学习产生不同影响以及出现多次的低频词和出现一次的高频词被习得的概率等问题。直接词汇学习与间接词汇学习对比实验主要研究加注释和查字典两种辅助手段对词汇学习的影响、两种词汇学习的优势对比。隐喻理解能力研究主要探讨隐喻的判断和解释能力与词汇水平的关系和中国英语学习者的隐喻能力发展水平。

本研究中的受试分四类:一类是高职四年制大三学生,共计 237 人参

加了本研究中的词汇知识广度、深度和阅读能力调查。其中 55 位参加了词汇知识的自动性教学实验研究;第二类是深圳某高职院校三个自然班三年制新生,共 70 位学生参加了直接词汇学习与间接词汇学习的对比实验研究。第三类是深圳某大学英语专业一年级学生,共 50 名学生参加了二语阅读中的聚焦和边缘词汇学习实证研究和阅读附带词汇习得中的频率效应研究。第四类是某大学外语教师和深圳某大学英语专业三年级学生,共 40 名教师和 60 名学生参加了二语阅读水平和隐喻识解能力调查。

本研究中的词汇知识广度测试工具有两种,一是直接采用内森(Nation, 2001)的词汇水平测试,二是根据研究需要自行设计的题型。词汇知识深度测试主要采用帕里巴克特和威士奇(Paribakht & Wesche, 1993)的词汇知识五级量表和里德(Read, 1998)的词汇联想测试工具。阅读词汇知识教学实验以及直接词汇与间接词汇学习对比实验中的工具基本上为笔者自己设计(详见附件)。注意假设和频率效应研究中的阅读材料选自《圣经》(创世记)。边缘学习与聚焦学习方式问卷调查工具部分采用内森的教学设计,其他为笔者自行设计,主要包括多项选择题、词干补笔等。隐喻能力测试工具中的前两套源自莉特莫尔(Littlemore, 2001)的发现隐喻意义能力测试,后两套为笔者自行设计,材料取自英文版《圣经》、《简爱》、《绿野仙踪》、《心灵鸡汤》、《组织行为学》和《希腊罗马神话》。文本分析工具为词频概貌软件,数据处理工具为 SPSS。

通过数据分析,本研究得出以下主要结论:

一、词汇知识广度和深度与阅读能力成正相关关系。线性回归分析显示,词汇广度和深度均对阅读能力具有预测力,词汇知识越丰富,阅读能力越强。词汇量较高的学习者的词汇知识广度与深度在词义、搭配关系的识别能力和在语篇水平上的认知能力方面发展几乎是同步的,表明该组学生在词汇学习过程中比较重视词汇深度的习得,词汇量低的学习者的广度和深度之间不存在相关性,表明其深度并没有随着词汇量的扩大而加深,二者的发展明显不在同一水平上。

二、直接词汇学习可以有效地促进识别词汇和识解词汇知识的快速增长。在完成识解任务时,语境越复杂、任务难度越大,识别词汇转化为识解词汇能力的可能性就越小。在翻译句子和篇章时,词汇知识只有 25% 和 18% 的表现,说明迅速增长的词汇知识绝大部分仍然停留在识别词汇水平上,只有少部分达到了识解水平。在阅读词汇知识发展的过程

中,即使对完整的词汇知识进行了集中的系统接触和处理,如果没有后续的语境化自动加工处理,词汇知识仍然会停留在辨认的水平上,达不到识解水平。

三、在没有具体的词汇学习任务时,基于理解的阅读任务可以不同程度地促进三种词汇知识的习得,词义优先,其次是搭配或词形知识。边缘学习无助于阅读附带词汇习得,只有当学习者采取聚焦学习方式去处理目标词时,才有词汇附带习得的可能。

四、一次阅读可以产生部分词汇知识的学习,主要是词义知识。其他维度的知识保持率较低;多次接触对词义习得影响显著,对词形和搭配的习得没有影响;本研究中接触频次为1的目标词习得率不如其他频次的目标词,一次习得现象不过是一种偶然现象,不足以推翻频率效应。本研究中的低频词的出现频率和习得率都高于高频词,充分说明频率效应对阅读附带词汇习得具有积极影响。

五、对低水平学习者来说,直接词汇学习比间接词汇学习更能促进词汇习得,但没有学习者课外的努力,有限的课时无法强化直接词汇学习效果;生词表和词典并不能自动地促进低水平学习者间接习得词汇,需要明确、合理、适量的学习任务进行引导。

六、二语词汇水平与二语隐喻能力紧密相关;发现隐喻、隐喻解释与语境中隐喻的识别和理解能力成正相关;中国二语学习者隐喻发展水平明显偏低,在单句水平上,发现隐喻意义和隐喻解释能力要比语境中的隐喻识别和理解能力强。研究结果充分说明,中国英语专业习得者,即使词汇能力能力达到了较高水平,但二语隐喻能力发展水平明显滞后。

以上研究结果对我国的二语习得研究以及公共外语教学实践具有一定的启示意义。第一,两项词汇知识调查和直接词汇学习教学实验结果均说明,词汇知识与词汇能力是两个完全不同的概念,公共外语的教和学在提供词汇知识输入时,应同时强调新学得词汇知识在真实语境中的内化。直接词汇学习对于未达到二语阈限水平的外语学习者来说是一种有效的途径,其作用不可忽视。如果缺乏充足的语言运用环境,学生掌握的词汇知识无论多么丰富都只能停留在知识的层面上,无法转化为理解和表达能力。第二,注意假设虽然意识到语言的多维性需要不同的注意水平,但既没有说明何种语言知识需要何种程度的注意,也没有说清注意与学习行为之间的关系。张文忠和吴旭东的聚焦学习和边缘学习概念为

阅读附带词汇习得研究提供了新的视角,但要根据具体的研究进一步确定其操作定义,这样更助于阅读附带词汇习得产生机制的实证研究和理论探讨。对于二语习得来说,基于阅读的附带词汇习得并不会完全是无意识的、隐性的,在某种程度上需要意识的参与。因此,泛读课上如果能适量增加一些显性的学习任务会更有效地促进伴随性词汇习得。第三,本次频率效应研究显示,一次性的阅读活动只是频率对语言学习的累积效应的开始,而不是结束。多维的词汇知识习得应该靠长期的、大量的阅读活动才能产生。高频词具有优先习得的优势,这对于我国目前的大学英语教学改革具有一定的启示意义。第四,词汇量水平影响词汇学习效果。低水平学习者缺乏基本的单词拼读能力和构词知识,词汇学习完全靠死记硬背。因此,公共英语教学应加强低水平学习者的语音、构词知识的辅导,培养低水平学习者最基本的词汇学习能力。第五,语言和文化因素可以提升隐喻理解能力,但仅靠传统的语言文化教学显然不足于有效提高英语隐喻能力。

一言以蔽之,以上结论对我国公共外语教学具有一定的借鉴价值。非英语专业习得者的词汇知识远远还没有达到满足自由阅读的保障水平,专业英语习得者包括博士水平的学习者,其隐喻理解水平明显发展滞后,这充分说明,中国二语学习者的概念流利水平依然还存在较大的发展空间。在 EGP 能力发展不足的情况下,大力推进 ESP 或 EAP 教学改革,是否缺乏充分的依据?这值得国内公共外语教学同仁进一步思考。

contents 目录

中国习得者
二语阅读词汇能力
发展研究

第一章 词汇知识

- 一 | 词汇的概念/ 002
- 二 | 词频统计/ 005
- 三 | 产/接词汇与主动/被动词汇/ 009
- 四 | 词汇知识框架/ 017

第二章 词汇能力模型与心理词汇表征模型/ 022

- 一 | 词汇能力模型/ 022
- 二 | 心理词汇表征模型/ 028

第三章 词汇知识与词汇能力评估/ 036

- 一 | 词汇量/词汇宽度评估/ 036
- 二 | 词汇知识深度测试/ 040
- 三 | 词汇能力测试/ 045
- 四 | 隐喻能力测试/ 048

第四章 注意假设与输入假说/ 052

- 一 | 注意假设/ 052
- 二 | 克拉申的输入假说/ 056

第五章 词频效应假设/ 062

- 一 频率及频率效应的定义/ 062
- 二 累积习得假设/ 062
- 三 支持频率效应说的观点/ 063
- 四 反对频率效应说的观点/ 066
- 五 折中派的观点/ 069
- 六 频率效应的局限性/ 070
- 七 结束语/ 071

第六章 信息处理理论/ 074

- 一 工作记忆和长期记忆/ 074
- 二 自上而下与自下而上的加工/ 077
- 三 自动加工和受控加工/ 080

第七章 语境效应与二语词汇习得/ 085

- 一 语境效应/ 086
- 二 二语词汇习得中的语境效应之争/ 094
- 三 影响语境猜词的知识类别/ 101
- 四 语境效应研究中应注意的问题/ 106
- 五 高职英语学习者对生词的处理策略/ 107
- 六 阅读材料中的真实性问题/ 109
- 七 国外语境中的词汇学习实证研究/ 112
- 八 基于阅读的词汇教学策略/ 118
- 九 对外语教学的启示/ 125

第八章 阅读词汇知识阈限假设/ 128

- 一 阅读词汇知识阈限假设/ 128

- 二 | 语料库中的词汇组织特征/ 130
- 三 | 二语学习者与本族语者的心理词汇差异/ 131
- 四 | 阅读理解的认知过程/ 132
- 五 | 结语/ 134

第九章 二语词汇知识与阅读能力的相关性实证研究/ 135

- 一 | 英语词汇知识广度与阅读能力的相关性/ 135
- 二 | 词汇知识深度习得对阅读能力的影响/ 142
- 三 | 词汇知识、词汇能力与阅读能力的关系/ 150

第十章 注意假设与频率效应实证研究/ 159

- 一 | 二语阅读中的聚焦和边缘词汇学习实证研究/ 159
- 二 | 阅读附带词汇习得中的频率效应/ 169

第十一章 间接词汇学习与直接词汇学习的对比实验研究/ 180

- 一 | 相关文献/ 180
- 二 | 实证研究/ 183
- 三 | 数据分析/ 185
- 四 | 结果与讨论/ 188
- 五 | 结语/ 191

第十二章 二语阅读词汇水平与隐喻识解能力相关性研究/ 192

- 一 | 相关研究/ 192
- 二 | 术语界定/ 193
- 三 | 研究问题/ 194
- 四 | 研究方法/ 195

五 | 结果与讨论/ 198

六 | 结论与启示/ 202

第十三章 公共英语教学中的阅读词汇学习问题与对策/ 203

一 | 公共英语教材的词汇量与阅读难度之关系/ 203

二 | 公共英语教学大纲词汇表的研制思路/ 211

三 | 研究问题/ 217

附录/ 221

一 | 附录 1: 直接词汇与间接词汇学习对比研究附件/ 221

二 | 附录 2: 阅读词汇知识发展的实证研究附件/ 249

三 | 附录 3/ 287

参考文献/ 295

术语对照表/ 327

第一章

词汇知识

词汇在二语习得中起何作用?作用有多大?如果在上个世纪五六十年代,词汇习得研究刚刚起步的阶段。恐怕没人回答得上来。中国人对英语语法的认知程度已经深到本族语者为之汗颜的地步。可你要问及词法,他们一定会说:“只听说语法,从没听说还有词法。”词法在中国甚至在全世界的外语教学中基本上登不了大堂,一直不为人们所重视。潘文国教授(2008:2-3)在给陈万汇博士的专著《中国学习者二语词汇习得研究》的序言中对其背景做了精辟独到的分析:“第一,汉语研究起源于母语教学,西方语言理论起源于外语教学;第二,西方传统外语教学又基本上是在同一语系诸语言之间进行的。”潘教授特别指出,由于同一语系具有相同的基督教传统和希腊拉丁传统,英、法、德等欧洲语言具有同质性,同源词很多。因此,只要具有了一定的语法规则和读音规则,学习第二种语言就能驾轻就熟。随着认知语言学和语料库语言学的兴起,人们发现,以语法为中心的外语教学并没有像传统语言学家所推定的那样一劳永逸。因此,随着二语词汇习得研究的深入,词汇在语言习得中的作用越来越受到人们的关注。威尔金斯(Wilkins,1972:111)对于词汇在语言习得中的重要性有句精辟的论断:“没有语法,很多东西难以表达;没有词汇则什么也表达不了。”近半个世纪以来,词汇习得研究受到越来越多的重视。其研究范围涉及到词汇量评估、心理词库、词汇网络、概念网络、概念流利等诸多层面。研究者从语用学、语言学、语料库语言学、神经科学、计算机语言学、心理语言学、认知语言学等多种视角提出了多种理论模型。

英语在中国是典型的外语,与二语有着本质的区别。学术界把母语外的语言习得统称为二语习得,但在具体研究中,区分二语和外语具有十分重要的意义。环境造就语言。香港的英语学习环境要比内地丰富得多,不仅因为香港是个国际都市,英语的使用频率很高,而且英语也是第二官方语言,很多场合必须用英文交际。这种区分在国内已然成为学界

共识,这一点在我国国内相关研究文献中可见一斑。由于外语的特殊语用环境,汉语作为母语对于与自己差异甚大的另一种语言系统具有较强的排斥力和干扰效应。本章将从词汇的最基本概念入手,逐步介绍词汇、词频统计以及词汇知识方面的一些基本概念。

一 词汇的概念

(一) 汉语中“字”与“词”的关系

关于汉语“字”与“词”的关系,潘文国教授等(2004:1)在其专著《汉语的构词法研究》绪论中是这样表述的:

“中国的传统研究以‘字’为中心,从文字的形、音、义分别衍生出文字、音韵、训诂三大语言研究部门。由于古汉语中字、词不甚区分,很少有人从许慎的《说文解字》中体会到构词法的价值。”

赵克勤(2005:5)在其专著《古代汉语词汇学》第二章中就“字”和“词”之间的复杂关系作了十分详细的阐述。赵在其绪论中对字、词的关系论述言简意赅,笔者特引述如下:

“汉语里有字,有词,有些字等于词,有些字不等于词,同一个字,有时候是词,有时候又不是词。字与词怎样区分,区分到什么程度,对单音词、复音词、连绵字、同义词、反义词、多义词、同音词、通假字、古今字的研究至关重要。上述带‘词’的术语,虽曰‘词’,但与‘字’不无关系。‘字’与‘词’的界限分明,问题就豁然而解;界限不清,就会做出错误的结论。上述带‘字’的术语,虽曰‘字’,实际上就是‘词’,从文字学的角度讲是‘字’,而运用到语言中就是‘词’了……”

字和词的关系在现代汉语日常会话中同样是个混淆不清的概念,难分彼此。比如,

① 我母亲过去没有受过什么教育,大字不识几个。

② 她的那几个字总是写得歪歪扭扭的,除了她自己,没人认得清。

③ 你都会些什么词儿啊?

④ 瞧她常看的漫画书上就没有几个正儿八经的词儿!

以汉语词典释义为例。《现代汉语词典》是这样给“词”下定义的:“语言里最小的、可以自由运用的单位”;而“字”的定义却十分模糊。我们不

妨仔细对比一下汉语关于“字”、“词”的定义、搭配及用法,再对比一下英文中的“word”,便可体会到中国人学英语时强调“背单词”中的“词”会是什么概念?存在哪些误区?

字:①文字:汉字、识字、字体、字义(对比:词义)、常用字(=word)。

②字音:咬字、字正腔圆、她说话字字清楚。

③字体:篆字、美术字。

④书法作品:字画、一幅字。

⑤字眼;词:革命人民的字典中没有“屈服”这个字儿。

⑥字句:立字为凭、收到款子,写个字儿给他。

词:①说话或诗歌、文章、戏剧中的语句:戏词、义正词严、词不达意、她问得我没词儿回答。

②一种韵文形式:宋词。

③语言里最小的、可以自由运用的单位。

汉语中的“字”既是上义词也是自己的下义词,它既是一个“字”的总体名词,是“音”、“形”、“义”的总称如⑤,它同时又分指“音”如②、“形”如③。如果我们说,“词”是“语言里最小的、可以自由运用的单位”,按照“字”的第五个释义⑤,“字”也应算是“语言里最小的、可以自由运用的单位”。由此可见,我们自己对母语中的“字”和“词”并没有明确的区分。但我们会说“别咬文嚼字!”,而不会说“别咬文嚼词!”;我们会说“她的词汇十分丰富,”而不会说“她的字很丰富”。我们能正确使用,靠的是语言习惯。约定俗成是语言的共性之一。从我们出生开始,就已经潜移默化地在影响着我们的表达习惯。一旦习惯养成,就会对另一种语言形式产生排斥和干扰。下面我们拟就汉语中的字词概念对英语学习造成的干扰进行讨论。

(二) 英语词汇学习的误区

由于缺乏正确的词汇意识,只背单词而不记短语和固定表达法的现象在中国学生中十分普遍。有学者通过调查发现,中国英语学习者无论是做单项选择、词汇再现(recall)还是翻译题,总是回避动词短语而倾向使用单个动词,特别是当翻译题中没有选项时尤其如此。比如,中国学生倾向使用“disappoint”而不喜欢用“let down”(Liao & Fukuya, 2004)。单词就是单个的词,不仅中国学生对英语词汇持有误解,许多普通外语教师同样不置可否。当老师建议学生多背单词时,一般都难以明确告诉学

生“背单词”应该注意哪些问题。只要看一看中国的英语教材就能发现其中的问题。中国人自己编写的英语教材在处理生词时,通常只对生词释义;语言点讲解往往以课文中的语境意义为主,尽管所举例子中有明显的搭配关系,这种搭配关系也不会引起重视。生词表中列出短语是因为排除阅读理解障碍的需要,而不是将它们设定为学习目标。如果通过语境可以猜测的或语义比较透明的短语或语块(从字面上可以得出语义的),基本不在其列。即使生词表中列了短语,也不在教师听写范围内。与其说中国学生“回避使用英语动词短语”,倒不如说他们轻视短语学习。在思维的概念层面,由于短语知识匮乏,中国学习者通常以单个的词而不是用短语来表征概念。因此,当某个概念被激活时,能被提取的只有单个的词。

“单词”是单个的词(individual words)吗?“单词”是指“词汇”吗?

卡特(Carter, 1987: 4-5)把单词定义为由左右两个空格将之与其他文字隔开的一个独立的字符串。比如“I appreciate your compliment.”是由四个单词组成的。这个定义并没有说明一个单词的不同形式(屈折和派生形式)以及那些表达同一个意义或概念的习语和复合词等是否算同一个单词。例如“I’m so pleased to hear that.”和“It’s my pleasure.”两句中的“pleased”和“pleasure”是同一个词的不同形式,还是应该单独算一个词?卡特和麦卡锡(Carter & McCarthy, 1988: 18-19)认为,一个自由存在的单词是由一个或多个有意义的“块”组成的。比如单词“propel”是由前缀“pro-”(朝前)和词根“-pel-”(推动)两部分构成。我们可以看到前缀和词根虽有意义却不是单词。这个定义同样也没有解决卡特定义存在的问题。

复合词是词吗?从语义学上来说,复合词指的是真实世界的实体,这个实体造就了这个词并决定了它的复合词组织结构(Carter & McCarthy, 1988: 18-19)。换言之,复杂的心理世界需要人们用两个和两个以上的单词个体去表现。语言学习者往往对词汇抱有很强的心理现实性。由多个单位构成的复合词在心理词库中因为其语义关系通常会被看作是一个单词。这就意味着多单位组合的复合词在心理词库中形成的是一个词位(Carter & McCarthy, 1988: 18-19)。当一个成功的英语学生听到或看到“round-the-clock”这个动词短语时,他并不是从心理词库中分别提取每个单词的词义,然后进行词义组合,而是直接提取这个复合词的词义,因为他是把它当作一个独立的词汇单位储存在心理词库中的。显然,对于像复合词类的词块,其认知方式决定了记忆存取的方式。如果

知道词汇包含了复合词、习语等,学习者就会把它们当成一个单词那样去学习,这样,他在提取有关信息时就十分的快捷。否则,提取就会失败。

哪些属于词汇范畴?为了更清楚地回答这个问题,语言学家提出了另一个术语“词位”(lexeme)。词位和词素(morpheme)、义素(sememe)形成相对概念。一个词位就是一种语言中的最基本的有意义的词汇单位。它能以多个不同形式出现在口语或书面语句子中,其不同形式包括其屈折形式,也包括派生形式。例如,“throw”的屈折形式“threw”、“thrown”、“throwing”,派生形式“thrower”都属于一个词位。其短语“throw out”、“throw a party”,甚至像“throw caution to the wind”、“throw the baby out with the bath water”这样的习语都分别属于一个独立词位。在当代英语词典中,如《朗文高阶英汉双解词典》(2006版,外语教学与研究出版社),基本都把这样的短语和习语作为一个独立词条列出,而不是像《英华大词典》(郑易里,曹成修,1996,商务印书馆)将其列在头词“throw”词条解释中。里昂斯(Lyons, 1981: 145)更明确地将词位分为单词词位(word-lexeme)和短语词位(phrasal lexeme)。单词可以被看作是一个核心抽象形式,包括变体或屈折形式。不仅单个的词可以被看作是一个词位,习语和复合词也可以。比如“takes”、“taking”、“taken”都属于一个单词词位。同样,像“kick the bucket”、“pass away”、“cat burglar”等表达法,每个都是一个短语词位。但是像“a dirty dog”这样的表达法就需要根据语境区别对待。如果有位女人骂自己丈夫是“A dirty dog!”,意思是说他是个“下流坯”。这句话中的“a dirty dog”是一个短语词位而不是三个单词词位。如果骂她的狗太脏,“a dirty dog”就是三个单词词位。

基于词位概念,词汇学习就不再只是学习单个的词。正确的词汇概念应该包括任何具有词汇意义的词汇单位,不管是单个的词还是复合词、短语还是习语,只要其表征一个独立的语言单位或概念,都必须当成一个心理词汇单位来学习和存储。只有这样的记忆和存储,才会有相应的表征和提取。

二 词频统计

在论述词汇知识前,我们先必须弄清几个概念:一是统计语言学中的词汇知识(lexical knowledge/vocabulary knowledge),它涉及到词类

(word type)、文本覆盖率(text coverage)等概念;二是二语习得视角下的词汇知识,强调二语词汇学习目标、启动词汇量(start-up vocabulary)、词汇知识的广度与深度、词汇能力等;三是认知语言学中的词汇知识概念,强调词汇各种知识点在认知个体记忆中的表征、可及性以及概念网络的建立等。

统计语言学视角下的词汇相关概念

语言中的词汇到底有多少?这个问题不是一个简单地依靠人工统计就可以完成的。首先要弄清楚的问题是:什么是词汇?如果你要问中国学生“take”算不算词汇,答案是肯定的。但如果要问“taken”、“taking”是否需要另算一个词时,答案就不一致了。这还只是涉及到词汇的屈折变化形式,如果要考虑词的派生词、复合词、转类词、缩略词等,问题就更复杂了。本节中,我们只对几个核心概念做简单的介绍。

词汇可以归为几类:根词(base)、派生词(derived word)、专有名词(含复合专有名词)(proper words)、复合词(compounds)。更宽泛地说,词汇还包括符号、词缀、缩写形式、古词以及方言词等。复合词也包括三种形式,松散式如“express mail”、半紧密式“know-how”、紧密式“knockout”。专有名词通常是指字典中那些首字母大写或斜体的词。专有名词特别不受中国公共外语学习者的重视,人名、地名等词汇均不在词汇学习目标之列。派生现象给词汇统计增添了不少麻烦。首先,同样都是派生词,有的被算作一个独立的词条,有的则只算作根词的一个派生形式,比如多数带-ness的名词,带-ly的副词。其主要标准是看派生词的词义,如果派生词的词义与根词发生偏离,且产生了认知差异,需要付出一定的认知努力方能习得,这样的派生词应算独立的词条。不规则屈折形式同属于派生词。有些通过前缀派生出来的词被归为根词,尽管其构词成分的词义具有较高的语义透明度。另外,有些粘着词根构成的派生词,哪一个算根词,哪些是其派生词,也没有定论,比如 congruity 被看做是根词,而 congruent、congruous、congruism 则被视为其派生形式。这背后的理据依然具有争议。

可见,词汇概念看似简单,要字斟句酌起来就会发现它似一团乱麻,理不出头绪。

1. 词汇标记(word tokens)、词类、类型-标记比(type-token ratio)

词汇标记是指在语料中被实际观察到的词汇单位。比如,一个自然