

Research Hotspots and Trends of French Learner Corpus: Visual Analysis Based on Citespace

Chang Zhang, Yuhan Wang

Dalian University of Foreign Languages, Dalian 116044, Liaoning province, China

Abstract: With the help of Citespace visualization software, based on the bibliometric method, this paper analyzes the development context, core authors, research hotspots and frontier trends of the research on the construction and application of French learner corpus in Scopus database from 1999 to 2024. It is found that the international corpus of French learners has shown a positive trend of continuous growth, which can be divided into two stages: germination stage and fluctuation growth stage. The core research focuses on second language acquisition and computer-assisted French teaching, with special attention to error analysis, verbal communication, pronunciation training, data-driven learning and so on. The topics of cross-language comparative research, data-driven instructional design and corpus annotation automation are the research frontiers. Through analysis, there are some problems in the existing research, such as the small scale of French corpus, the lack of research on Asian French learner corpus, the lack of integrated research on other factors such as multimodal data and context, and the need to strengthen interdisciplinary research.

Keywords: French; Learner corpus; Corpus application; Bibliometrics

法语学习者语料库研究热点及趋势——基于Citespace的可视化分析

张畅 汪裕涵

大连外国语大学, 中国·辽宁大连 116044

摘要: 借助 Citespace 可视化软件, 基于文献计量法, 对 1999—2024 年斯高帕斯 (Scopus) 数据库中关于法语学习者语料库建设与应用的相关研究的发展脉络、核心作者、研究热点与前沿趋势进行分析。研究发现国际上法语学习者语料库研究整体呈持续性增长的积极态势, 可分为萌芽期和波动增长期两个阶段; 核心研究多集中在二语习得和计算机辅助法语教学方面, 特别关注偏误分析、言语交际、发音训练、数据驱动学习等内容; 跨语言对比研究、数据驱动教学设计及语料标注自动化等话题为研究前沿。通过分析, 现有研究存在法语语料库规模较小、针对亚洲法语学习者语料库的研究不足、对多模态数据和语境等其他因素的整合研究较少、跨学科研究有待加强等问题。

关键词: 法语; 学习者语料库; 语料库应用; 文献计量学

1 引言

随着全球信息化的快速发展和大数据技术运用的不断深化, 语料库已经日益成为语言学及其跨学科研究中的重要工具。语料库作为“一组用于语言学研究的电子化的言语样本集合^[1]”或可以将其形容为“机器可读的真实文本 (包括口语数据的转录本) 集合^[2]”, 为我们从更为宏观的角度观察研究语言提供

了新的视角和方法。而学习者语料库作为语料库中的一员, 指“通过收集外语或二语学习者的自然语料, 利用计算机软件工具进行加工整理而成的语言数据库^[3]”, 这一数据库不仅记录了学习者的语言使用情况, 还能够帮助我们更好地观察学习者的不同阶段, 分析其语言特征。

自 20 世纪 90 年代初国际学习者英语语料库 (ICLE) 项目启动以来, 学习者语料库的建设逐步发展, 相关研究也受到越来越多的关注。根据鲁汶大学英语语料库语言学研究制定的《世界学习者

项目信息: 辽宁省社会科学规划基金青年项目“中国法语学习者语料库建设与应用研究” (L20CYY007)。

语料库一览表》统计，截止到2025年3月，全球范围内已经完成或正在建设的学习者语料库项目超过201个，目标语主要为英语（超过2/3），其次为西班牙语、德语、法语、意大利语等。法语方面，已经建成或在建的规模较大的学习者语料库已有20个，如比利时鲁汶大学的“FRIDA”法语中介语语料库，建立于1998年，包含约20万词的法语学习者书面材料；英国法语学习者口语语料库（FLLOC），建立于2001年，该语料库由七个子库构，各子库在进行纵向研究的同时，也可进行横向间的对比研究；加拿大“Dire Autrement”法语学习者语料库，库容约为5万词，采集母语为英语的中高级法语学习者作文语料，主要进行法语学习词典的研究与编纂。

根据《全球法语现状2022年综述》，截至目前全球有超过3.21亿法语使用者，作为一门遍布五大洲，为32个国家与地区作为官方语的语言。法语在国际外交、文化交流等领域有重要的地位，根据联合国相关预测数据，到2050年，法语有望发展成为仅次于中文的第二大语言。此外，随着中国法语专业高等教育的发展，截至2025年3月，中国共有148所高校开设了法语本科专业，面对这一趋势，对全球范围内法语学习者语料库的研究情况进行综述，不仅是法语专业教师开展教学研究的题中之义，也可为中国法语专业高等教育发展提供参考。

鉴于此，本文将通过斯高帕斯（Scopus）平台和比利时鲁汶大学的学习者语料库文献库（Learner Corpus Bibliography），检索二十五年间（1999–2024）发表于期刊及国际会议的有关法语学习者语料库的相关文献，回顾法语学习者语料库研究的发展历程，对其研究现状、研究热点以及未来发展趋势进行梳理与总结，以期为中国的相关研究提供参考。

2 研究设计

2.1 研究问题

本研究拟回答以下三个问题：(1) 1999–2024年国际法语学习者语料库研究的发文情况(年发文量、作者、国家)如何？(2) 法语学习者语料库的研究热点是什么？(3) 法语学习者语料库研究的前沿及趋势有哪些？

2.2 数据来源

通过检索斯高帕斯（Scopus）数据库，以“french” “learner” “corp*”为关键词进行检索，主题词检索逻辑确定为“TITLE-ABS-KEY (french

AND learner AND corp*) AND PUBYEAR > 1998 AND PUBYEAR < 2025” 文献时间跨度为1999年1月1日至2024年12月31日，共得到338条文献。从中剔除重复以及与主题不相关的文献，最终纳入统计的有效文献为127篇。

2.3 研究方法

本研究运用文献计量法，以Citespace软件（v.6.4.R1版本）作为主要数据处理工具。该软件是由美国德雷赛尔大学华人学者陈超美教授研发的信息可视化软件，通过生成科学知识图谱，可直观呈现科学知识结构、规律和分布情况^[4]。文献计量法与可视化技术的结合有助于分析研究文献，探析某一领域的研究进展及发展趋势。

研究的具体步骤如下：（1）将Scopus中检索到的127条文献的题录信息下载为TXT格式文档，导入至Citespace软件中，进行数据预处理；（2）设置时间跨度为1999–2024年，时间切片为2年，比例系数 $k=25$ ，阈值（top N per slice）为50，依次选择节点为作者、国家、关键词，并对不同类型节点进行可视化图谱绘制；（3）采用文献计量方法对使用Citespace获得的数据进行研究热点及趋势的分析，并尝试探讨目前研究存在的问题和未来发展方向。

3 研究结果

3.1 时间分布

为了全面把握法语学习者语料库研究的文献数量变化规律，我们对相关文献进行了数量统计（图1）。从发文量整体变化来看，法语学习者语料库受到国内外学者的持续关注，该领域的文献呈现总体增长趋势。研究发展可分为两个阶段：萌芽期（1999–2016）、波动增长期（2017–至今）。



图1 1999–2024年法语学习者语料库研究年度发文量分布图

3.2 核心作者及合作关系

基于研究文献，利用Citespace软件以“作者”为节点类型，生成法语学习者语料库研究领域核心作者及合作关系网络图谱（见图2）。图中的节点大小代表了作者发表论文的数量，节点越大，表明学者的发文量越多。根据生成的图谱，我们统计了发文量前五位的国内外学者（见表1）。

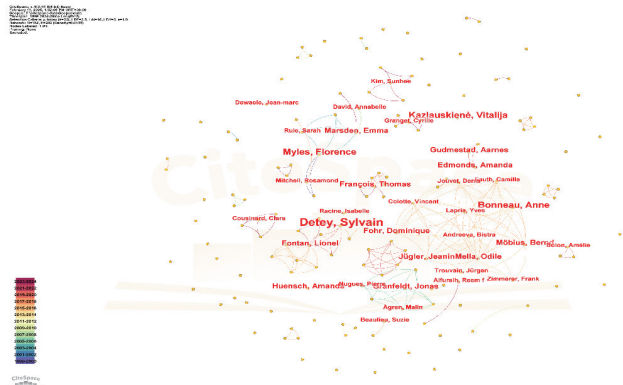


图2 法语学习者语料库研究（1999–2024）作者共现图谱

表1 法语学习者语料库研究（1999–2024）高发文量作者统计

序号	学者姓名	发文量
1	Detey, Sylvain	7
2	Myles, Florence	4
3	Kazlauskienė, Vitalija	4
4	Bonneau, Anne	4
5	Marsden, Emma	3

Myles, Florence的研究方向主要集中在第二语言习得（SLA），特别强调语料库语言学和计算机技术在二语习得研究中的重要性，例如如何使用计算机工具（如CHILDES组件）来分析法语学习者语料库^[13]。Kazlauskienė, Vitalija的研究则集中在对法语学习者词汇习得^[14]、句法搭配^[15]的情况，具体而言，通过关注学习者如何掌握词汇搭配，尤其是名词–形容词搭配（nomino–adjectival collocations）和名词短语（noun phrase），探讨学习者搭配运用能力与语言能力和语言表现之间的关系。Bonneau, Anne则更多关注二语习得中学习者的语音差异，共同参与的研究者还有Laprie, Yves, Bernd等，他们关注母语与二语之间的语音干扰，特别是在法语和德语的对比中，例如音素误发和发音位置的干扰^[16]，并且通过双语语音学习者语料库IFCASL，进行半人工的标注和分句，研究学习者在

在不同音素（如喉塞音、阻塞音）下的不同表现^[17]。

Detey, Sylvain作为统计中最高产的作者，发文数达到7篇。与他一起研究的研究员还有Fontan, Lionel、Ferran é, Isabelle等。他们的研究涉猎广泛，跨声学、感知语音学、计算机科学与二语教学等学科，主要研究方向包括法语学习者法语发音和语言特征的量化研究，其中包括口语流利度^[18]、语音连贯性^[19]等。总体来看，该团队研究学习者在发音问题、语言流利度和语音特征的分析，同时研究如何通过自动化处理工具和语料库技术促进语言习得过程中的精确评估和外语教学的改进。

3.3 国家分布

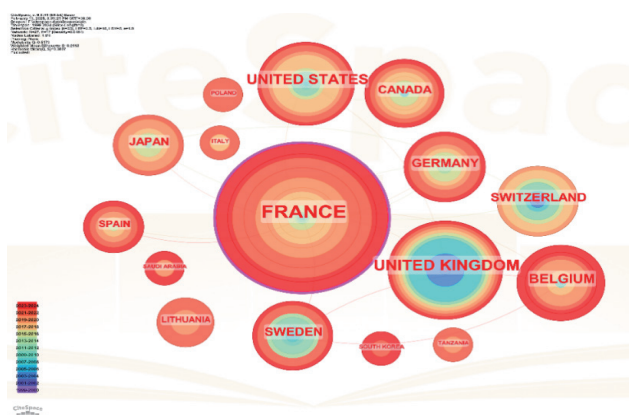


图3 法语学习者语料库研究（1999–2024）国家共现图谱

将127篇相关文献导入引文空间分析，对文献发表国家整理并可视化后发现，法语学习者语料库研究呈现“一核多边，协同多元”的特点。

“一核多边”是指法语学习者语料库建设以欧洲作为核心，但并不是“一家独大”的发展模式，而是多国合作、多元化发展的模式。核心国家主要有法国、英国、比利时等，其中法国以37篇发表量居首（中心性0.22），英国以17篇发表量次之，两个个国家研究多关联“学习者语料库”、“语音水平”等传统语言学主题。通过观察两个国家发表相关文献的年份构成，我们可以看出英国虽然相关研究起步早（最早一篇于1999年发表），但发展的势头随着时间有所减弱，法国起步虽略晚于英国（最早一篇于2009年发表），但随着时间的推移，发展势头依然强劲。与此同时，法国周边国家也积极参与法语学习者语料库的相关研究中，例如比利时、德国、瑞士、西班牙等，形成了多边合作的研究格局。

表2 法语学习者语料库研究高发表国家

序号	国家	发表篇数	序号	国家	发表篇数
1	法国 (FRANCE)	38	6	瑞典 (SWEDEN)	9
2	英国 (UNITED KINGDOM)	17	7	瑞士 (SWITZERLAND)	9
3	美国 (UNITED STATES)	14	8	加拿大 (CANADA)	8
4	比利时 (BELGIUM)	11	9	日本 (JAPAN)	7
5	德国 (GERMANY)	9	10	西班牙 (SPAIN)	5

“协同多元”则指随着对法语学习者语料库研究的不断深入，参与国家越来越多，研究地域也在同步扩展，国际合作不断深化，协同相应逐步显现，数据也趋于多元化，在法语学习者语料库的起步阶段，研究以欧美国家为主，但在近期，更多非欧洲国家（如日本）以及发展中国家（如尼日利亚、坦桑尼亚）也逐渐参与到法语学习者语料库的建设中来，这些国家研究的主题也多与“自动化”等技术应用领域相关联。

3.4 研究热点分析

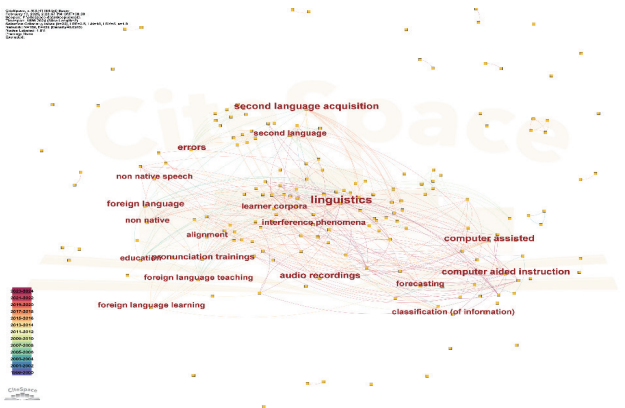


图4 法语学习者语料库研究（1999—2024）关键词共现图谱

表3 法语学习者语料库建设（1999—2024）前十位高频关键词统计

序号	关键词	频次	序号	关键词	频次
1	语言学 linguistics	7	6	发音训练 pronunciation trainings	4
2	二语习得 second language acquisition	6	7	学习者语料库 learner corpora	4
3	计算机辅助教学 computer aided instruction	6	8	计算机辅助 computer assisted	3
4	学习者语料库 learner corpus	5	9	言语交际 speech communication	3
5	偏误分析 error analysis	4	10	数据驱动学习 data-driven learning	3

关键词是作者对文章内容要点的高度概括和提炼，因此常用高频关键词来分析考察某学科领域的发展动向和研究热点。本文利用CiteSpace 对1999—2024 年法语学习者语料库研究的关键词进行数据处理，绘制关键词共现图谱（图4），我们得到法语学习者语料库研究前二十为高频关键词（表3）。通过表3，并结合对已有文献的系统化阅读和梳理，将相关研究内容主要归结为以下几个方面：二语习得、偏误分析、话语研究、计算机辅助教学。

(1) 二语习得

学界关于二语习得的研究由来已久，以二语习作为关键词在斯高帕斯（Scopus）进行检索可得，这一术语最早于20世纪60年代就在相关论文中出现，那时主要研究的内容包括一语习得与二语习得对比研究^[20]、二语习得在不同条件下的迁移效应^[21]、语言习得过程的认知和情感因素等^[22]，主要与之结合的理论有行为主义^[23]、生成转换语法^[24]等。对于传统二语习得研究而言，学习者的产出是了解学习者潜在心理活动的重要途径。随着法语学习者语料库的不断发展，二语习得领域的研究也有了新的发展空间，借助基于错误标注下的语料库，研究者可以从更大层面概括学习者的特征，建立有关二语生成的心理表征和发展模型^[13]；或以语料库的形式收集整理学习者的中介语系统，创建自动分析器^[25]、探讨语言使用倾向^[26]、提出教学建议^[27]等。

(2) 偏误分析

在收录的文献中，关于学习者偏误分析集中体现在对其产出的句法形态进行研究。Escoubas-Benveniste, M.-P., Di Domenico, S., 2017, Manyasa, J., 2019, Vi é mon, M., 2021，借助的理论有错误分析法、中介语理论^[27]等。尽管具体关注的都是学习者的笔语产出，但具体研究的学习者或研究问题有所不同，涉及语言迁移^[28]、学习策略及语言习得研究，有一定的参考价值。其中Escoubas-Benveniste, M.-P.和 Di Domenico, S 两位学者在2017年研究的结论表明不规则动词占的拼写偏误比例较少，而规则动词在屈折语素上产生了更多的拼写偏误，某些言语语境对言语形式识别的切分也会造成阻碍^[29]。

(3) 计算机辅助教学

随着计算机技术的不断进步，学习者语料库的建设和应用也与之结合得越来越紧密，诞生出越来越多的新课题，计算机辅助教学便是其中之一。计算机辅

助教学的内涵是利用信息技术和自动化系统为语言学习提供支持,语料库在其中是重要的资源。语料库视角下的计算机辅助教学则能够提供大量真实的语言资源,帮助学习者与教师提供语音评估和发音训练的支持^[30]、错误的自动化识别^[31]、词汇与句法复杂度分析等^[32]。

(4) 言语交际

通过检索鲁汶大学英语语料库语言学研究中心制定的《世界学习者语料库一览表》发现,目前已被收录在一览表中的法语学习者语料库有 20 个,其中专门包含口语语料库8个、专门包含笔语语料库9个、多模态语料库2个,以及关注电脑辅助沟通(Computer-mediated Communication, CMC)语料库一个。这一数据从侧面印证了在法语学习者语料库的建设与发展中,学者们对于学习者口语发展和笔语发展同样重视,共同发展。同样,“语音训练”、“言语交际”也作为高频关键词出现在统计数据中。这表现出法语学习者语料库研究者对话语研究的关注,例如Detey, S., Fontan, L., Le Coz, M., Jmel, S.和Robertson, S., Munteanu, C., Penn, G.两个学者团队各自在2020年的研究中,对计算机辅助发音训练(Computer-Assisted Pronunciation Training, CAPT)工具的开发提出了建设性的建议,前者为其中语言流畅性的评估提出了行之有效的算法^[12];后者为专门为法语初学者开发此类工具提供了语料库支撑^[33]。

3.5 研究趋势与展望

表4 法语学习者语料库建设(1999-2024)突变术语统计

序号	关键词	起始年	结束年	强度
1	错误诊断	2003	2008	1.15
2	语料库标注	2003	2012	1.01
3	外语学习	2012	2016	1.07
4	非母语语言	2014	2016	1.34
5	言语交际	2014	2016	1.63
6	发音训练	2016	2020	1.31
7	计算机辅助教学	2015	2016	1.17
8	二语习得	2017	2018	2.11
9	偏误分析	2017	2022	1.56
10	二语英语	2019	2020	1.69
11	学习者语料库	2021	2024	1.55
12	数据驱动学习	2022	2024	1.49
13	自动化	2022	2024	1.14

运用Citespace的突变检测算法(burst detection)能从文献的题目、摘要和关键词中提取突变术语,以

便检测某一学科领域研究的突然增长,从而考察该领域的研究前沿和发展趋势^[34]。运行Citespace生成关键词共现图谱后,我们通过“突变检测”(burstness)探测到法语学习者语料库研究的13个突发术语(表4),从中可以辨识近三十年来该研究的前沿动态及演化历程。

通过表3与表4可以看出,这些突现词与法语学习者语料库高频关键词具有较高的吻合度,如“言语交际”、“发音训练”、“计算机辅助教学”、“二语习得”、“偏误分析”和“数据驱动学习”。它们不仅是研究的热点问题,也是前沿问题,说明近30年来法语学习者语料库研究主题具有相对稳定性。2017—2022年该领域的最新前沿问题为“偏误分析”“二语英语”“数据驱动学习”“自动化”。

偏误分析。目前来看,在法语学习者语料库研究中,偏误分析的研究内容主要体现在对不同母语背景的学习者的笔语产出情况进行分析,研究语际迁移或学习者学习风格与偏误之间的相关性,对学习者的口语产出的研究目前尚有不足。从事该领域研究的学者主要有Manyasa, J., Vi é mon, M., Escoubas-Benveniste, M.-P.和Di Domenico, S.。其中以Manyasa, J.较为典型,其主要收集了坦桑尼亚四所大学的大学生法语笔头产出,于2019年和2021年分别调查了61名学习者语法阴阳性偏误和形态句法偏误,在研究中主要采取定性或定性与定量结合的研究方法,发现了学习者母语对法语二外语法阴阳性偏误存在影响^[28],并且对如何改进坦桑尼亚大学生学习法语提出了可行的学习策略^[35]。

二语英语。随着法语学习者语料库建设的发展和二语习得领域研究的不断深化,研究的内容已经不仅仅局限于对学习者的学习某一种语言的情况,一些学者着手进行语际对比研究。将两种语言的学习者情况进行对比,研究者能够更好地探索不同学习者、不同语言间的共性,尤其是将语言距离相近或形态学上近似的不同语言进行对比,可以帮助其更好地理解语言间的交互作用和差异,提供语际研究的新视角。这一类的研究诸如通过探讨研究法语与英语二语学习者在形态学复杂度上的发展,有学者分析了涵盖4个不同水平的100名法语二语学习者和100名英语二语学习者的口头产出数据,进而得出结论,法语二语学习者在语言形态复杂性的发展表现出了比英语二语学习者更连续的增长,这有助于提升我们对语言形态的复杂性对二语习得影响的了解^[36]。

言语交际与发音训练。法语学习者语料库研究对象十分多元,不仅体现在结合学科的多样性,也体现在关注学习者产出时,同时关注了学习者的笔语和口语。通过大量的学习者数据,研究者不仅能够较为科学地归纳出学习者在真实语境中遇到的语言使用问题,提高学习者的交际能力。比如对进行语音标注后的法德双语语料库进行分析,探讨学习者因为音位误用导致的理解问题,同时通过对话分析,揭示学习者的发音问题,并帮助其改正^[37]。另外,也能够与信息技术相结合,开发计算机发音自动评估系统^[31]。

数据驱动学习。在外语教学中,语料库为其提供了“数据驱动”的视角,这有助于从学习者自身角度来解释语言学习者在学习时出现的各种语言现象,更便于研究者发现语言使用背后的因素。通过语料库,研究者能够从定量分析的角度分析学习者在词汇、搭配等方面的使用情况。数据驱动学习(Data-Driven Learning, DDL)是一种基于真实语言数据的语言学习方法,通过语料库收集人们在真实语境下的语言产出,可以提升学习者对语言构式的理解和运用^[38],对于教学理念的转变和教学内容与形式的多样化也有所帮助^[39]。语料库作为数据驱动学习的核心要素,不仅为教师提供了丰富的教学资源,也为学习者提供了直观的学习资料。尽管作为一种新兴的研究课题起步较晚,但法语相关语料库建设已经有了一些建树,例如ORFEO平台(法语书面与口语工具与研究)为面向语言学和自然语言处理领域的研究者和学习者提供了现代法语语料库资源,学生通过这一工具,能够更好地与实际语境数据互动,增强学习的有效性和实用性^[40]。

自动化。“自动化”也成为目前法语学习者语料库建设的一个重要课题,随着信息技术的发展和研究的深入,不仅仅是法语学习者语料库,语料库建设越来越离不开信息技术和自动化的支持,也有越来越多的学者以实现更高的自动化水平而努力,例如有的学者结合组合范畴语法(Applicative Combinatory Categorical Grammar, ACCG)理论,分析出现在学习者文本中的宾语附着代词,以期建立能够处理学习者中介语的自动化系统^[41];另外,在外语教学中的测评中,也有一些学者利用TCFLE-8语料库,探讨学习者笔语产出的特征与学习者熟练度之间的关系,从而争取开发作文自动化评分系统(Automated Essay Scoring, AES)^[42]。

4 结论与展望

本文借助斯高帕斯(Scopus)平台和比利时鲁汶大学的学习者语料库文献库(Learner Corpus Bibliography),对法语学习者语料库研究的相关情况进行了总结,梳理了法语学习者语料库建设的历程,分析了该领域研究的特征。研究得出:法语学习者语料库研究历经起步阶段(1999-2016)、与深化阶段(2017-至今)两个阶段,已经取得了一定的成果,并且呈现加速发展的趋势。目前,法语学习者语料库研究以法国作为核心,以欧洲国家为主,呈现“一核多边、协同多元”的特征,近几年也有一些非欧洲国家(如日本)以及发展中国家(如尼日利亚)逐渐参与到其中。另外,从研究内容来看,法语学习者语料库研究热点集中在二语习得、偏误分析、话语研究和计算机辅助教学几个方面,研究对象较为多元,研究问题比较广泛,具备不同的理论与实践价值。最后,通过“突变检测”的方式,我们考察了法语学习者语料库研究领域的发展前沿和趋势,可以看出当前法语学习者语料库研究具有相对稳定性,一些关键词,如“偏误分析”及“数据驱动学习”,不仅是研究的热点,同时也是前沿问题。

但从目前来看,法语学习者语料库研究仍存在一些值得继续深化改进之处。

从源头出发,现有法语学习者语料库资源和规模有限,标注标准不统一。相比之英语,法语学习者语料库建设起步晚、规模小,研究人员可供选择的语料库资源受限。另外,标注与转录格式的不统一,限制了数据的互操作性和跨库比较。

以研究为导向来看,目前绝大多数研究主要关注文字或语音数据,对多模态数据和语境等其他因素的整合研究较少。其次,目前在采用定量研究的论文中,多采用统计分析的方式,这种方式可能无法捕捉到学习者在产出过程中的语用和心理变化,导致对学习过程的理解不够全面。最后,研究往往关注单一层面(词汇、句法等),但在实际学习过程中,各层面存在密切联系,如何客观科学地整合多层次数据有待解决。

随着有关自动化研究的不断深入,自动化工具不断为学者开发与使用,这将有助于建设更大规模的语料库资源。同时,自动化工作的开展将有助于丰富语料库元数据维度,扩展语料库领域与类型,标准化数据采集与标注,为法语学习者语料库研究进一步松

绑。最后,推动加强学习语料库研究与先进统计与技术的融合,以便于研究者能够捕捉学习者语言中的微妙变化和跨层次互动。尽管目前法语学习者语料库的研究尚有待发展和深化,但作为一门较为新兴的研究,法语学习者语料库研究大有可为。

参考文献

[1] Hunston, S., 2022. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

[2] McEnery, T., 2006. *Corpus-Based Language Studies: An Advanced Resource Book*. London and New York: Routledge.

[3] 尤易, 曹贤文. 20年来国内外学习者语料库建设及应用研究分析. *国际中文教育 (中英文)*, 2022, 7(02):5-14.

[4] 李杰, 陈超美, 2016. *CiteSpace: 科技文本挖掘及可视化*[M]. 北京: 首都经济贸易大学出版社.

[5] Myles, F., Mitchell, R. and Hooper, J., 1999. Interrogative Chunks in French L2: A Basis for Creative Construction? *Studies in Second Language Acquisition*, 21(1): 49-80.

[6] Dewaele, J.M. and Regan, V., 2001. The Use of Colloquial Words in Advanced French Interlanguage. In: *Proceedings of the Annual Conference of the European Second Language Association*, Krakow, Poland, September 7-10, 2000. Amsterdam: John Benjamins.

[7] Rézeau, J., 2007. L'apport du concordancier à l'analyse et à la remédiation des erreurs des apprenants dans les forums de discussion en ligne. *Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication*, 10(2): 27-43.

[8] Hamel, M.-J. and Milicevic, J., 2007. Analyse d'erreurs lexicales d'apprenants du FLS: démarche empirique pour l'élaboration d'un dictionnaire d'apprentissage. *Canadian Journal of Applied Linguistics*, 10(1): 25-45.

[9] David, A., 2008. Vocabulary Breadth in French L2 Learners. *The Language Learning Journal*, 36(2): 167-180.

[10] Marsden, E. and David, A., 2008. Vocabulary Use During Conversation: A Cross-Sectional Study of Development from Year 9 to Year 13 Among Learners of Spanish and French. *The Language Learning Journal*, 36(2): 181-198.

[11] Pillot-Loiseau, C., Amelot, A. and Fredet, F., 2010. Contributions of Experimental Phonetics to the Didactics of the Pronunciation of French as a Foreign Language Stage 1:

Establishing a Corpus. *Recherche et Pratiques Pédagogiques en Langues de Spécialité - Cahiers de l'APLIUT*, 29(2): 75-88.

[12] Detey, S., Fontan, L., Le Coz, M., 2020. Computer-Assisted Assessment of Phonetic Fluency in a Second Language: A Longitudinal Study of Japanese Learners of French. *Speech Communication*, 125: 69-79.

[13] Myles, F., 2005. Interlanguage Corpora and Second Language Acquisition Research. *Second Language Research*, 21(4): 373-391.

[14] Kazlauskienė, V., 2019. The Apposition in the Written Production of Lithuanian Learners in French. *Sustainable Multilingualism*, 15: 89-104.

[15] Kazlauskienė, V., 2022. Nature et acquisition des collocations lexicales en FLE. *Darnioji Daugiakalbystė*, 20: 194-209.

[16] Bonneau, A., 2021. Voicing Assimilations by French Speakers of German in Stop-Fricative Sequences. In: *INTERSPEECH 2021*.

[17] Bonneau, A., 2018. Impact of Fluency and Segmental Categorization in L2: The Case of French Final Fricatives Uttered by German Speakers. In: *Speech Prosody 2018*.

[18] Detey, S., Fontan, L., Le Coz, M. and Jmel, S., 2020. Computer-Assisted Assessment of Phonetic Fluency in a Second Language: A Longitudinal Study of Japanese Learners of French. *Speech Communication*, 125: 69-79.

[19] Fontan, L., Le Coz, M. and Detey, S., 2018. Automatically Measuring L2 Speech Fluency Without the Need of ASR: A Proof-of-Concept Study with Japanese Learners of French, 2018:2544-2548.

[20] Spolsky, B., 1969. Attitudinal Aspects of Second Language Learning. *Language Learning*, 19(3-4): 271-275.

[21] Jakobovits, L.A., 1969. Second Language Learning and Transfer Theory: A Theoretical Assessment. *Language Learning*, 19(1-2): 55-56.

[22] Brown, H.D., 1973. Affective Variables in Second Language Acquisition. *Language Learning*, 23(2): 231-244.

[23] Ewing, W.K., 1972. The Mentalist Theory of Language Learning. *Foreign Language Annals*, 5(4): 455-462.

[24] Reid, C.L., 1973. Discovery of Latent Structure as the Major Process in Language Acquisition: Implications for the Second Language Program. *Foreign Language Annals*, 6(4): 481-486.

- [25] Granfeldt, J., Nugues, P., Persson, E., Thulin, J., Agren, M. and Schlyter, S., 2006. CEFLE and Direkt Profil: A New Computer Learner Corpus in French L2 and a System for Grammatical Profiling, 2006:565–570.
- [26] Lemmens, M. and Perrez, J., 2010. On the Use of Posture Verbs by French–Speaking Learners of Dutch: A Corpus–Based Study.
- [27] Tracy–Ventura, N., Huensch, A., Mitchell, R., Le Bruyn, B. and Paquot, M., 2021. Understanding the Long–Term Evolution of L2 Lexical Diversity: The Contribution of a Longitudinal Learner Corpus. *Learner Corpus Research Meets Second Language Acquisition*, 148.
- [28] Manyasa, J., 2021. When Language Transfer is Negative: Analysis of Morpho–Syntactic Interference Errors by Learners of French in Tanzanian Higher Learning Institutions. *Journal for Foreign Languages*, 13(1): 165–190.
- [29] Escoubas–Benveniste, M.P. and Di Domenico, S., 2017. La Dictée de Français et les Étudiants Italophones. *Analyse Explorative des Erreurs de Morphologie Verbale. Bulletin VALS–ASLA*, 105: 77–94.
- [30] Park, S., Chun, J., Kim, M.H., Lee, H., Lee, S.H. and Kim, S., 2022. Designing a Korean French–Learners’ Speech Corpus (KFLSC) for Spoken Language Assessment. In: 2022 25th Conference of the Oriental COCOSDA, 1–6.
- [31] Laborde, V., Pellegrini, T., Fontan, L., Maclair, J., Sahraoui, H. and Farinas, J., 2016. Pronunciation Assessment of Japanese Learners of French with GOP Scores and Phonetic Information. In: *INTERSPEECH 2016*, 2686–2690.
- [32] Aleksandrova, D. and Pouliot, V., 2023. CEFR–Based Contextual Lexical Complexity Classifier in English and French. In: *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, 518–527.
- [33] Robertson, S., Munteanu, C. and Penn, G., 2020. FAB: The French Absolute Beginner Corpus for Pronunciation Training. In: *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 6613–6620.
- [34] Kleinberg, J., 2003. Bursty and Hierarchical Structure in Streams. *Data Mining and Knowledge Discovery*, 7(4).
- [35] Manyasa, J., 2019. Analysis of French Grammatical Gender Errors Committed by Learners in Tanzanian Universities. *Journal for Foreign Languages*, 11(1): 65–86.
- [36] De Clercq, B. and Housen, A., 2019. The Development of Morphological Complexity: A Cross–Linguistic Study of L2 French and English. *Second Language Research*, 35(1): 71–97.
- [37] Jouvét, D., Bonneau, A., Trouvain, J., Zimmerer, F., Laprie, Y. and Mobius, B., 2015. Analysis of Phone Confusion Matrices in a Manually Annotated French–German Learner Corpus. In: *Workshop on Speech and Language Technology in Education*.
- [38] 许群爱, 邓云华. 基于数据驱动学习模式的英语工具名词带宾构式习得研究. *外语电化教学*, 2024(01):77–84+123.
- [39] 李德超, 王克非. 基于双语旅游语料库的 DDL 翻译教学. *外语电化教学*, 2011(01):20–26.
- [40] Lafontaine, F., 2021. Proposition d’exploitation du corpus d’étude pour le français contemporain en didactique du FLE. *Jazykovedny Casopis*, 72(4): 951–966.
- [41] Affes, A., Biskri, I. and Jebali, A., 2022. French Object Clitics in the Interlanguage: A Linguistic Description and a Formal Analysis in the ACCG Framework. In: *International Conference on Computational Collective Intelligence*, 220–231.
- [42] Wilkens, R., Pintard, A., Alfter, D., Folny, V. and Francois, T., 2023. TCFLE–8: A Corpus of Learner Written Productions for French as a Foreign Language and Its Application to Automated Essay Scoring. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 3447–3465.