阿尔兹海默症患者语言研究的文献计量分析

廉紫蓉　郭爱萍
（太原理工大学外国语学院　山西　晋中　030600）

【摘　要】本文以 Web of Science 核心数据库所收录的阿尔兹海默症患者语言研究的 291 篇文献为数据来源，借助 CiteSpace 的信息可视化软件工具，从文献共被引、关键词两方面，对阿尔兹海默症（AD）语言研究的趋势进行可视化分析。研究发现：AD 语言研究知识基础表现在原发性进行性失语、颞部语言参数、双语、语义表达、话语等方面；热点及前沿集中在言语表现、连贯语言、诊断准则等方面。本研究较为客观地揭示了二十年来国际对于 AD 患者语言研究的现状，旨在对我国 AD 语言研究提供一定的参考和借鉴。

【关键词】阿尔兹海默症；语言；可视化分析；CiteSpace

1 引言

阿尔茨海默症被称为老年性痴呆，是一种最常见的神经退行性疾病，患者因大脑逐渐退化而出现记忆减退、日常生活能力和行为障碍等症状，多发于 60 岁以上人群[1]。2020 年一项人口学研究估计，我国 60 岁以上的痴呆患者有 1507 万人，其中阿尔茨海默病占到了 983 万人。面对与日俱增的患者数量，一个世纪以来，研究者从 AD 的发病机理、诊断、防治等角度进行了广泛研究并取得了一定成果[2]。为了更好了解 AD 患者语言研究的最新进展、前沿热点和发展趋势，本文拟采用文献计量的方法，运用 CiteSpace 软件对 2002-2021 年共 20 年间有关 AD 患者语言研究的国际文献进行分析，为国内将来开展相关研究提供理论依据。

2 研究方法

研究者在 Web of Science 核心合集中以“Alzheimer Disease”“language”“linguist*”为关键词进行主题检索，共检测到文献 1313 篇，经人工多次筛选，排除与语言研究无关文献，共计采集文献 291 篇。本文采用的为 CiteSpace 5.7.R5W 版本，对收集到的文献进行关键词分析及文献共被引分析。

3 计量结果及分析

3.1 基于文献共被引分析的知识基础

本文采用经筛选的文献数据，绘制基于文献共被引网络的 AD 语言研究知识图谱，旨在呈现 AD 语言研究的核心领域。经运行后得到可视化网络图谱，包括 600 个网络节点和 1840 条连接线。聚类值 Modularity Q 为 0.8711，剪影度为 0.9557。根据共被引网络图谱，核心研究领域从大到小依次为：原发性进行性失语、颞部言语参数、双语、语义表达、话语、习得年龄、一致性、心智理论、平均距离依赖、音系错误、朗读。为进一步探索和挖掘国际 AD 语言研究的知识基础和研究前沿演变过程中起关键作用的文献，本研究选取前五大聚类中高中心性文献进行深度解读。

第一个聚类为“原发性进行性失语 (PPA)”，这是由中枢神经系统变性疾病导致的一组综合征，其临床特征为语言功能呈慢性进行性损害，而其他认知功能在早期相对保留[3]。这一聚类下，文献数量最多共 69 篇，研究者主要通过区分失语症类型及其相应不同的语言表现，以期推动形成失语症临床评估量表、AD 诊断指南、及探究脑组织结构的病理改变与语言障碍之间的对应关系。这一聚类中心度最高的一篇文献为 Classification of primary progressive aphasia and its variants. 作者 Gorno-Tempini 在这篇文献中提供了原发性进行性失语症及其 3 个主要亚型的分类：非流利 / 语法进行性失语，语义进行性失语和少词型进行性失语。患者被诊断为 PPA 后，会根据每种亚型的特定言语和语言特征进行分类。非流利 / 语法进行性失语的语言特征主要表现为语法缺失及言语失用（指缓慢、费力的语音产出）；语义进行性失语（也叫语义性痴呆）核心特征为失语症和独词理解能力不足；少词型进行性失语的主要特征为单词检索缺陷（在即兴演讲和对证命名）及句子重复缺陷。

第二个聚类为“颞部言语参数”，文献数量有 56 篇，研究者大多基于数据库中痴呆症患者与健康老年人的语料对比，通过计算机技术等自动化分析手段从语音表现、词汇选取、叙述性语篇、口语理解与表达等方面得出可用于区分患者与健康人的语言特征。其中中心度最高的一篇为 Deficits in narrative discourse elicited by visual stimuli are already present in patients with mild cognitive impairment. 作者 Cláudia Drummond 要求研究对象根据以正确顺序呈现的七个场景来讲述一个故事（“车祸”事件），通过前测及实验引入并验证了评判叙述性语篇表现的七类定量参数：叙述时间、叙述字数、语篇类型（叙述性或描述性）、连贯性、指称衔接、话语有效性指标及叙事结构。这些定量参数可用于有效辨别健康老年人及认知障碍症患者，其中最显著的参数指标为话语有效性指标。

第三个聚类为“双语”，这一聚类的剪影度最高为 0.99，表明该聚类下内部文献相似性较高，这些文献通过实验数据证明双语学习对于延迟痴呆症发作、促进认知储备起到了积极作用。这一聚类中中心度最高的一篇文献为 Bilingualism tunes the Anterior Cingulate cortex for conflict monitoring. 作者 Jubin Abutalebi 对比了双语学习者及单语学习者的语言表现，研究发现双语者不仅可以用较少的神经资源解决认知冲突，而且他们的大脑还可以更好地适应冲突情况。大脑中前扣带皮层（ACC）是语言控制和解决非语言冲突的常见部位。研究发现双语者比单语者能够更加有效地运用该结构。早期曾有学习两种语言会导致儿童智力延迟的误解，但这篇文献却证明早期学习和终身学习使用双语对人类新皮质发育产生了极大的效果。

第四个聚类为“语义表达”，该聚类主要关注 AD 患者语义表现、双语患者的语言流畅性等，其中心度最高的一篇文献为 Bilingual language production: The neurocognition of language representation and control. 作者 Jubin Abutalebi 借助神经影像数据来探讨双语产出与单语产出之间的联系。作者发现双语语言的产出是一个动态过程，涉及皮质和皮下结构，可利用抑制来解决词汇竞争和选择预期语言。

第五个聚类为“话语”，此类研究数量相对较少共有 27 篇，核心文献为 Kathleen C. Fraser 所发表的 Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. 作者提出使用计算技术对原发进行性失语症中的连贯发音进行评估和分类，即从三组研究人员（语义痴呆（SD），进行性非流利性失语症（PNFA）和健康对照组）叙事语音的转录中自动提取句法和语义特征，差异较大的特征则被用来调整机器学习分类器。对影响特征的分析表明，与对照组相比，两个患者组都倾向于使用频率较高的单词（SD 倾向于使用名词，PNFA 则使用动词）。 SD 患者倾向于使用熟悉度更高的单词（尤其是名词），并且与对照组相比，他们产生的名词更少，但指示词和副词却更多。PNFA 小组的语言往往较对照组慢，并且所用单词较短。

3.2 前沿热点及发展趋势

表一：高频关键词及高中心度关键词表（2002-2021）

排名	高频关键词		高中心度关键词	
	关键词名称	频次	关键词名称	中心性
1	memory	77	task	0.27
2	aphasia	73	comprehension	0.22
3	diagnosis	44	aphasia	0.2
4	impairment	43	agraphia	0.19
5	deficit	38	working memory	0.19
6	age	29	impairment	0.18
7	speech	26	attention	0.18
8	disease	21	context	0.17
9	working memory	18	age	0.17
10	performance	18	association workgroup	0.17

关键词一般被认为是对文献核心内容的总结，对高频关键词的研究则可揭示某一学科领域的研究热点和重要主题。表一为软件统计得出的 2002-2021 年间高频关键词及高中心度关键词。在 CiteSpace 图谱中，一般将中心性大于 0.1 的节点视为关键节点。可以看出，AD 语言研究主要通过语言任务 (task) 来探究患者的语言理解 (comprehension)、语言缺陷 (aphasia)、记忆 (memory\working memory) 等方面的缺陷（deficits \impairment）。笔者同时分析了 20 年间突发性强度排名前五的关键词，旨在揭示一定时期内 AD 语言研究的热点：AD 患者的知识水平对语言的影响在 2007 年到 2011 年研究较多；在 2010 年到 2015 年研究者研究者更侧重于患者语言表现的研究；言语、连贯语言、诊断准则则在近些年内热度较高。在这五个突现词中，突现度最高的是“言语”（3.98），其次为“诊断准则”（3.39），这两词表明患者的话语，尤其是口语表现及 AD 的诊断准则颇受学界青睐。从突现词影响周期来看，“连贯语言”作为突现词的影响周期最长（6 年），说明语言学家较为重视患者话语的连贯表现。

另外，笔者分析了依照 LLR 算法自动生成的关键词时间线图，共产生 13 大聚类，依次为物品、语义性痴呆、注意力、少词型进行性失语、双语、老龄化、大脑皮层、沟通策略、颞部言语参数、血管性痴呆、音乐、自然口语及 sci。聚类顺序按照关键词数量由高到低排列。该聚类聚类值 Modularity Q 为 0.7816，剪影度 S 为 0.9196，说明聚类结果高效令人信服。经研究发现前五类聚类时间线较长，且连线紧密，说明此些领域研究具有连贯性且长久不衰，而其余聚类连线显著不似之前密集，说明成果相对较少。如沟通策略这一聚类中连线密集集中在 2007-2013 年间，说明这一时期学界对其关注较多。笔者还发现节点较大的关键词（代表出现频率较高）所出现的年代也都较早，位于时间线图前端，且与近年间的节点仍有密切连线，如 2002 年 aphasia 与 2021 年 mental state examination 仍有连线。这些现象说明高频关键词（见表一）时间跨度极大，在二十年的研究历史中不断被提及、具有重要的研究意义，并将会继续对今后的研究产生影响，属于 AD 语言研究热点。值得一提的是，在近五年中并无中心性大于 0.1 的节点出现。近些年中出现的关键词有：流利度、衔接、连贯、机器学习等，但并未形成较大节点。这一现象说明 AD 语言研究还未形成具有影响力的前沿文献，研究者们仍处于探索阶段。

4 启示

研究国际 AD 语言可视化知识图谱，反思我国研究现状，笔者总结如下。第一：学界仍亟待跨学科交流合作。语言学家需加强与神经学家、病理学家、心理学家等的交流合作以进行更加专业的研究。第二：夯实语料库建设，在记录语料时，需详细记录被试的真实语言表现及健康状态。鉴于 AD 患者大多为老年人，身体状况无法做到控制变量一致，因此语料收集过程中，应记录详尽，完善多模态语料库建设。

参考文献：

[1] 肖义军，陈莉莉．阿尔茨海默症及其预防措施概述 [J]．生物学教学，2020,45 (10):2-4.

[2] 刘红艳．阿尔茨海默症患者语言障碍研究现状和进展——基于病理语言学的实验研究综述 [J]．外语电化教学，2020(05):72-78+11.

[3] 崔博，谭纪萍，王鲁宁．logopenic 型进行性失语的研究进展 [J]．中华神经科杂志，2013(03)

作者简介：

廉紫蓉（1997-），汉族，女，山西吕梁人，太原理工大学外国语学院硕士，研究方向为语言学及应用语言学。

郭爱萍（1967-），汉族，女，山西临猗人，博士、太原理工大学外国语学院教授，研究方向为语言学及应用语言学。