

# 人工智能驱动的语言学研究演进探索

## ——基于CNKI与WOS的可视化比较研究

文嘉仪

四川大学外国语学院，四川 成都

收稿日期：2025年12月8日；录用日期：2026年1月15日；发布日期：2026年1月27日

### 摘要

近年来人工智能飞速发展，语言学研究正在经历深刻的变革。本文基于CiteSpace工具，对中国知网(CNKI)与Web of Science (WOS)数据库中的相关文献进行对比文献计量分析。从年度发文趋势、作者与机构合作网络、关键词共现及关键词演化四个维度展开研究，系统呈现国内外语言学研究在人工智能语境下的发展趋势。研究结果表明，中文文献更关注AI在语言教学、语料库技术等教育场景中的应用，英文文献则侧重于技术实现、计算模型及语言认知研究。本研究揭示了中外研究重点的差异，并探讨了差异形成的原因，同时为未来语言学与人工智能跨学科融合的方向提出了建议。

### 关键词

人工智能，语言学研究，文献计量分析，可视化

# Exploring the Evolution of AI-Driven Linguistic Research

—A Visualization-Based Comparative Study from CNKI and WOS

Jiayi Wen

School of Foreign Languages, Sichuan University, Chengdu Sichuan

Received: December 8, 2025; accepted: January 15, 2026; published: January 27, 2026

### Abstract

In recent years, the rapid development of artificial intelligence brings many changes to linguistics. In this study, we perform comparative bibliometric analysis based on literature from CNKI and Web of Science by using CiteSpace method. We compare these literatures from four aspects: annual

**publication, author and institution collaboration, co-occurrence of keywords and evolution of research topics. The results show that, compared with international research, Chinese studies are more interested in applying AI in education, such as intelligent language teaching and corpus-based learning, while international research pays more attention to technical development, computational modeling and cognitive linguistics. Finding out these differences in research focus, we try to explore some possible reasons for it and put forward some suggestions for future research on cross-disciplinary cooperation between AI and linguistics.**

## Keywords

**Artificial Intelligence, Linguistic Research, Bibliometric Analysis, CiteSpace Visualization**

Copyright © 2026 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. 简介

近年来，人工智能的发展深刻地改变了语言学研究的方法与关注重点，尤其体现在自然语言处理、语义识别和机器翻译等领域。人工智能的广泛应用不仅拓展了语言学研究的范围，也推动了语言学与计算机科学的深度融合。随着大型语言模型及 BERT、GPT 等预训练模型的不断发展，人们开始重新思考语言的本质，以及语言的传递和理解的过程。随着 BERT 及其衍生模型的迅速进步，自然语言处理任务在阅读理解、文本分析、语言翻译和情感分析等应用中也取得了突破性的创新[1]。语言是人类智慧的结晶，而自然语言处理被视为人工智能领域的“皇冠上的明珠”，在人工智能的发展中“掌握语言者将掌握未来”[2]。这一技术变革使学界重新审视语言学的研究路径与根本问题。

数据被视为继土地、劳动力、资本和技术之后的第五大生产要素，并逐渐成为新的关键动力来源[3]。人工智能与大数据、深度学习紧密相关，人工智能正是基于海量数据的训练，加上 Transformer 架构学习中涌现出的语言模型[4]。作为数据的重要组成部分，语言数据的重要性不言而喻。另一方面，语言学的研究已不再局限于传统的理论探索，而是逐渐向更多跨学科和应用型研究领域延伸。语言技术、语料库语言学、情感分析等新兴方向的兴起，推动了语言学研究的多样化以及数据驱动的语言学发展。计算语言学作为一门交叉学科，近年来经历了巨大的变革。过去的计算语言学主要依赖基于规则的模型和符号逻辑，而如今，统计方法与深度学习架构成为推动其发展的重要力量[5]。这些变化也促使语言研究者重新审视关于语言结构、意义与使用的传统问题。Church 和 Liberman 同样指出，深度学习不仅为语言学提供了新的研究方法，也为“如何科学地研究语言”确立了新的方向[6]。

另一个值得关注的趋势是，越来越多的研究者开始将目光投向低资源语言与多语模型。近期，Sahin 证明了可以运用人工智能的方法来采集语言数据，从而缓解语言数据分布不均的问题，并提升模型中弱势语言的表现[7]。因此，研究者呼吁推动语言技术的民主化，使语言技术更加具有包容性。

除了技术层面的变革，我们还可以看到，关于人工智能与语言学融合的研究文献呈现爆炸式增长。为了可视化这一增长领域的知识结构与主题演化趋势，我们采用 CiteSpace 等文献计量工具对相关研究进行追踪与分析。这类工具能够帮助我们识别不同时间与地区的核心研究群、主题聚类以及演化路径，进一步探讨 AI 与语言学研究在不同语言情境和学术语境中的差异。例如，中国数据库(如 CNKI)中的研究可能呈现不同的研究主题、热点或研究重点，而国际数据库(如 WOS)中的研究则可能体现另一种关注方

向。因此，我们通过对两类文献的比较研究，考察人工智能驱动的语言学研究在全球化趋势与本土化路径上的异同。

本研究基于 CiteSpace，对 2015~2025 年间发表在 CNKI 与 WOS 的相关文献进行跨数据库文献计量分析。通过对年度发文量、作者合作网络及关键词演化的分析，本文旨在追踪人工智能推动下的语言学研究主题演变，并探讨中国与国际学界在研究焦点及方法论取向上的共性与差异。

## 2. 数据来源与研究方法

为追踪人工智能领域中语言学研究的主题演化趋势，本研究选取中国知网(CNKI)与 Web of Science(WOS)两大学术数据库作为数据来源。为了获取过去十年全球与本土学术话语的整体图景，本研究将这两个数据库分别作为国内与国际学术界的代表。因此，我们从两大数据库中收集了 2015~2025 年期间发表的相关文献。

检索所用的核心关键词包括：“人工智能”“计算语言学”“自然语言处理”“智能写作”“语言技术”等。对于 CNKI，文献通过高级检索获取，限定为同行评审的期刊论文，不包含报纸与会议论文。对于 WOS，检索范围限定在核心合集(Core Collection)，并选择 Linguistics、Computer Science 以及 Artificial Intelligence 等相关类别下的期刊论文。最终，共获得 CNKI 中文文献 100 篇与 WOS 英文文献 100 篇，并分别以可用于分析的格式导出。CNKI 文献以 Refworks 格式下载后再转换为纯文本文件；WOS 文献则直接以“Plain Text File”形式导出，并包含全部记录及被引文献。

本研究的主要分析工具为 CiteSpace (版本 6.1.R6)，这是一款广泛应用的文献计量可视化软件。借助 CiteSpace，可以生成年度发文趋势、关键词共现图、作者合作网络，并进行新兴术语的突现检测。通过 CiteSpace 对来自 CNKI 和 WOS 的文献进行分析，以揭示年度发文变化、关键词共现网络、时间线可视化结果，以及机构与作者的合作结构。

此外，为增强两个数据集之间的可比性，本研究在两组文献中采用了相同的分析参数设置：节点类型(作者、机构、关键词、参考文献)、时间切片(2015~2025，每片一年)、每片选取被引或共现频次最高的前 50 项。聚类分析基于对数似然比(LLR)算法完成。通过这一系列方法，可以深入探索国内与国际学术界人工智能相关语言学研究的知识结构与发展趋势。

## 3. CNKI 与 WOS 文献年度发文趋势比较

李宇明(2020)指出，从数量与质量来看，语言数据是最重要的数据类型之一，也是关键的生产要素。包括语言数据在内的语言产业，未来将成为数字经济的重要支柱。年度发文量是衡量研究领域发展状况及关注重点变化的重要指标，尤其是核心期刊的发文数量，能够反映学界对该主题的关注度以及研究领域的发展成熟度。通过比较 CNKI (2015~2025)与 WOS (2016~2025)两大数据库的年度发文趋势，我们能够洞察中外学者在该领域的成熟度、研究重点以及关注度差异。

如图 1 所示，展示了国内期刊中人工智能与语言学交叉研究的年度发文趋势。如图 1 所示，CNKI 在 2015 至 2019 年增长缓慢，年发文量维持在 0 至 3 篇之间。从 2020 年起，发文量开始稳步且快速增长：2024 年达到 27 篇，2025 年为 28 篇。发文量的迅速上升可能得益于中国近年来发布的人工智能发展规划以及国家对人工智能跨学科研究的政策推动。

如图 2 所示，WOS 的发文趋势更具波动性。2016 年发文量为 11 篇，此后逐年下降至 2019 年的 3 篇，随后再次上升，并在 2024 年达到峰值 27 篇，然而 2025 年又下降至 10 篇。这一趋势可能反映了国际研究在该领域具有较高成熟度但呈现阶段性波动；2024 年的峰值可能与全球对大语言模型与生成式人工智能的高度关注相关，而下降趋势则可能与研究主题饱和或研究重心转移有关。

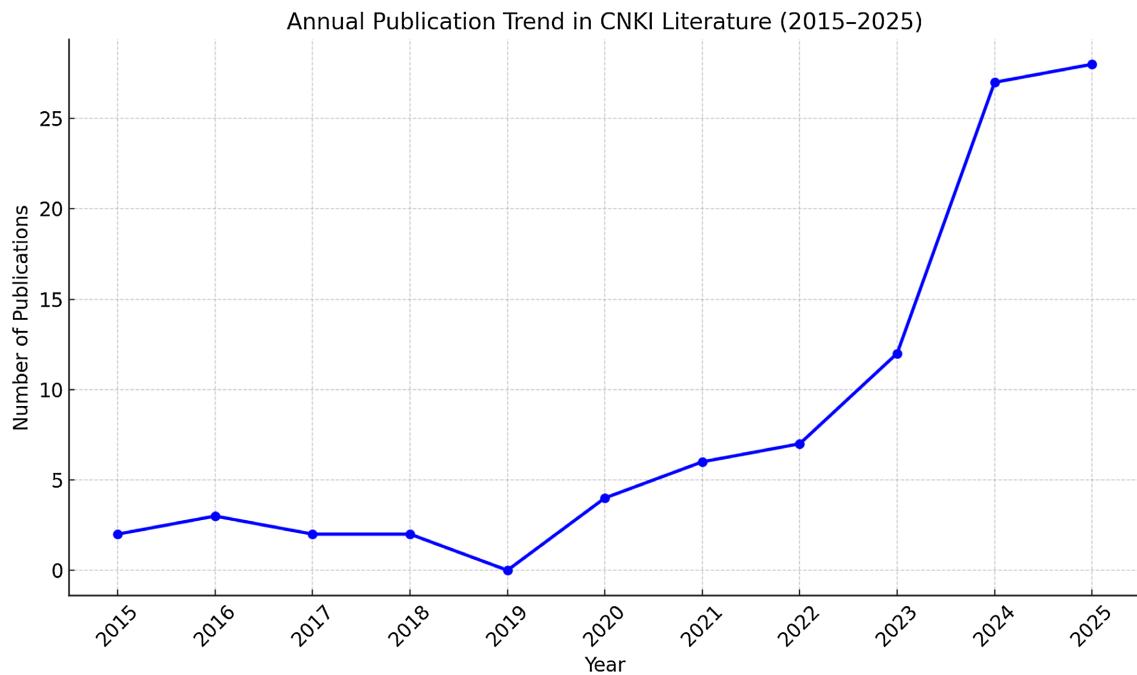
**Figure 1.** Publication trend of CNKI

图 1. CNKI 发文趋势

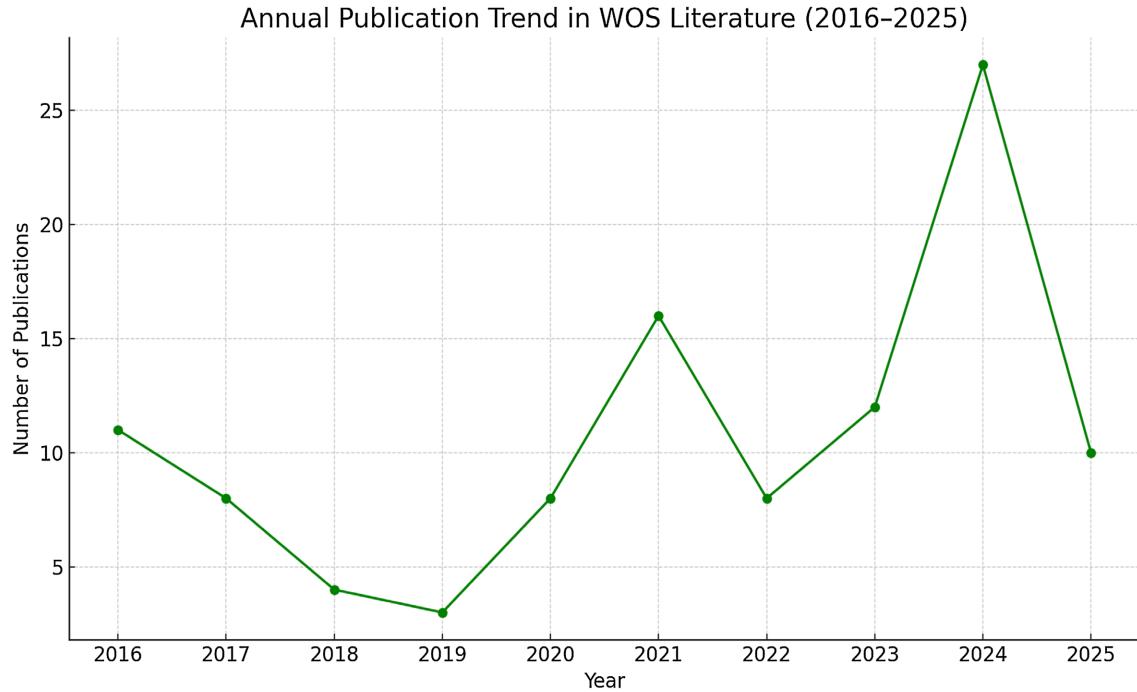
**Figure 2.** Publication trend of WOS

图 2. WOS 发文趋势

值得注意的是，CNKI 和 WOS 在 2024 年均出现显著增长，说明人工智能与语言学研究在全球范围内形成共振。这与全球语言学研究向计算化方法转向的趋势一致。Bender 等(2021)指出，大语言模型近年来引发了关于语言理解及语言泛化的重要学术争论。

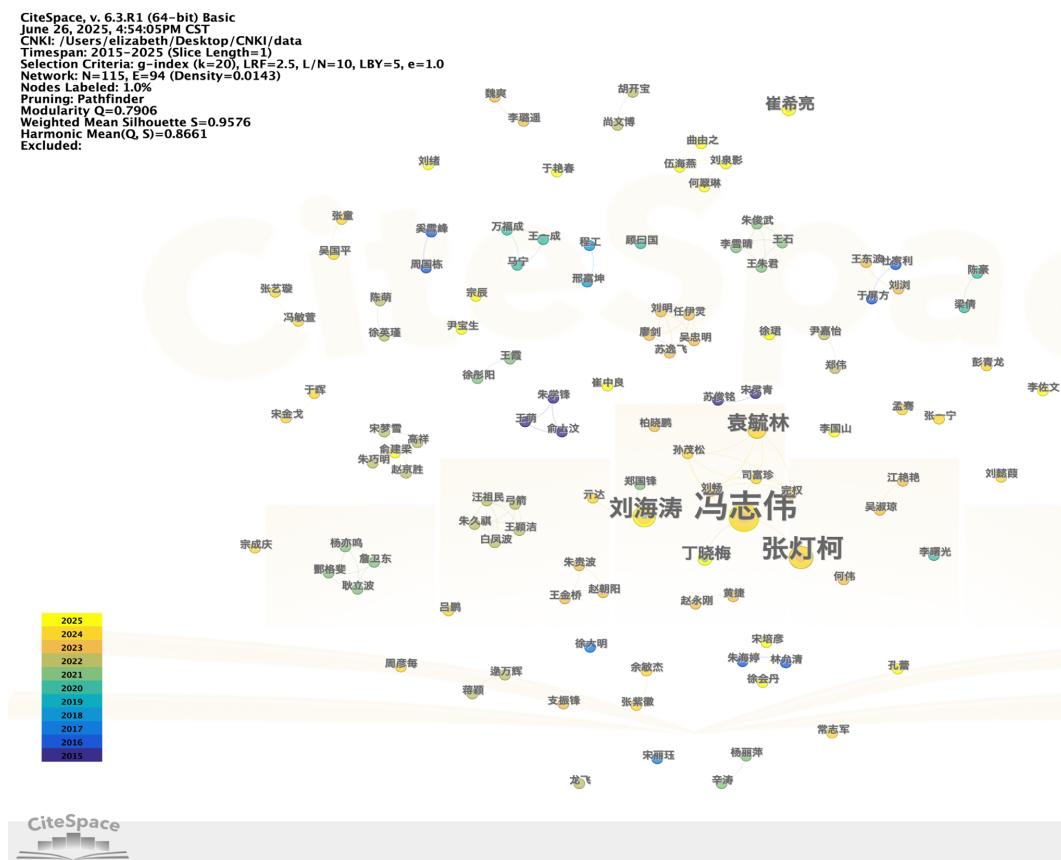
然而，CNKI 的增长明显更为集中且呈爆发式。这表明中国正处于该领域的追赶阶段，同时也是快速内部发展时期。国内研究的显著上升说明人工智能与语言学的交叉研究正处于快速扩张阶段。如前文所述，中国文献的增长在一定程度上体现了全球趋势在教育和语言应用领域的本土化落地。相较之下，国际研究则呈现出较高成熟度及主题细化的趋势。因此，我们应抓住契机，从西方获得理论启发与方法论借鉴，同时以大规模实证研究与语境化应用回馈国际学术界。

#### 4. 人工智能与语言学研究现状

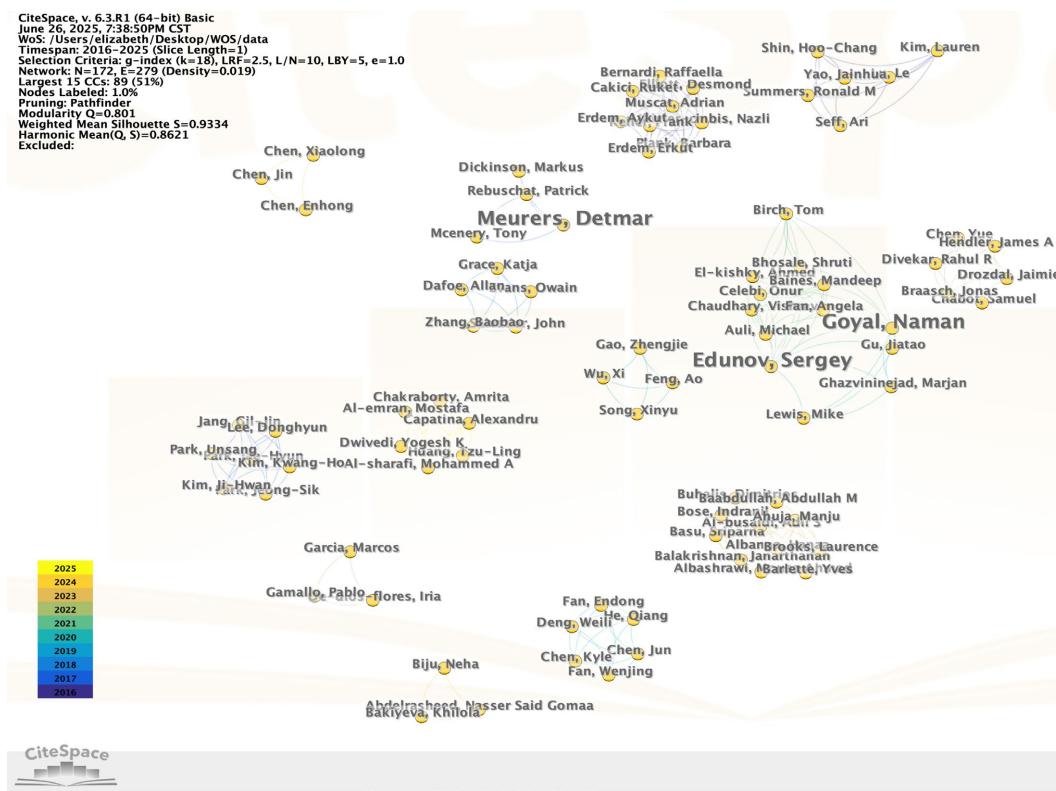
#### 4.1. 作者合作网络比较

研究者共现分析能够识别某一学科领域的核心作者，同时也能反映作者之间的合作强度与引文关系。在作者合作共现图中(图 3)，研究者姓名的字体越大，代表其发表的论文数量越多；作者之间的连线表示合作关系，连线越粗合作越紧密。基于 CiteSpace 对 WOS 和 CNKI 数据库的可视化结果，本节呈现 2015 至 2025 年间人工智能与语言学领域的作者合作状况。

在 WOS 的作者共现网络中(图 4), Goyal、Naman、Edunov、Sergey 与 Meurers、Detmar 具有较高的中心性与较大的合作半径, 表明他们在该领域具有高产出与较强影响力。由这些作者构成的聚类可能对应区域性研究团队或项目小组。Goyal 与 Edunov 与 Auli、Michael 及 Lewis、Mike 存在紧密联系, 这可能表明他们本人或其所属机构(如 Meta AI 或其他大型 NLP 研究中心)与其他学者存在合作关系。其节点呈现明显的时间颜色渐变, 从蓝色到黄色, 说明这些学者持续贡献至 2025 年, 研究活跃度较高。



**Figure 3.** CNKI author co-occurrence network  
**图 3.** CNKI 作者共现网络



**Figure 4.** WOS author co-occurrence network

**图 4.** WOS 作者共现网络

相较之下, CNKI 的作者合作图结构较为分散,但仍可观察到若干规模较大的聚类。冯志伟、臧登科、刘海涛等学者表现为中心节点。与 WOS 的作者网络相比, CNKI 的合作模式更具本土性或机构依附性,即团体内联系紧密,但整体规模较小。不同团簇之间的连线较弱,反映出跨机构合作程度较低。

总体而言,基于 WOS 的作者合作网络呈现出更广泛的国际合作、更紧密的引文联系以及更持续的研究产出;而 CNKI 的作者网络则表现出明显的本土化特征与较强的内部凝聚力。研究结果显示,尽管国内学界在该领域研究活跃且持续增长,但仍有必要进一步加强跨国合作,以提升与全球研究体系的联动性。

## 4.2. 机构合作网络比较

机构共现分析能够反映研究机构在某一领域的学术影响力与科研能力。本节比较了 CNKI (2015~2025)与 WOS (2016~2025)在人工智能与语言学领域的机构共现图(图 5、图 6)。结果显示,虽然整体合作网络的密度仍然较低,但两张图谱中显现的主要机构存在明显差异。

在 CNKI 的机构共现图中,浙江大学是最显著的节点,意味着其中心性较高且在该领域的发文量较大。其他可见机构包括北京外国语大学、上海交通大学以及中国科学院自动化研究所。其余网络较为模糊,表明国内机构间的合作仍相对有限,且多集中于同一区域或同一学科内部。

相比之下,WOS 的机构图呈现出更清晰的国际合作网络。图中包含具有显著影响力的国际研究中心,如香港科技大学、中国科学院、普林斯顿大学及哥伦比亚大学。此外,代表产业研究力量的 Alphabet Inc.与 Google 也出现在网络中,表明该研究领域具有较强的跨学科与产业应用特征。

与 WOS 相比,CNKI 的机构合作网络仍然较为分散,且合作主体主要为学术机构。作为国内研究网络的体现,CNKI 的合作模式更可能局限于区域性或同学科内部的探索。然而,WOS 所展示的则是国际

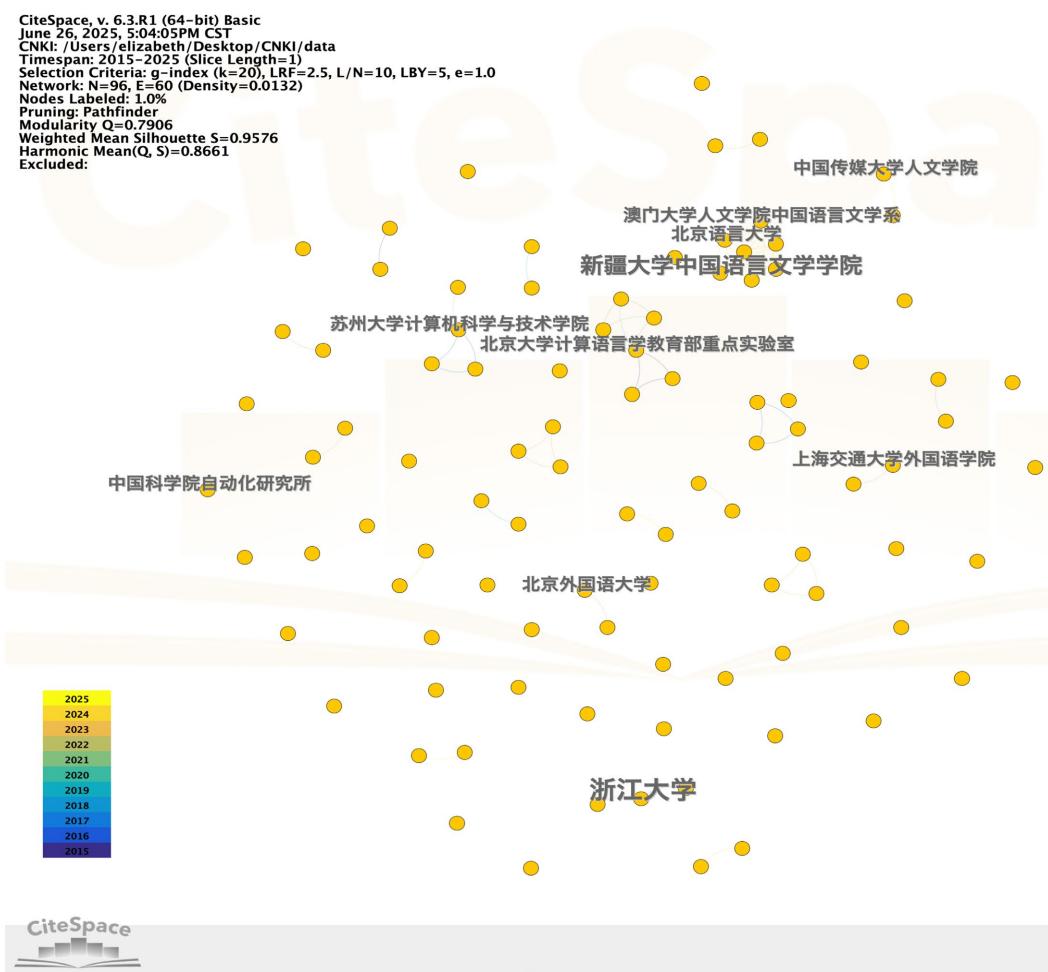
化且产学合作显著的网络结构。这一差异可由中国当前的研究文化与资金体系解释：目前国内的大部分研究仍由高校团队主导，而被 WOS 收录的机构则逐渐加强产业合作。尤其是 Alphabet Inc.与 Google 等科技公司，能够在 NLP 与智能语言教育等实际问题中提供技术支持[8]。

此外，CNKI 网络的模块化程度较低，合作团簇数量较少，这可能限制创新，因为知识更容易在同质化群体中循环。正如 Edunov 与 Goyal 所指出的，跨机构与跨学科合作对于解决语言学中的跨界问题至关重要，尤其是在计算语言学与人工智能教育等领域[8]。因此，中国学术机构不仅需要拓展国内合作，也应强化与国际机构及产业界的联系。例如，通过参与全球开源项目、多语种 NLP 项目以及人工智能教育项目，中国研究团队可以探索更多新的研究与应用可能性。

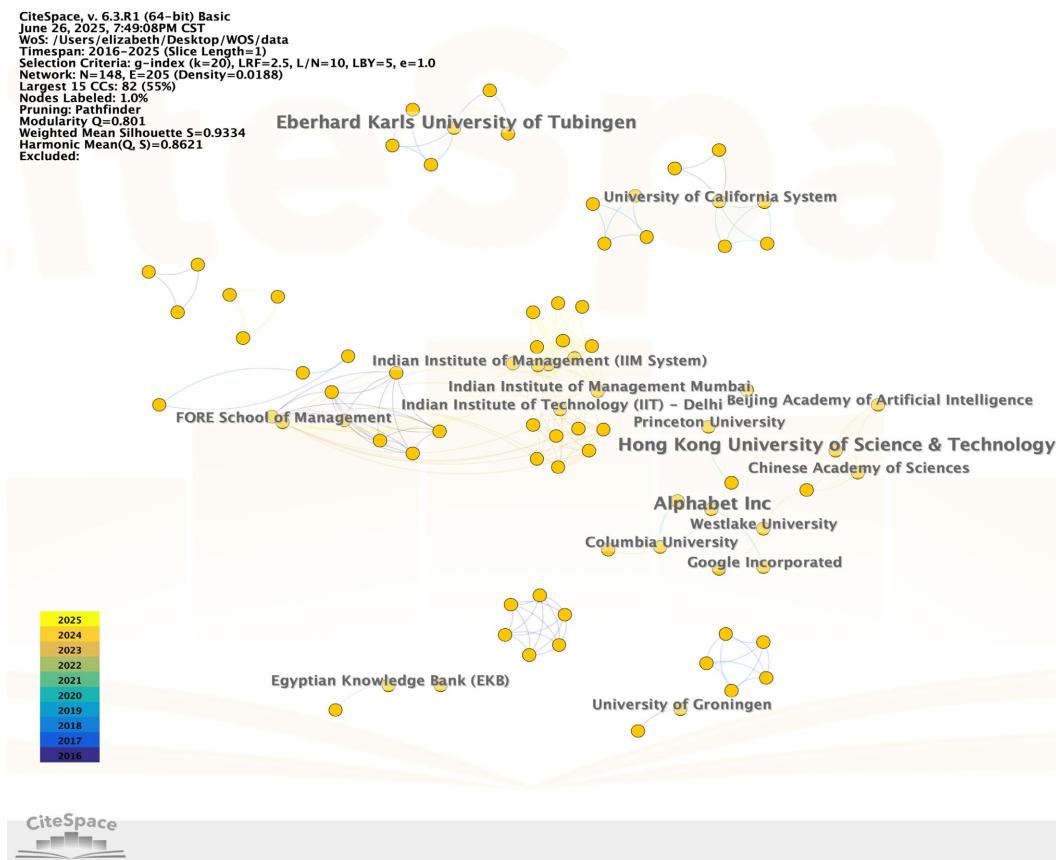
### 4.3. 关键词共现图比较

尽管两者的比较为人工智能与语言研究领域的相对研究趋势提供了有价值的观察，但 CNKI 与 WOS 均凸显了“人工智能”作为研究主题的重要性。然而，由这些数据集构建的关键词网络及其背后的学术共同体，在研究重点、方法论取向和理论基础方面呈现出显著差异。

在 WOS 的关键词网络中，不仅“artificial intelligence”“large language models”“natural language processing”“computational linguistics”等高频出现，而且它们之间具有高度互联性。这种高度整合的



**Figure 5.** CNKI institutional map  
**图 5.** CNKI 机构分布图



**Figure 6.** WOS institutional map  
**图 6.** WOS 机构分布图

研究范式反映了技术发展、语言建模与应用任务在同一研究空间内并行发展的特征。“deep learning”“algorithms”“decision making”等术语的出现进一步表明，WOS 文献具有强烈的技术导向性，firmly grounded in 计算语言处理的机制与模型优化框架。而诸如“2nd language”“learners”“writing”等术语出现在技术核心附近，则显示出一个新兴子领域：将第二语言习得与人工智能结合，用于自动化写作评价与智能个性化辅导。

相较之下，CNKI 的关键词网络呈现完全不同的形态。虽然“人工智能”处于中心位置，但其周围环绕着“语言能力”“语言学习”“生成语法”“语言哲学”等词汇。CNKI 网络呈辐射状结构，即以单个中心外扩连接其他相对独立的学科主题。与 WOS 中多个关键词集群高度融合不同，CNKI 的结构更为碎片化。这表明，在中国学术语境中，人工智能更常被用作辅助工具，以服务语言学或语言教育的问题，而非作为独立的技术或理论对象。

WOS 文献通常表现为技术驱动，其研究主题多围绕拓展大语言模型的能力边界、改进计算语义模型，或优化算法在自然语言任务上的对齐方式。这些主题天然涉及模型解释性、泛化性与可扩展性等问题，反映了计算机科学、信息论与应用语言学的交叉融合。相比之下，CNKI 研究中更具教育导向的主题主要使用人工智能工具促进语言学习、支持语法教学，或借助新技术反思既有语言学理论。

从理论取向来看，这种分歧同样明显。WOS 文献通常基于经验建模框架，如 Transformer [9]，研究重点放在技术优化与实际应用上。例如，某项研究探讨如何通过实时反馈系统提升二语写作[8]，这与 WOS 图谱中“2nd language”“writing”等关键词高度一致。而 CNKI 文献更倾向于援引教育理论，如建构主

义或最近发展区理论，将人工智能视为课堂中的支持性技术或提高学习成效的认知工具。

此外，对关键词图谱的比较也揭示了两种不同的跨学科视角(图 7、图 8)。WOS 的网络显示出计算机科学、语言学与认知科学等成熟学科之间的深度融合；其中“computational linguistics”“semantics”“natural language processing”等聚集成跨领域、兼具理论与应用的研究群体。而 CNKI 虽包含哲学、语言学与教育学等术语，但不同团簇之间融合度较低，表明这些领域尚未真正整合，仍处于并行状态。

可以说，与西方学术界相比，中国学术界虽然对人工智能融入语言研究的兴趣迅速增长，但尚未达到“语言”与“人工智能”在认识论与方法论层面相互融合的阶段。2018 至 2023 年间多数 CNKI 文献关注工具有效性、系统设计或教师态度，而较少探讨人工智能如何改变语言理解的本质或语言学理论本身。

总体来看，WOS 与 CNKI 的关键词共现模式不仅揭示了研究内容的差异，也体现出研究哲学的差异：WOS 将人工智能视为语言研究的驱动力与研究对象，而 CNKI 则将人工智能纳入既有的语言理论或教育框架中，作为支持性工具。或许，中国学界需要进一步吸收人工智能的认识论与方法论，以弥合这一差距，实现技术与语言的双向互动，让语言与技术在相互作用中共同演进。

#### 4.4. 关键词演化趋势比较

下面展示的关键词时间线分别呈现了人工智能与语言学研究的关注重点及其发展方向(图 9、图 10)。在 CNKI 的时间线上，2015~2018 年的研究主要涉及“人工智能”“深度学习”“语言智能”等基础

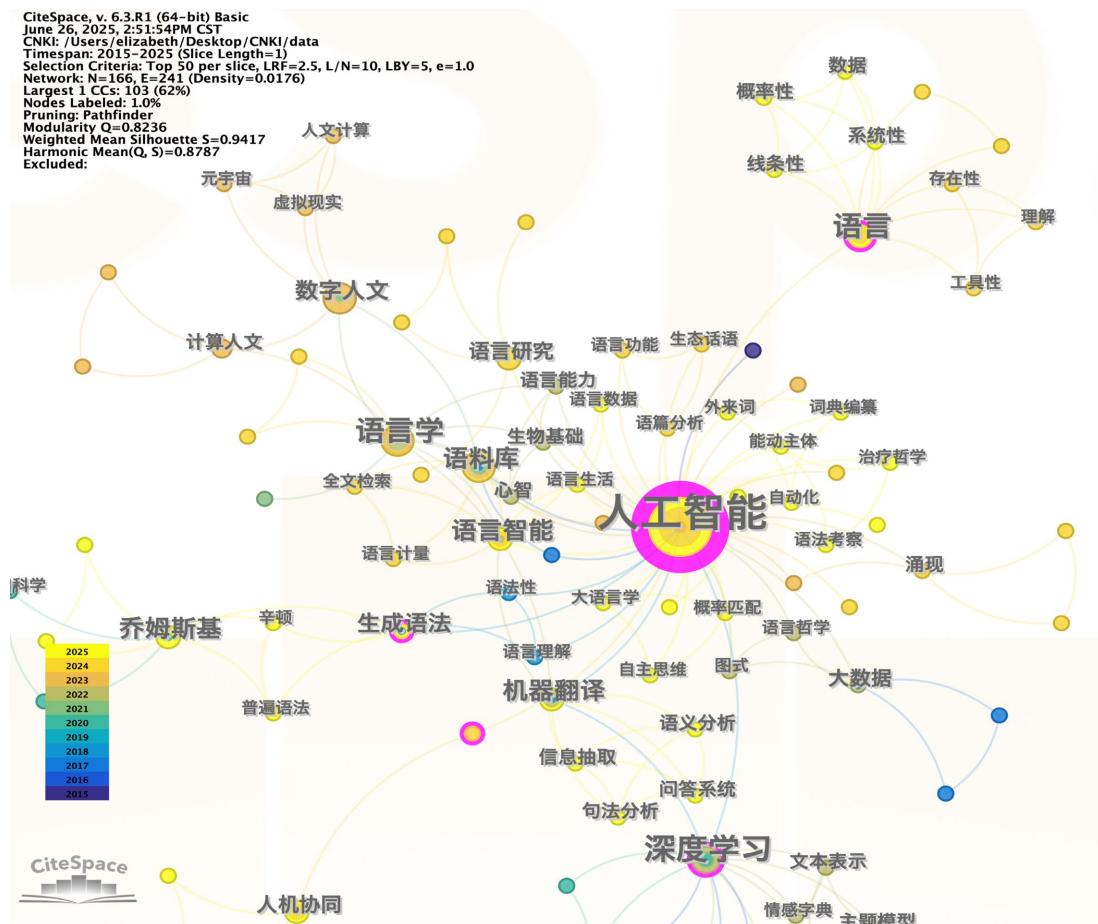
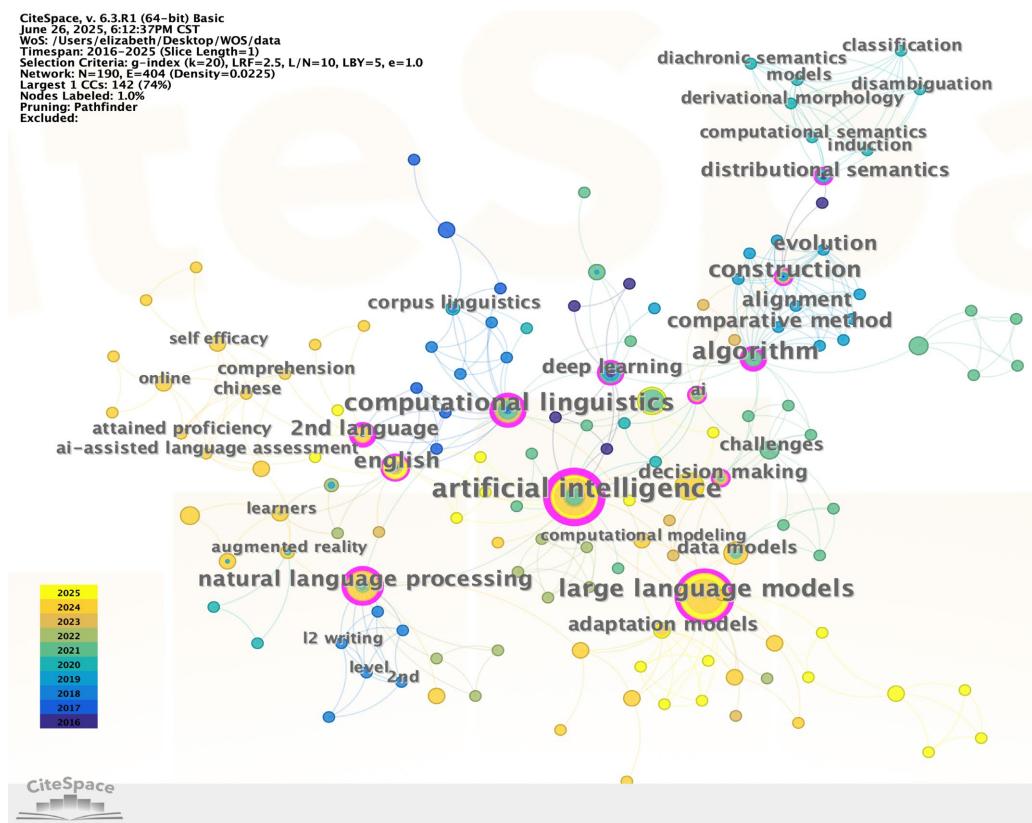


Figure 7. Keyword co-occurrence map of CNKI

图 7. CNKI 关键词共现图



**Figure 8.** Keyword co-occurrence map of WOS

**图 8.** WOS 关键词共现图

概念，说明国内研究在早期更关注人工智能在汉语语言处理及教育领域的应用。自 2020 年起，“提示工程”“系统性”等新概念开始出现，这些概念的涌现表明：算法的发展并不是终点，而是人工智能在语言学与教育领域中实现实际应用的起点。

与此不同，WOS 的时间线呈现出更为明显的技术驱动趋势。早期研究关注“deep learning”“computational linguistics”“corpus annotation”等主题。从 2016 年起，研究主题逐渐转向“large language models”“data modeling”以及“artificial intelligence ethics”。“explainability”“few-shot learning”等聚类的出现说明，全球范围内的研究者不仅致力于新技术的开发，也关注其在应用、伦理及教育中的影响。

这两条研究路径反映了两类学术共同体的不同研究理念。CNKI 的研究更强调人工智能在教育语境中的应用以及社会语言学层面的解释。这与 Zhang 等的观点一致：当前中国教育 AI 的研究重点包括两个方向——其一是 AI 在教学场景中的应用，如智能反馈下的自适应语言学习环境；其二是 AI 在语言学中的应用，如自动语料标注[10]。相比之下，国际学界则从认知与计算的视角讨论人工智能，探讨模型的可解释性、性能及其在多语种 NLP 中的公平性[11]。

此外，关键词演化差异也反映了两类语料库的关注重点。在 CNKI 文献中，研究者更关注语言能力、语义以及人工智能与人类之间的哲学意义；而在 WOS 文献中，前沿技术发展与实证评估占据主流，这与西方学界侧重实验验证与系统扩展性的研究传统一致。

综上，尽管两类语料库均体现出对人工智能与语言学交叉研究的关注，但 CNKI 更倾向于教育应用与理论反思，而 WOS 则更侧重技术创新与伦理思考。只有通过此类跨文化比较研究，我们才能为人工智能在语言学中的未来发展奠定更坚实的基础。

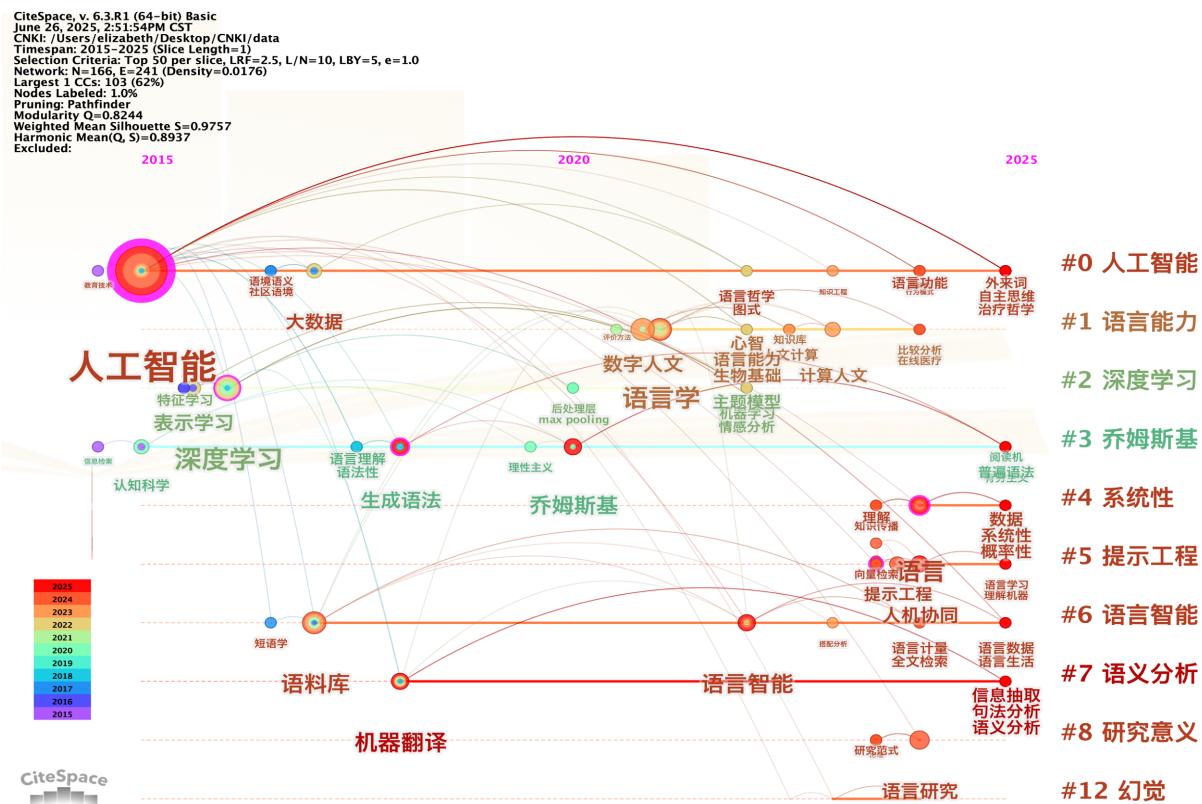


Figure 9. CNKI keyword timezone map  
图 9. CNKI 关键词时间线图

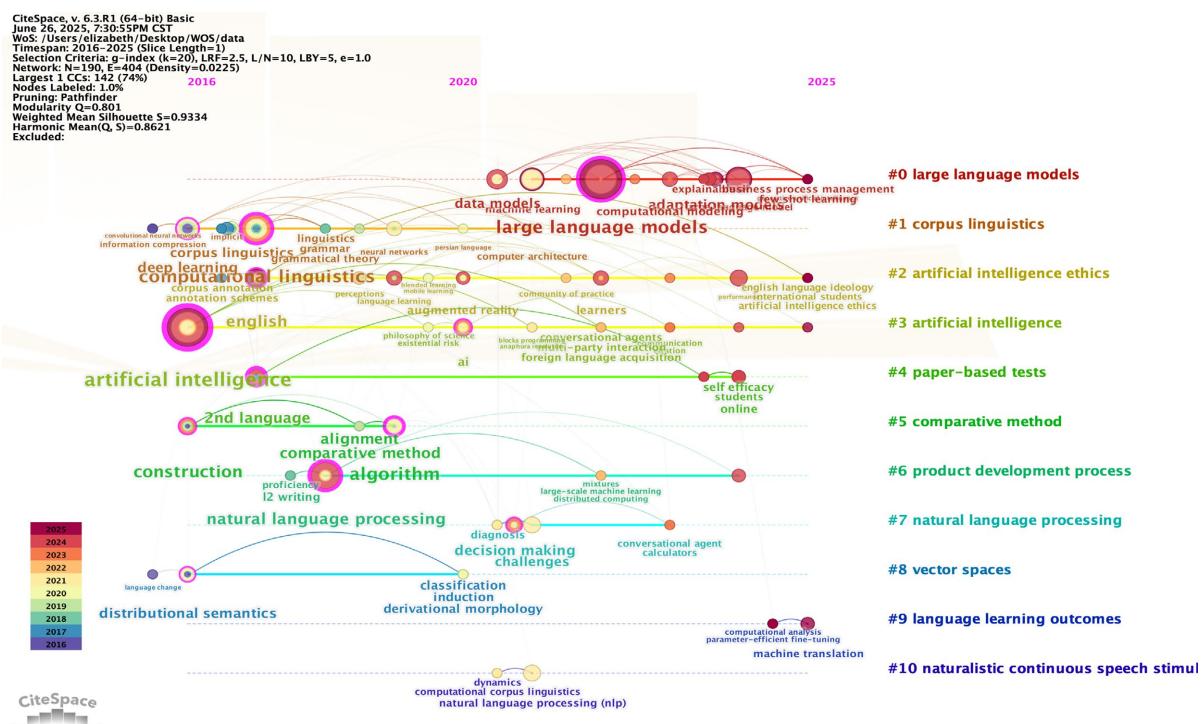


Figure 10. WOS keyword timezone map  
图 10. WOS 关键词时间线图

## 5. 总结

本文基于 CiteSpace 文献计量与可视化分析方法，对 2015~2025 年间 CNKI 与 Web of Science 数据库中人工智能驱动的语言学研究进行了系统比较，从年度发文趋势、作者与机构合作网络、关键词共现结构及主题演化路径等多个维度，揭示了中外学术共同体在研究取向、知识结构与发展阶段上的显著差异。中文文献更多受教育应用驱动，多数研究聚焦于如何利用人工智能工具支持外语教学、基于语料库的教学以及智能辅导系统[12]。国际研究则更具技术性或计算导向性，学者的关注点包括方法创新、大语言模型以及跨学科应用等内容。

在理论层面，本研究主要作出以下三方面贡献：

从科学知识图谱视角出发，本文通过跨数据库比较，揭示了人工智能与语言学交叉研究在不同学术语境中的知识结构差异，补充了以往多集中于单一数据库或单一语种研究的不足。研究表明，WOS 所呈现的知识图谱更具高度整合性与技术导向，而 CNKI 的图谱则体现出应用驱动与学科并行发展的特征，这一发现为理解全球语言学研究的结构性分化提供了实证依据。

在计算语言学研究层面，本文从宏观计量角度展示了人工智能技术如何逐步从工具角色转变为语言研究的核心对象。通过关键词演化与共现模式分析，研究进一步验证了国际学界在模型机制、可解释性与算法伦理等方向上的理论深化趋势，同时指出国内研究在方法论整合与技术内化方面仍有较大发展空间，从而为计算语言学未来的理论发展提供了比较参照。

在跨文化学术研究层面，本文揭示了不同学术传统、研究文化与制度环境对人工智能语言学研究路径的塑造作用。中外研究在研究问题选择、理论引入与评价标准上的差异，反映的不仅是技术发展阶段的不同，也体现了教育需求、科研评价体系与学术话语结构的差异。这一发现有助于推动跨文化视角下对人工智能语言研究的反思与对话。

在实践方面，本文提出以下更具针对性的建议：

对于中国学界而言，未来研究可在保持教育应用优势的基础上，进一步加强对计算模型原理、方法论创新及实验设计的关注，推动人工智能从“教学工具”向“理论驱动因素”转变。应通过参与国际合作项目、开源语料库建设及多语种 NLP 研究，提升国内研究在全球计算语言学网络中的可见度与影响力。在语言教育实践中，研究者与教师可结合国际研究中成熟的智能反馈、自动评估与个性化学习模型，将其与本土教学情境相结合，提升人工智能语言应用的科学性与可持续性。未来研究还应更加重视人工智能语言技术的伦理、公平性与社会影响问题，避免单纯以效率或效果为导向，推动负责任的语言技术发展。

总体而言，本文不仅从宏观层面描绘了人工智能驱动下语言学研究的演进图景，也为理解不同学术共同体在技术融合过程中的路径差异提供了新的解释框架。通过科学知识图谱与跨文化比较的结合，本研究为人工智能与语言学的进一步融合提供了理论参考与实践启示。

## 参考文献

- [1] Matthew, M. (2020) AI, Analytics, Machine Learning, Data Science, Deep Learning Research Main Developments in 2020 and Key Trends for 2021.  
<https://www.kdnuggets.com/2020/12/predictions-ai-machine-learning-data-science-research.html>
- [2] 冯志伟. 人工智能领域：得语言者得天下[J]. 语言战略研究, 2018(5): 85-90.
- [3] Slaughter, M.J. and McCormick, D.H. (2021) Data Is Power: Washington Needs to Craft New Rules for the Digital Age. *Foreign Affairs*, **100**, 54-62.
- [4] 刘海涛. 从语言数据到语言智能——数智时代对语言研究者的挑战[J]. 中国外语, 2024(5): 5-13.
- [5] Kim, W. and Zheng, J. (2023) Compositionality in Computational Linguistics. *Annual Review of Linguistics*, **9**, 463-481.
- [6] Church, K.W. and Liberman, M. (2021) The Future of Computational Linguistics: On Beyond Alchemy. *Frontiers in*

- Artificial Intelligence*, **4**, Article ID: 625341. <https://doi.org/10.3389/frai.2021.625341>
- [7] Şahin, G.G. (2022) To Augment or Not to Augment? A Comparative Study on Text Augmentation Techniques for Low-Resource NLP. *Computational Linguistics*, **48**, 5-42. [https://doi.org/10.1162/coli\\_a\\_00425](https://doi.org/10.1162/coli_a_00425)
- [8] Goyal, N., Edunov, S. and Lewis, M. (2023) Enhancing L2 Writing Feedback with Transformer-Based Language Models. *Transactions of the Association for Computational Linguistics*, **11**, 233-247.
- [9] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N. and Polosukhin, I. (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- [10] Zhang, E.Y., Cheok, A.D., Pan, Z., Cai, J. and Yan, Y. (2023) From Turing to Transformers: A Comprehensive Review and Tutorial on the Evolution and Applications of Generative Transformer Models. *Sci*, **5**, Article No. 46. <https://doi.org/10.3390/sci5040046>
- [11] Blasi, D., Anastasopoulos, A. and Neubig, G. (2022). Systematic Inequalities in Language Technology Performance across the World's Languages. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, Volume 1, 1-14. <https://doi.org/10.18653/v1/2022.acl-long.376>
- [12] 张博, 董瑞海. 自然语言处理技术赋能教育智能发展——人工智能科学家的视角[J]. 华东师范大学学报(教育科学版), 2022, 40(9): 19-31.