

基于LDA模型的生态话语主题识别及演化的顺应性分析

杨亚平

新疆大学外国语学院, 新疆 乌鲁木齐

收稿日期: 2026年4月21日; 录用日期: 2026年5月25日; 发布日期: 2026年6月4日

摘要

本研究以2016~2025年CNKI数据库生态话语期刊文献为语料, 采用潜在狄利克雷分配(Latent Dirichlet Allocation, LDA)主题模型量化挖掘, 结合顺应论四维框架系统探析其语用策略。研究发现: 1) LDA模型识别出7个核心主题, 分别为多模态生态话语、生态文明话语、生态语法、评价理论、企业生态话语、生态哲学、生态教育, 其中评价理论、及物性系统为核心热点; 2) 主题演化可划分为萌芽探索期(2016~2018)、多元发展期(2019~2022)、聚焦成熟期(2023~2025)三阶段; 3) 语用预设策略呈现多维顺应特征, 高频词与主题统计体现语言结构顺应, 主题阶段性演化体现三维语境的动态顺应, 主题强度凸显文化意识顺应。本研究实现计量方法与语用理论的深度融合, 可为生态语言学学科建设、话语传播及相关研究提供参考。

关键词

顺应论, LDA模型, 生态话语, 主题演化, 语用策略

Analysis of Ecological Discourse Theme Identification and Evolution Based on LDA Model from Linguistic Adaptation

Yaping Yang

School of Foreign Languages, Xinjiang University, Urumqi Xinjiang

Received: April 21, 2026; accepted: May 25, 2026; published: June 4, 2026

Abstract

This study takes the journal literature on ecological discourse in the CNKI database from 2016 to

2025 as the corpus, and uses the Latent Dirichlet Allocation (LDA) topic model for quantitative mining. It combines the four-dimensional framework of the Adaptation Theory to systematically analyze its pragmatic strategies. The findings are as follows: 1) The LDA model identified seven core themes, namely multimodal ecological discourse, ecological civilization discourse, ecological grammar, Appraisal Theory, corporate ecological discourse, ecological philosophy, and ecological education, among which Appraisal Theory and the transitivity system are the core hotspots. 2) The theme evolution can be divided into three stages: the nascent exploration period (2016~2018), the diversified development period (2019~2022), and the focused maturity period (2023~2025). 3) The pragmatic strategies exhibit multi-dimensional adaptation features. The high-frequency words and theme statistics reflect the adaptation of language structure, the phased evolution of themes reflects the dynamic adaptation of the three-dimensional context, and the theme intensity highlights the adaptation of cultural awareness. This study achieves a deep integration of quantitative methods and pragmatic theories, which can provide references for the discipline construction of ecological linguistics, discourse dissemination, and related research.

Keywords

Adaptation Theory, LDA Model, Ecological Discourse, Theme Evolution, Pragmatic Strategies

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. 引言

生态话语分析是系统功能语言学与生态语言学交叉的核心领域，旨在探究语言如何参与生态认知、态度塑造与价值建构[1]-[3]。作为生态语言学的重要分支，生态话语分析以韩礼德系统功能语言学为理论根基，聚焦生态语篇的语言表征、意义建构与社会功能，成为应用语言学研究热点[4]-[6]。系统分析其研究主题的演进有利于厘清学科发展脉络与动态规律。

LDA (Latent Dirichlet Allocation)主题模型作为无监督文本挖掘工具，已广泛应用于教育学、图书馆学、应用语言学等领域，为科研热点识别、主题演化分析提供量化支撑[7]-[10]。在生态话语研究中，LDA模型已用于生态科普、环境政策文本的主题提取，但现有研究存在两大局限：其一，方法论融合度不足，缺乏与语言学核心理论的深度结合；其二，研究视角相对单一，多侧重主题内容的描述性分析，未充分阐释语用策略的内在逻辑。

基于此，本研究将LDA主题模型与语言顺应论相结合，以CNKI期刊生态主题论文为语料，一方面通过LDA模型量化识别生态研究主题、划分演化阶段；另一方面从顺应论视角，剖析生态话语在语言结构、语境适配、文化意识层面的语用策略，揭示主题演化的语用逻辑，丰富生态话语分析的研究。

2. 理论与方法

2.1. 理论基础

20世纪90年代，比利时语言学家维索尔伦创立语言顺应论[11][12]。该理论认为，语言使用是交际意图和语言驱动下一个不断选择的过程，是一个语言选择的顺应过程。理论核心包含四个维度：语境顺应、语言结构顺应、动态顺应、顺应的意识程度。

2.2. 研究方法

2.2.1. 数据来源

为探究国内生态话语研究成果的主题演化规律、识别领域热点，本文基于文献检索平台中国知网(CNKI)数据，限定文献类型为期刊论文，以“生态话语/生态话语分析/生态语言学”为主题词检索文献，检索时间跨度为2016年至2025年。

经检索共获取382篇期刊文献，筛选维度包括题名、作者、来源、发文时间、关键词、摘要。剔除无摘要、会议通知、不相关文献后，最终得到有效文献324篇，见图1。

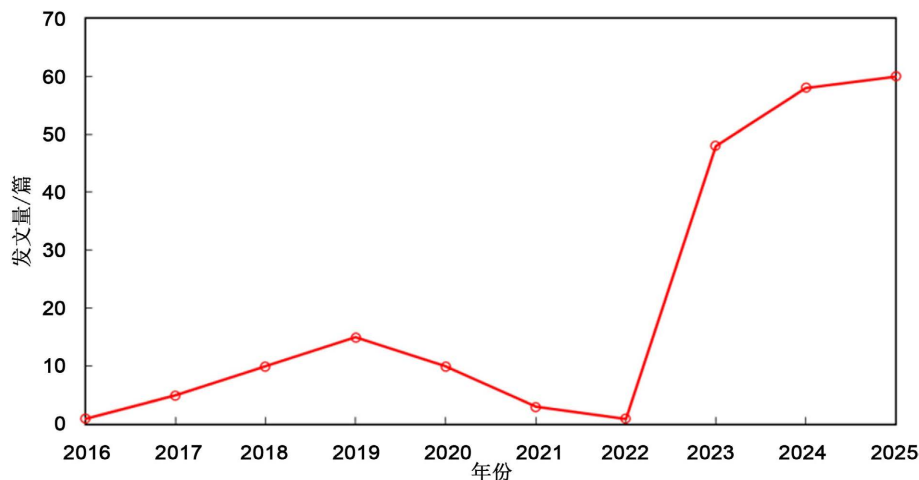


Figure 1. Trend of domestic ecological discourse publication volume from 2016 to 2025

图1. 2016~2025年国内生态话语发文量趋势

2.2.2. LDA 模型构建与评估

LDA (Latent Dirichlet Allocation)主题模型是无监督学习概率生成模型，核心逻辑为：文档由若干主题按概率组合而成，主题由特定词汇按概率分布构成，通过文档-主题-词语三层概率建模，计算主题与文档、主题与词语的语义关联度[13] [14]。

为提升主题识别效率与准确性，采用Python语言jieba库进行文本预处理，包括分词、去除停用词、归并同义词，删除乱码、特殊符号与冗余排版符号，构建纯净语料库。预处理阶段添加自定义术语词典(生态话语、及物性系统、态度资源等)，避免专业术语分词错误。

为确定最优主题数量，依托Gensim库搭建基础LDA模型，本研究依托Gensim库搭建LDA模型，设置主题数 $K=2\sim 12$ (步长为1)，每个主题输出15个特征词，以c-v一致性得分 ≥ 0.35 为标准确定最优主题数。最终主题数 $K=7$ 时c-v得分最高，模型效果最优。研究流程框架见表1。

Table 1. Research framework based on LDA topic model

表1. 基于LDA主题模型的研究框架

阶段序号	阶段名称	核心操作
1	文本收集	中国知网(CNKI)检索，下载并整理目标文献
2	文本预处理	jieba库分词、去除停用词、完成词频统计
3	模型参数优化	gensim库调整参数，用c-v一致性得分确定最优主题数
4	LDA主题建模	执行LDA主题建模，“主题-特征词”“文档-主题”“主题强度”分析
5	结果分析与结论	开展主题识别/强度演化分析，识别前沿热点、形成预测导向

3. 生态话语研究主题识别与演化分析

3.1. 核心主题识别

主题数 $K=7$ 时 $c-v$ 一致性得分达到峰值，模型最优，见图 2。7 个核心主题及特征词、占比见表 2。

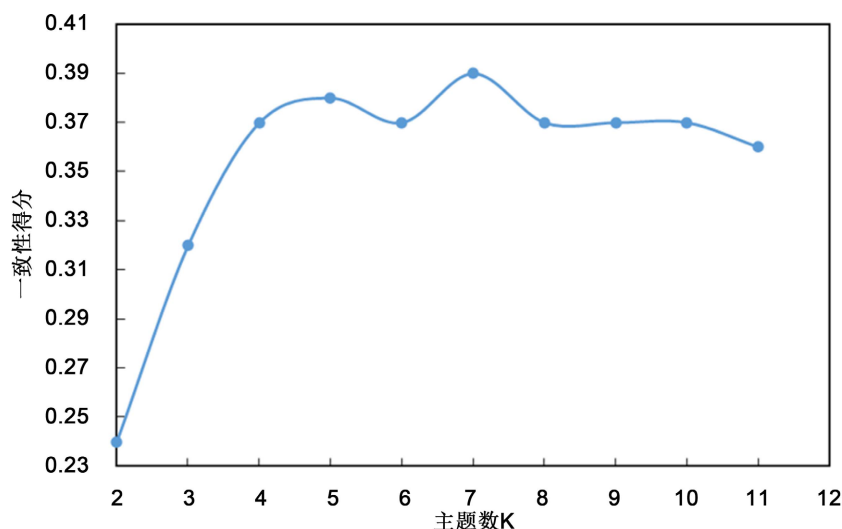


Figure 2. $c-v$ consistency score
图 2. $c-v$ 一致性得分

Table 2. Themes and keywords in ecological discourse research from 2016 to 2025

表 2. 2016~2025 生态话语研究主题 - 高频特征词

主题序号	主题概括	前 10 个高频特征词	整体占比%
1	多模态生态话语	意义、海洋、生命体参与者、新能源、隐喻、汽车、新闻、演讲、环境保护、气候变化、多模态	8.05
2	生态文明话语	中国、国际、发展、生态文明、能源、建构、和谐、绿色、体系、系统	13.00
3	生态语法	生命体参与者、自然、及物性系统、及物性、纪录片、动物、生态思想、物质过程、关系过程	20.43
4	评价理论	资源、态度资源、新闻话语、态度、新闻报道、媒体、评价理论、评价资源、情感	23.53
5	企业生态话语	企业、报告、资源、责任、态度、态度资源、可持续发展、生态意识、形象	13.00
6	生态哲学	自然、架构、及物性系统、动物、生命体参与者参与者、生态思想、及物性、生态文明	8.05
7	生态教育	教材话语、语篇、教师、学生、英语教材、人与自然生命共同体、及物性系统、态度资源	13.93

如表 2 所示，各主题占比排序为：评价理论(23.53%) > 及物性系统(20.43%) > 生态教育(13.93%) > 生态文明话语/企业生态话语(13.00%) > 多模态生态话语/生态哲学(8.05%)。从整体分布来看，生态话语研究领域已形成稳定的主题格局：领域已形成稳定主题格局：评价理论为核心研究方向，生态语法为重

要方法论支撑,生态教育、生态文明话语、企业生态话语为次要方向,多模态生态话语、生态哲学为补充性研究。

3.2. 主题演化三阶段

为精准分析生态话语主题演化的动态特征,依据 LDA 模型挖掘结果,整理得到 2016~2025 年文档 - 主题分布数据,分为三个阶段,明确各年份主题分布、主导方向及研究规模,见表 3。结合“年份 - 文档范围”对应关系及各文档主题概率数据,2016~2024 年各主题强度统计结果及变化趋势如下,见表 4。

Table 3. Document-Topic distribution from 2016 to 2025

表 3. 2016~2025 年文档 - 主题分布

年份	文档范围	文档数量	主题	主导主题
2016	文档 324	1	7	7 (100%)
2017	文档 320~323	4	6	6 (100%)
2018	文 311~319	9	1、3、4、5、6、7	1、3、4 (各 22%)
2019	文档 289~310	22	1、2、3、4、5、6、7	2 (27%)
2020	文档 268~288	21	2、3、4、5、6、7	2 (33%)
2021	文档 219~267	49	2、3、4、5、6、7	3 (31%)
2022	文档 170~218	49	1、2、3、4、5、6、7	3 (24%)
2023	文档 111~169	59	1、2、3、4、5、6、7	2 (31%)
2024	文档 53~110	58	1、2、3、4、5、6、7	2 (28%)
2025	文档 1~52	52	1、2、3、4、5、6、7	4 (35%)

Table 4. Statistical data of theme intensity from 2016 to 2025

表 4. 2016~2025 年主题强度统计

年份	主题 1 (多模态生态话语)	主题 2 (生态文明话语)	主题 3 (生态语法)	主题 4 (评价理论)	主题 5 (企业生态话语)	主题 6 (生态哲学)	主题 7 (生态教育)
2016	0.00	0.00	0.00	0.00	0.00	0.00	1.00
2017	0.00	0.00	0.43	0.00	0.00	0.81	0.45
2018	0.71	0.00	0.88	1.00	0.58	0.00	0.80
2019	0.79	0.72	1.00	0.90	0.00	0.88	0.84
2020	0.68	0.75	0.92	0.75	0.85	0.72	0.86
2021	0.71	0.68	0.73	0.78	0.75	0.61	0.86
2022	0.79	0.73	0.88	0.85	0.84	0.76	0.79
2023	0.75	0.64	0.76	0.82	0.72	0.88	0.69
2024	0.89	0.67	0.85	0.78	0.83	1.00	0.85

首先,在萌芽探索期(2016~2018 年),该时期文献总量仅 14 篇,研究主题呈现单点聚焦特征。2016 年仅 1 篇文献,集中在主题 7 生态教育;2017 年 4 篇文献全部聚焦主题 6 生态哲学,偏向理论思辨;2018 年主题首次延伸,主题 1、3、4、5、7 同步出现,研究边界快速拓展,多领域研究方向初步显现。

第二,在多元发展期(2019~2022年),该时期文献量快速增长,研究重心呈现动态轮换特征,2019~2020年主题2生态文明话语成为研究热点,契合生态文明发展需求;2021~2022年主题3生态语法取代主题2成为核心,研究回归生态语言学本体,及物性系统与生态语法分析占据主导。

第三,聚焦成熟期(2023~2025年),该时期文献量169占持续高位,主题4实现绝对主导且占比逐年攀升:2023年32.2%至2024年31.0%再至2025年34.6%。主题3稳居次席但占比下降(20.3%~20.7%~15.4%),显示研究从语法结构分析“向态度资源建构”的深度转型。主题7保持稳健发展(13%~15%),体现生态教育研究的持续性。此阶段形成了“评价理论为核心、及物性系统为支撑、生态文明与教育为两翼”的稳定格局。

4. 顺应论视角下生态话语研究

基于维索尔伦(Verschueren)的顺应论框架,生态话语研究的主题演化呈现四维顺应特征。

4.1. 语言结构策略的顺应

生态话语研究的主题分布与高频词聚类,充分体现对语言结构的顺应。结合表1主题-核心高频词数据可知,生态话语在语言结构层面通过精准的词汇选择、语义聚类与语体适配,主动完成对语言内部系统规则的主动顺应。

生态语法与生态哲学主题以生命体参与者、及物性系统、及物性、自然、动物等为核心高频词,聚焦句法编码规则与语义建构逻辑,其中生态语法主题的物质过程、关系过程等高频词,进一步强化了句法层面的适配性;评价理论主题聚焦态度资源、情感、评价资源、新闻话语等核心要素,严格遵循评价语义的层级结构与表达规范,实现评价性语言表达的系统性适配;多模态生态话语主题围绕隐喻、新闻、演讲、海洋、新能源、气候变化等高频词,顺应隐喻认知修辞规律与复合型语篇的语言组织结构,同时兼顾生态议题的核心语义;企业生态话语的企业、责任、可持续发展等高频词契合商务话语规范,生态教育话语的教材话语、教师、学生、英语教材等高频词适配教育话语逻辑。从主题占比分布来看,评价理论与生态语法主题占比居前,凸显评价语义结构与生态句法结构是生态话语语言结构顺应的核心维度,其余主题作为重要补充,共同构成层次完整、逻辑严谨的生态话语语言结构顺应体系,充分彰显了生态话语在语言结构选择上的适配性、系统性与自觉性。

4.2. 动态过程的顺应

主题的阶段性演化、更迭是对学科发展与社会需求的动态顺应。萌芽期顺应学科建制需求,以教育话语为切入点;发展期顺应社会发展,主题2占比骤升;成熟期顺应2020年提出的“双碳”目标与生态文明建设深化需求,研究重心转向对企业话语与媒体态度的精细分析。这种“教育话语”到“政策话语”再到“企业/媒体话语”的重心转移,展现了生态话语研究持续适配学术发展与实践需求的动态顺应过程。这一历时演化过程中,主题的更迭、主导主题的转移以及文档数量的变化,清晰呈现了生态话语在动态发展中不断调试表达策略、协商话语意义的过程,完全契合顺应论动态顺应的核心要求。

4.3. 语境的顺应

生态话语研究的语用策略精准顺应物理世界、心理世界、社会世界三维语境,实现了生态话语对外部语境的选择性顺应。在物理世界层面,研究场景从教材、课堂等固定教育场景,拓展至新闻媒体、企业报告等多元现实场景,顺应生态话语传播的客观时空与物质载体需求。在社交世界层面,多模态与生态哲学主题的低占比,契合国内应用语言学重实证、轻哲学思辨的学科文化规范。在心智世界语境层面,企业生态话语主题在2020年后保持稳定,顺应了“双碳”目标下公众对企业环境责任的认知需求与社会

关切，贴合社会主体的生态心智需求。

4.4. 意识程度的顺应

各主题强度占比和强度演化可量化表征生态话语元语用意识的凸显程度，见表4。从主题整体强度占比来看，占比越高，表明对应话语主体的交际意图越鲜明、元语用意识越凸显：评价理论主题占比最高，凸显话语主体对生态态度表达、价值评价的高度重视，元语用意识最为强烈；生态语法主题占比次之，体现对生态话语句法规范、语义建构的自觉关注，意识凸显程度较高；生态文明、企业生态、生态教育主题占比均衡，反映出对社会生态治理、企业责任、生态教育等领域的重点关注，意识凸显程度中等；多模态生态话语、生态哲学主题占比最低，表明其作为补充性主题，话语主体的相关意图相对隐蔽，意识凸显程度较低。

从年度演化来看，不同年份主导主题的变化进一步印证了意识凸显的动态差异：2016~2017年单一主导主题的凸显，体现当时话语主体对生态教育、生态哲学的集中关注；2019~2020年、2023~2024年主导主题生态文明的凸显，反映话语主体对社会生态治理的意识强化；2021~2022年主导主题生态语法的凸显，体现对生态话语规范建构的意识提升；2025年主导主题评价理论的凸显，彰显对生态价值评价的意识升级。这种主题强度占比的差异分布与年度演化特征，清晰呈现了生态话语主体元语用意识的层级分化与动态变化，与顺应论意识凸显维度的核心内涵高度契合。

5. 结语

5.1. 主要结论

研究基于LDA主题模型，系统识别2016~2025年国内生态话语研究主题并分析演化规律，初步厘清该领域的研究脉络与演化特征。主要结论如下：

第一，主题结构呈“双核引领、多元并进”的格局。评价理论与生态语法合计占比超四成，初步形成以评价理论为核心、及物性系统为重要支撑的研究范式，辅以生态文明传播、生态教育、企业生态话语等多个分支研究方向。

第二，主题演化轨迹可划分为萌芽探索、多元发展、聚焦成熟三个阶段。整体呈现单点萌芽至多元竞争再至范式初步定型的较为清晰的发展脉络。研究重心从早期的生态教育、哲学思辨逐步转向生态文明传播、生态语法分析，最终定型于评价理论视角的生态态度资源分析。

第三，生态话语研究的主题演化与顺应论逻辑基本契合。在语言结构、动态过程、语境与意识程度四大维度实现多维顺应，这一演化在一定程度上印证了国内生态话语研究从理论探索到实证探究、从宏观思辨到微观分析的成熟发展路径，但这种顺应性仍需更多跨样本、跨方法的研究进一步验证。

5.2. 研究不足和展望

本研究的结论受到研究方法与样本的限制，存在以下三个方面的不足：第一，研究方法存在固有局限，传统LDA模型内在缺陷导致难以精准处理跨主题融合类文献，导致部分融合及物性与评价理论的文献被单一归类，无法准确反映国内生态话语研究中“理论融合”的真实研究现状；第二，语料样本选取存在局限，仅单一数据库来源为分析对象，未纳入其他核心期刊、学位论文及外文相关文献，缺乏跨平台、跨语种、跨载体的语料对比，结论的普适性与代表性不足；第三，研究侧重文献计量与主题梳理，未结合具体的生态话语实践场景开展案例分析，导致应用性有所欠缺。

针对本研究的不足，结合国内生态话语研究现状，未来研究可从以下三方面开展更具针对性、可操作性的探索。

一是优化研究方法, 弥补模型固有缺陷。针对 LDA 模型跨主题识别不准的问题, 可引入动态主题模型(DTM)与贝叶斯网络模型相结合的方式, 一方面通过 DTM 捕捉主题热度与研究重心的时序连续演变, 另一方面通过贝叶斯网络模型优化跨主题文献的归类精度, 重点解决“及物性 - 评价理论”交叉研究的精准识别问题, 进一步细化生态话语研究的主题分类体系。

二是拓宽语料来源, 增强结论普适性。针对单一数据库的局限, 未来可补充国内核心期刊、学位论文及外文权威期刊的相关文本, 构建多语种、多载体的生态话语语料库。

三是强化实证导向, 推动理论与实践结合。针对缺乏实证案例的不足, 未来可聚焦具体实践场景, 选取典型生态话语文本如上市公司环境报告等开展个案分析, 解读实践中的生态话语建构策略; 同时, 深化生态哲学与多模态话语分析的融合应用, 结合视觉、听觉等多模态符号, 为生态话语的实践应用提供具体指导。

参考文献

- [1] Stibbe, A. (2020) *Ecological Discourse Analysis*. Cambridge University Press.
- [2] Verschueren, J. (1999) *Understanding Pragmatics*. Arnold.
- [3] 程淑佳, 刘琳琳, 夏昉. 基于 LDA 主题模型的我国职业教育研究主题演进分析[J]. 职业技术教育, 2025, 46(36): 64-71.
- [4] 段志敏. 及物性系统国内近十年研究综述[J]. 海外英语, 2024(16): 58-60.
- [5] 姜萍. 从顺应论看学术论文微信标题的语用策略——以社会学 C 刊为例[J]. 海外英语, 2025(13): 5-8.
- [6] 姜威, 于琦, 于彤, 等. 基于 LDA 模型的中医药新闻主题识别及其热点演化分析[J]. 中国中医药图书情报杂志, 2025, 49(6): 112-117.
- [7] 金媛. 基于顺应论的河南省旅游景区公示语翻译的生态构建[J]. 河北旅游职业学院学报, 2015, 20(2): 55-57.
- [8] 李雪梅. 语言顺应论视角下医学生传播中国抗疫故事的策略研究[J]. 海外英语, 2025(17): 37-41.
- [9] 李战子, 韩泽婷. 近十年 CSSCI 期刊社交媒体话语研究现状及热点趋势分析[J]. 外国语(上海外国语大学学报), 2025, 48(5): 102-111.
- [10] 尚永欣. 从顺应论看贵州毕节方言中“一大啷”的语用演变[J]. 中原文学, 2025(14): 183-185.
- [11] 王伊蕾, 冯德正. 深度学习技术在生态话语分析中的应用——基于 BERTopic 模型的弹幕生态主题分析[J]. 外语与外语教学, 2025(6): 19-29+144-145.
- [12] 吴恺. 近十年我国研究生教育研究领域的研究概况与主题分析——基于教育学 CSSCI 来源期刊数据[J]. 西南民族大学学报(人文社科版), 2020, 41(8): 234-240.
- [13] 张家缙, 王延敏, 赵玲. 基于 LDA-Word2vec 国内草原生态补偿研究主题演化与热点主题识别[J/OL]. 草业科学, 2025. <https://link.cnki.net/urlid/62.1069.S.20251211.1029.004>, 2026-04-20.
- [14] 赵紫琦, 亢琦. 基于 LDA 主题模型的 AI 赋能智慧图书馆研究热点分析[J]. 情报探索, 2025(12): 119-126.