

《计量语言学学报》的计量特征分析

——基于2007~2024年发表文章的研究

崔 超

内蒙古农业大学外国语学院, 内蒙古 呼和浩特

收稿日期: 2025年6月2日; 录用日期: 2025年7月31日; 发布日期: 2025年8月14日

摘 要

本研究运用文献计量学方法, 对《计量语言学学报》2007~2024年发表的342篇文章的标题、摘要及Scopus引文信息展开分析。结果显示, 期刊研究主题集中在语言、文本、英语、词汇和汉语等方面, 计量研究方法占主导。研究趋势上, 文本分析热度下降, 依存和句法分析成近年热点。中国研究人员数量显著增加, 居该领域领先地位, 论文作者和机构分布呈现学科多元化与国际化趋势。

关键词

文献计量学, 语料库分析, Scopus数据库, 研究趋势

Analysis of the Quantitative Characteristics of the Journal of Quantitative Linguistics

—A Study Based on Articles Published from 2007 to 2024

Chao Cui

College of Foreign Languages, Inner Mongolia Agricultural University, Hohhot Inner Mongolia

Received: Jun. 2nd, 2025; accepted: Jul. 31st, 2025; published: Aug. 14th, 2025

Abstract

This study uses bibliometric methods to analyze 342 articles published in the Journal of Quantitative Linguistics from 2007 to 2024, including their titles, abstracts, and citation information from Scopus. The results show that the journal's research topics focus on language, text, English, vocabulary, and Chinese, with quantitative research methods dominating. In terms of research trends, the popularity of text analysis has declined, while dependency and syntactic analysis have become hotspots in recent years. The number of researchers from China has increased significantly, taking

a leading position in this field, and the distribution of authors and institutions of the papers shows a trend of disciplinary diversification and internationalization.

Keywords

Bibliometrics, Corpus Analysis, Scopus Database, Research Trends

Copyright © 2025 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

《计量语言学学报》作为国际计量语言学学会会刊,自2007年起在国际学术界地位重要。文献计量法在语言学研究中的应用广泛[1][2],众多学者借此多维度探索该领域。在跨学科交叉、各分支学科及不同地域与语种研究方面均有成果[3],但部分研究存在范围窄、对新兴与交叉领域分析不深入等局限[4][5]。

鉴于此,对《计量语言学学报》进行文献计量分析意义重大。当前对该专门期刊的分析缺乏,通过分析可系统掌握计量语言学领域研究热点等发展态势,助力发现研究薄弱环节[6]-[8],完善学科理论方法,加强学术交流合作,还能为研究者选题和学术评价提供参考。

本研究旨在用文献计量学方法,分析该期刊2007~2024年342篇文章的计量特征,包括研究人员、研究对象等方面,利用Scopus数据库分析引文信息,为该领域发展提供参考。

2. 研究方法

本研究采用文献计量学方法[9]-[11],对《计量语言学学报》2007~2024年342篇文章的标题和摘要深入分析。

构建标题和摘要语料库,对其中实词进行词频和邻近词分析,按频率排序揭示高频词汇和邻近词。

利用Scopus数据库分析期刊引文信息,包括CiteScore等指标。

分析研究论文的国家地区、作者和机构分布,以及研究对象、方法和目的等计量特征。

3. 研究过程

3.1. 语料库构建与分析

下载2007~2024年所有期刊,构建标题和摘要语料库,分为2007~2009、2010~2012等6个时间段。

对标题语料库分析,17年中高频词有“of”“the”“language”“text”等,反映研究主题;高频邻近词有“analysis of”“of the”等,揭示研究方法和目的。

对摘要语料库分析,2007~24年高频词汇如“the”“of”“and”等,高频相邻词汇如“of the”“in the”等。

3.2. 研究结果

分别对六个时间段的标题和摘要语料库的高频词、高频相邻词进行统计分析,结果如下表所示(部分)。

3.3. Scopus 检索

通过“JQL-only”和“JQL-inclusive”在Scopus检索,了解整体发展趋势。2022年该期刊CiteScore为2.4,最高百分位数85.0%等,结果见表1~3。

Table 1. High-frequency words and adjacent words in different time periods
表 1. 不同时间段的高频词及高频相邻词

时间段	标题高频词	标题高频相邻词	摘要高频词	摘要高频相邻词
2007~2009	of、the、and 等	in the、of the 等	the、of、a 等	of the、in the 等
2010~2012	of、the、a 等	analysis of、in the 等	the、of、and 等	of the、in the 等

Table 2. Top 18 adjacent words
表 2. 排名前 18 临近词

Type	Rank	Freq	Range	NormFreq	NormRange
analysis of	1	26	6	7256.489	1
of the	2	25	6	6977.393	1
in the	3	20	6	5581.915	1
distribution of	4	14	6	3907.34	1
based on	5	13	6	3628.244	1
on the	5	13	5	3628.244	0.833
authorship attribution	7	11	6	3070.053	1
word length	7	11	5	3070.053	0.833
altmann law	9	10	6	2790.957	1
menzerath altmann	9	10	6	2790.957	1
on a	9	10	5	2790.957	0.833
s law	9	10	4	2790.957	0.667
study of	9	10	6	2790.957	1
to the	14	9	5	2511.862	0.833
zipf s	14	9	5	2511.862	0.833
a quantitative	16	8	3	2232.766	0.5
of text	16	8	5	2232.766	0.833
a statistical	18	7	3	1953.67	0.5
evaluation of	18	7	3	1953.67	0.5
in language	18	7	4	1953.67	0.667
of speech	18	7	5	1953.67	0.833
probability distribution	18	7	5	1953.67	0.833
properties of	18	7	4	1953.67	0.667
study on	18	7	3	1953.67	0.5
the menzerath	18	7	6	1953.67	1

Table 3. Analysis of Scopus citation information in 2022
表 3. 2022 年 Scopus 引文信息分析

CiteScore	最高百分位数	2019~2022 引文	2019~2022 文献	%被引用	SNIP	SJR
2.4	85.0% 144/1001 Language and Linguistics	169	71	63	1.137	0.596

4. 分析和讨论

4.1. 总体介绍

2007~2024 年，学报发表文章类型多样，其中论文占比 84.73%，书评占 9.94%，风格接近理工科期刊，注重实证研究。

各阶段论文数量总体上升，第三和第六阶段最多，第四和第五阶段受疫情等影响降幅明显，疫情后第六阶段反弹显著。

内容分析显示，研究围绕语言和文本现象，以计量研究为主，符合期刊发文范围和宗旨。

文章长度差异大，分类为 article 的文章长度均值为 18.52 页。

从表 4 结果可见，词频前 100 的关键词中，涉及语言本体和认识的词，排名前三的是 language、text 和 word。

从邻近词看，计量语言学注重模型、统计等，具有跨学科性质。

Table 4. Top 100 words in terms of word frequency that involve language ontology and recognition
表 4. 词频前 100 中涉及到语言本体和认识的单词

单词
language
text
english
word
chinese
length
languages
texts
linguistic
speech
lexical
linguistics
words
discourse
structure
phonetic
sentence
vocabulary
grapheme
japanese
phonological

其中，排名前三的是 language，text 和 word，即语言、文本和词汇。

应证科勒所言“研究目标和认识论意义而言，计量语言学和其他语言学分支并无大的区别。”其中，词汇为最多涉及语言结构，其次是语音。如果将关键词连起来，则得到一个陈述“英汉：(言语的语音，语言的文本词汇、句子、义素)*长度。”

4.2. 研究论文的标题、摘要分析

4.2.1. 标题分析

表 5 标题聚类结果中高频实词显示词汇研究是 2007~2024 年计量语言学主要对象，分布特征是主要计量特征。文本分析热度呈下降态势，疫情期间曾逆向上扬。

Table 5. Title clustering (2007~2024)
表 5. 标题聚类(2007~2024 年)

Cluster	Rank	Freq	Range	NormFreq	NormRange
length distribution in	1	3	2	0.071	0.333
for word length	2	2	2	0.048	0.333
of sentence length	2	2	1	0.048	0.167
length an entropy	4	1	1	0.024	0.167
length and clause	4	1	1	0.024	0.167
length and frequency	4	1	1	0.024	0.167
length distribution based	4	1	1	0.024	0.167
length distribution model	4	1	1	0.024	0.167
length distributions based	4	1	1	0.024	0.167
length empirical evidence	4	1	1	0.024	0.167
length entropies and	4	1	1	0.024	0.167
length for reliable	4	1	1	0.024	0.167
length frequencies for	4	1	1	0.024	0.167
length frequencies in	4	1	1	0.024	0.167
length frequency profiles	4	1	1	0.024	0.167
length in academic	4	1	1	0.024	0.167
length in chinese	4	1	1	0.024	0.167
length in spoken	4	1	1	0.024	0.167
length minimization in	4	1	1	0.024	0.167
length position and	4	1	1	0.024	0.167
length vocabulary size	4	1	1	0.024	0.167
and clause length	4	1	1	0.024	0.167
and text length	4	1	1	0.024	0.167
and word length	4	1	1	0.024	0.167
between the length	4	1	1	0.024	0.167
chinese dependency length	4	1	1	0.024	0.167

续表

function word length	4	1	1	0.024	0.167
linguistics text length	4	1	1	0.024	0.167
logic word length	4	1	1	0.024	0.167
measure word length	4	1	1	0.024	0.167
minimum document length	4	1	1	0.024	0.167
of word length	4	1	1	0.024	0.167
on sentence length	4	1	1	0.024	0.167
polish word length	4	1	1	0.024	0.167
practice word length	4	1	1	0.024	0.167
stochastic word length	4	1	1	0.024	0.167
the clause length	4	1	1	0.024	0.167
vocabulary word length	4	1	1	0.024	0.167

4.2.2. 摘要分析

摘要中高频词和近义词特征与标题基本相同。计量语言学在一定程度上遵循自然科学科研方法论，但适用程度需进一步研究。语料库是其研究基础。各阶段有新的分析视角，依存和句法在第五阶段成热点。表 6 显示树库使用在 2013~2015 年最频繁，阶段划分方式影响趋势呈现。

Table 6. Abstract clustering (2007~2024)
表 6. 摘要聚类(2007~2024 年)

Cluster	Rank	Freq	Range	NormFreq	NormRange
word length distribution	1	5	2	0.089	0.333
of the distribution	2	4	4	0.071	0.667
distribution of dependency	3	2	2	0.036	0.333
the probability distribution	3	2	1	0.036	0.167
distribution are they	5	1	1	0.018	0.167
distribution based on	5	1	1	0.018	0.167
distribution classification analysis	5	1	1	0.018	0.167
distribution in german	5	1	1	0.018	0.167
distribution in mongolian	5	1	1	0.018	0.167
distribution in scientific	5	1	1	0.018	0.167
distribution in us	5	1	1	0.018	0.167
distribution in zhuang	5	1	1	0.018	0.167
distribution interrelation and	5	1	1	0.018	0.167
distribution is valid	5	1	1	0.018	0.167
distribution model probability	5	1	1	0.018	0.167
distribution models in	5	1	1	0.018	0.167

续表

distribution of dependencies	5	1	1	0.018	0.167
distribution of discourse	5	1	1	0.018	0.167
distribution of elemental	5	1	1	0.018	0.167
distribution of initial	5	1	1	0.018	0.167
distribution of parts	5	1	1	0.018	0.167
distribution of phonetic	5	1	1	0.018	0.167
distribution of represented	5	1	1	0.018	0.167
distribution of sentence	5	1	1	0.018	0.167
distribution of sortal	5	1	1	0.018	0.167
distribution of textual	5	1	1	0.018	0.167
distribution of three	5	1	1	0.018	0.167
distribution of variants	5	1	1	0.018	0.167
distribution on the	5	1	1	0.018	0.167
distribution sometimes explode	5	1	1	0.018	0.167
characteristics and distribution	5	1	1	0.018	0.167
greek probability distribution	5	1	1	0.018	0.167
guide probability distribution	5	1	1	0.018	0.167
hierarchy depth distribution	5	1	1	0.018	0.167
in the distribution	5	1	1	0.018	0.167
its probability distribution	5	1	1	0.018	0.167
jun probability distribution	5	1	1	0.018	0.167
language probability distribution	5	1	1	0.018	0.167
letter frequency distribution	5	1	1	0.018	0.167
load diachronic distribution	5	1	1	0.018	0.167
mandelbrot s distribution	5	1	1	0.018	0.167
of discrete distribution	5	1	1	0.018	0.167
on the distribution	5	1	1	0.018	0.167
probability word distribution	5	1	1	0.018	0.167
the bose distribution	5	1	1	0.018	0.167
the exponential distribution	5	1	1	0.018	0.167
zipf mandelbrot distribution	5	1	1	0.018	0.167

4.2.3. 国家地区分布

342 篇文章中, 中国作者文章数量最多, 达 163 次, 美国 64 次, 中国从 1994~2013 年的第七位升至第一位。

4.2.4. 作者分布

表 7 作者分布中雷蕾、Macutek jan 等较为突出, 计量语言学呈现强者林立态势。单从两个字的名字

排名看来，我们中国的作者占了前五位中的两个，其余是斯洛伐克，印度和西班牙。即亚洲和欧洲。如果进一步详细分析，应该可以看到阿尔特曼学派的团队及其学生们的名字，而且，可以大胆推测：最近十几年的发展，计量语言学已经是强者林立的状态，不再是某个团队的一枝独秀。这说明计量语言学已经在世界生根发芽开花结果，之前的所有专家学者及其团队的努力都成功了，而且在继续扩大。

Table 7. Distribution of authors
表 7. 作者分布

Type	Rank	Freq
Lei lei	1	8
Macutek jan	1	8
Hemlata dhami	3	6
Liu haitao	3	6
Pande hemlata	3	6
Cancho ramon	6	5
Ferrer i	6	5
I cancho	6	5
Dhami h	9	4
Jiang jingyang	9	4
Kelih emmerich	9	4
Qu yunhua	9	4
Andres jan	14	3
Cech radek	14	3
Chen ruina	14	3
Fan fengxiang	14	3
Haoran lei	14	3

4.2.5. 机构分布

从表 8 可见浙江大学发表文章最多，达 157 篇，浙江外国语学院 37 篇，具有一定地域和机构集中性。

4.3. 书评

Table 8. Book review statistics
表 8. 书评统计

Collected work	
语音	Spanish phonology and morphology. Experimental and quantitative perspectives
方法论	Exact methods in the study of language and text. Dedicated to Gabriel Altmann on the occasion of his 75th birthday
统计	Statistics in language research: Analysis of variance
言语	The Czech Language in the Present-Day Private Correspondence. Letters, e-mails, SMS
记忆	Memory-based language processing

续表

文本言语	Evaluation of Text and Speech Systems
语料库	The Changing Face of Corpus Linguistics
书面语	Analyses of Script: Properties of Characters and Writing Systems
计量方法	History of the application of quantitative methods in the Russian language and literature
多模态语料库	Evaluation in Media Discourse Analysis of a Newspaper Corpus
计量方法	Quantitative Methods in Linguistics.
文本和语言	Text and Language
计量方法	Methods and Applications of Quantitative Linguistics.
统计方法	Statistical Methods in Language and Linguistic Research.
文本	Forms and Degrees of Repetition in Texts-Detection and Analysis
语言表层和深层结构	Language Classificatory Surface Pattern and Deep Structures in the lombardo-venetian dialect region and in the Raeto Romania
计量分析	Quantitative Analysis of Poetic Texts
R 软件分析	Text Analysis with R for Students of Literature
并列结构	English Coordinate Constructions. A Processing Perspective on Constituent Order
语料库聚类	Cluster Analysis for Corpus Linguistics
多元分析	Multi-Dimensional Analysis, 25 years on—A tribute to Douglas Biber
语言文本序列	Sequences in Language and Text
认知	Cognitive Linguistics: The Quantitative Turn
计量方法	Maths Meets Myths: Quantitative Approaches to Ancient Narratives
标注	Handbook of Linguistic Annotation
主题	Motifs in Language and Text
语料库英汉对比	Corpus-Based Contrastive Studies of English and Chinese
统计学方法	Statistics in Corpus Linguistics: A Practical Guide
计量分析依存句法	Quantitative Analysis of Dependency Structures
计量方法俄语	Quantitative Approaches to the Russian Language
语料库文体	Corpus Stylistics: Theory and Practice
词汇和计量	Words and Numbers. In Memory of Peter Grzybek (1957~2019)
计量方法	Quantitative Approaches to Universality and Individuality in Language
计量方法中文	Quantitative Studies on Vocabulary and Syntax (in Chinese)
依存句法	Dependency Structure from Syntax to Discourse. A Corpus Study of Journalistic English

- 1) 2007~2009: 西班牙语语音, 捷克语, 统计, 记忆, 语料库
- 2) 2010~2012: 书写系统, 计量*2, 语料库, 文本
- 3) 2013~2015: 计量, 统计
- 4) 2016~2018: 语料库, 多元分析, 序列, 计量*2, 标注, 语料库
- 5) 2019~2021: 语料库, 计量依存, 计量, 语料库文体

6) 2022~2024: 词汇, 计量*2, 依存。第一阶段涉及欧洲小语种, 统计方法论, 语言认知, 语料库; 其中, 语料库在第二段, 第四段和第五阶段出现频率上升为每个阶段两次; 从第二阶段到第六阶段, 计量平均每年出现两次, 算是传统; 文本, 词汇, 标注, 语音, 语篇, 依存句法等内容, 间接点缀每一年。

因此, 计量, 统计和语料库是频率最高的书评主题。因此可以认为, 计量语言学中, 这三件法宝是必须重点掌握的生命线。

4.4. Scopus 数据库引文信息

仅《学报》的 2022 年 CiteScore 为 2.4, 最高百分位数 85.0%等, 在领域内学术地位较高, 且 CiteScore 总体呈上升趋势。包含《学报》的检索中, 其在相关领域具有独特性。

5. 研究讨论

本研究发现《计量语言学学报》在研究对象、方法和目的上有明显计量特征, 国际影响力较大。中国在作者、机构分布上表现突出, 具有一定地域和机构集中性。

但研究存在局限性, 仅基于标题和摘要分析, 且仅关注该期刊, 未来研究可扩大范围以获得更全面认识。

参考文献

- [1] Lei, L. and Liu, D. (2019) Research Trends in Applied Linguistics from 2005 to 2016: A Bibliometric Analysis and Its Implications. *Applied Linguistics*, **40**, 540-561. <https://doi.org/10.1093/applin/amy003>
- [2] Lei, L. and Liu, D. (2019) The Research Trends and Contributions of System's Publications over the Past Four Decades (1973-2017): A Bibliometric Analysis. *System*, **80**, 1-13. <https://doi.org/10.1016/j.system.2018.10.003>
- [3] 单谊. 基于文献计量的国外实验语用学研究态势探析[J]. 外国语文, 2024, 40(2): 24-37.
- [4] Liu, Z. (2022) Mapping the Research Trends of Third Language Acquisition: A Bibliometric Analysis Based on Scopus. *Frontiers in Psychology*, **13**, Article 1021517. <https://doi.org/10.3389/fpsyg.2022.1021517>
- [5] Birkle, C., Pendlebury, D.A., Schnell, J. and Adams, J. (2020) Web of Science as a Data Source for Research on Scientific and Scholarly Activity. *Quantitative Science Studies*, **1**, 363-376. https://doi.org/10.1162/qss_a_00018
- [6] 袁国廷. 基于文献计量和形势概念分析语言学发展态势研究[J]. 宜春学院学报, 2019, 41(5): 73-80.
- [7] Georgas, H. and Cullars, J. (2005) A Citation Study of the Characteristics of the Linguistics Literature. *College & Research Libraries*, **66**. <https://doi.org/10.5860/crl.66.6.496>
- [8] Ellegaard, O. and Wallin, J.A. (2015) The Bibliometric Analysis of Scholarly Production: How Great Is the Impact? *Scientometrics*, **105**, 1809-1831. <https://doi.org/10.1007/s11192-015-1645-z>
- [9] CheshmehSohrabi, M. and Mashhadi, A. (2023) Using Data Mining, Text Mining, and Bibliometric Techniques to the Research Trends and Gaps in the Field of Language and Linguistics. *Journal of Psycholinguistic Research*, **52**, 607-630. <https://doi.org/10.1007/s10936-022-09911-6>
- [10] He, S. and Shaari, A.H. (2025) Hot Spots and Emerging Trends in 'Event' Research in Linguistics (2001-2023): A Bibliometric Analysis. *Cogent Arts & Humanities*, **12**, Article 2488086. <https://doi.org/10.1080/23311983.2025.2488086>
- [11] Pearson, W.S. (2021) Quoted Speech in Linguistics Research Article Titles: Patterns of Use and Effects on Citations. *Scientometrics*, **126**, 3421-3442. <https://doi.org/10.1007/s11192-020-03827-5>