

Analysis of the Scientific Knowledge Graph of Corpus-Based Discourse Analysis

Zongbo Zhang¹, Xinyu Hu¹, Shuyi Zhang¹, Hongxin Li^{1,2}

¹PLA Strategic Support Force Information Engineering University, Luoyang Henan

²State Key Laboratory of Cryptology, Beijing

Email: zhangzongbo09@163.com

Received: May 31st, 2019; accepted: Jun. 18th, 2019; published: Jun. 27th, 2019

Abstract

Scientific knowledge graphs which cover the literature of corpus-based discourse analysis in the past 28 years were generated in this paper with the help of a bibliometric tool, a piece of software named CiteSpace, and with the source data (1988~2018 (March)) from Web of Science database. The major knowledge graph employed in this paper is the co-citation graph of corpus-based discourse analysis generated by CiteSpace. According to this graph, we found 9 major literature clusters covering the main research aspects in the field of corpus-based discourse analysis and we have analyzed these 9 clusters in detail. Important articles or books in each cluster were analyzed in this process. Furthermore, we also analyzed those key points in the graph, which had high citation frequency, high centrality value, or high citation burstness value. Last but not least, we tried to explore potential research hotspot by using the timeline function of CiteSpace.

Keywords

Discourse Analysis, Corpus, CiteSpace, Knowledge Graph

基于语料库话语分析的科学知识图谱分析

张宗波¹, 胡鑫煜¹, 张书轶¹, 李宏欣^{1,2}

¹战略支援部队信息工程大学, 河南 洛阳

²密码科学技术国家重点实验室, 北京

Email: zhangzongbo09@163.com

收稿日期: 2019年5月31日; 录用日期: 2019年6月18日; 发布日期: 2019年6月27日

摘要

本文借助CiteSpace文献计量工具,以Web of Science数据库1988~2018年(3月)收录数据为来源,绘制基于语料库的话语分析近28年来的科学知识图谱。利用CiteSpace软件,我们绘制了基于语料库的话语分析领域的文献共被引图谱,根据图谱的聚类结果,对9个较大的文献聚类进行了深入分析,涵盖了基于语料库的话语分析的主要方向。此外,我们还对知识图谱中具有高被引、高中介中心性和高突变性的关键节点文献进行了分析,梳理了基于语料库的话语分析各个方向的重要文献,并结合图谱的时间线功能分析了潜在的研究热点。

关键词

话语分析, 语料库, CiteSpace, 知识图谱

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

话语分析自上世纪中叶兴起以来已历经半个多世纪的发展。如今的话语分析已经不再是语言学的专利,而是拓展到了政治学、社会学、传播学、哲学、历史学乃至自然学科等诸多学科领域[1]。然而,传统的话语分析方法被一些学者指出具有示范性分析的语料较少、较零散,分析缺乏全面性、代表性和完整性等局限[2] [3] [4] [5] [6]。话语研究界对此不乏回应和辩驳[7] [8] [9] [10],但也有学者尝试为话语分析寻求新路径。在此背景下,“以大量精心采集而来的真实文本为研究素材,主要通过概率统计等方法得出结论”[11]的语料库语言学在近年来受到了许多话语研究学者的青睐。在国际上,Hardt-Mantner、Stubbs、Baker [12] [13] [14] [15] [16]等学者率先垂范,倡导将语料库的方法与话语分析结合起来。在国内,也有不少学者从学理和方法的层面对运用语料库语言学进行话语分析做了论证和实践[17] [18] [19] [20] [21]。近年来,国内外基于语料库的话语分析成果不断涌现,可以说,将语料库语言学的方法运用于话语分析已经形成一种潮流。对此,学界不乏梳理和评述,但均从定性的角度入手。本文拟依托 Web of Science 检索平台数据,借助可视化分析工具,绘制基于语料库话语分析的知识图谱,利用定量和定性相结合的手段剖析该领域主要研究方向、重要节点文献和潜在研究热点。

2. 研究方法

2.1. 研究工具

科学知识图谱,是显示科学知识的发展进程与结构关系的一种图形,属于科学计量学范畴[22]。在以往,为了了解某个学科领域发展的状况和规律,学者们需要人工查阅该领域内几乎所有的文献。而到了信息技术高速发展的今天,只需要借助科学知识图谱分析工具,就可以将学者们从海量的文献数据中解放出来,以图形的方式高效直观地展示学科领域发展的整体状况和演进规律。绘制科学知识图谱的方法和工具比较多样。主要的数据来源包括 Web of Science (WoS)、CSSCI、CNKI、CSCD 等;分析的方法主要有共现分析、共(被)引分析、多元统计分析等;常用的绘制工具有 CiteSpace、Ucinet (Netdraw)、Pajek、

SPSS、VOSviewer、Bibexcel、Bicomb 等等[23]。

本文采用的主要分析工具是美国德雷塞尔大学(Drexel University)陈超美教授开发的 CiteSpace 软件。该软件是一款功能强大的引文可视化分析软件,通过对软件节点类型的设置,可以绘制不同类型的知识图谱,主要包含四大类:合作网络分析(作者、机构或国家的合作网络分析),共现分析(主题、关键词或 Web of Science 类别的共现分析),共被引分析(文献、作者和期刊的共被引分析),以及文献的耦合分析[24][25]。根据不同的研究问题,我们可以采用不同图谱绘制方法。

2.2. 数据来源

本研究利用如下检索策略在 Web of Science 平台进行检索:

“(TS = ((discursive OR discourse OR CDA OR critical lingu*) AND (corpus OR corpora))) AND 语种: (English)精炼依据: Web of Science 类别: (LINGUISTICS OR LANGUAGE LINGUISTICS) AND 文献类型: (ARTICLE)索引 = SCI-EXPANDED, SSCI, A & HCI, ESCI 时间跨度 = 所有年份”。

即,标题、关键词、摘要中不仅包含话语研究的字样(discursive 或 discourse 或 CDA 或 critical lingu*),还包含语料库字样(corpus 或 corpora);文献语种限定为英语;研究领域限定为 Web of Science 类别中的 Linguistics 或 Language Linguistics;文献类型只选取论文类(Article);检索数据库范围为 WoS 核心数据集 (SCI-EXPANDED, SSCI, A & HCI, ESCI);时间跨度为所有年份(实际上,笔者所使用的 WoS 端口只提供 1988~2018 年的数据)。检索日期为 2018 年 3 月 31 日,得到检索结果 1591 条。

借鉴陈超美[26]教授的做法,笔者对检索结果进行了扩展。通过 WoS 平台的“创建引文报告”功能,可以查看引用了这 1591 条结果的文献,即施引文献,共 6825 条。由于施引文献可能还包含 1591 条结果中的文献(即自引文献),我们选取去除自引的施引文献以避免数据重复,得到结果 6097 条。显然,1591 条结果是基于语料库的话语分析相关度较高的文献,而 6097 条去除自引的施引文献,由于引用了 1591 条结果中的文献,我们也将其归属到与基于语料库的话语分析相关的文献。所以本文所分析的文献数据实际有 7688 条。将这 7668 条文献的数据以纯文本格式批量下载,并利用 CiteSpace 的数据去重功能处理后,最终得到 7429 条独立结果,这 7429 条结果的年度分布情况如图 1 所示,主要来源出版物如表 1 所示。

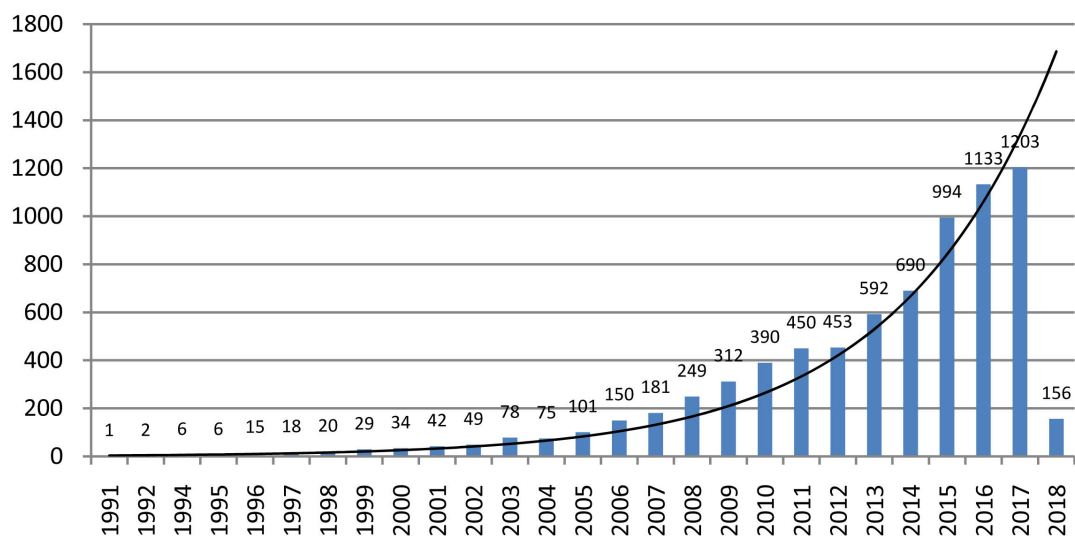


Figure 1. Each year's publishing amount of the data collected from WoS

图 1. WoS 检索结果年度发文量统计图

Table 1. Top 30 sources of the data collected from WoS
表 1. WoS 载文量前 30 位的来源出版物

序号	载文量	来源出版物名称	2017 年 影响因子	近 5 年 影响因子	Linguistics 类别 中的排序 ¹
1	545	<i>Journal of Pragmatics</i>	1.039	1.521	71/181
2	191	<i>English for Specific Purposes</i>	1.362	1.829	47/181
3	146	<i>Journal of English for Academic Purposes</i>	1.42	2.093	42/181
4	112	<i>International Journal of Corpus Linguistics</i>	0.976	1.475	75/181
5	106	<i>Lingua</i>	0.864	1.101	87/181
6	97	<i>Applied Linguistics</i>	<u>3.225</u>	<u>3.899</u>	<u>3/181</u>
7	90	<i>Linguistics</i>	0.644	0.878	106/181
8	77	<i>Discourse Studies</i>	1.016	1.601	54/84 ²
9	76	<i>System</i>	1.547	1.826	36/181;
10	76	<i>Iberica</i>	0.694	0.521	103/181
11	73	<i>Pragmatics</i>	0.5	0.587	121/181
12	72	<i>Language Sciences</i>	0.832	0.847	91/181
13	72	<i>Cognitive Linguistics</i>	1.902	2.297	23/181
14	65	<i>Computational Linguistics</i>	1.319	2.202	50/181
15	59	<i>English</i>	无数据	无数据	无数据 ³
16	56	<i>Language</i>	<u>2.34</u>	<u>2.966</u>	<u>9/181</u>
17	53	<i>Language Learning</i>	1.655	3.297	32/181
18	52	<i>World Englishes</i>	0.75	1.081	99/181
19	52	<i>Journal of Language and Politics</i>	0.69	0.631	104/181
20	50	<i>Intercultural Pragmatics</i>	1.125	1.154	65/181
21	50	<i>Corpus Linguistics and Linguistic Theory</i>	1.2	1.386	60/181
22	48	<i>Journal of Memory and Language</i>	<u>2.829</u>	<u>5.148</u>	<u>5/181</u>
23	47	<i>Journal of Second Language Writing</i>	<u>3.324</u>	<u>4.137</u>	<u>2/181</u>
24	46	<i>Metaphor and Symbol</i>	0.862	1.373	88/181
25	44	<i>Tesol Quarterly</i>	<u>2.256</u>	<u>2.828</u>	<u>10/181</u>
26	44	<i>Pragmatics and Society</i>	0.34	0.576	147/181
27	43	<i>Modern Language Journal</i>	<u>2.789</u>	<u>2.578</u>	<u>6/181</u>
28	43	<i>Language and Literature</i>	0.525	0.602	118/181
29	43	<i>Journal of Historical Pragmatics</i>	1.063	0.56	68/181
30	43	<i>Folia Linguistica</i>	0.324	0.616	149/181

¹数据来自第 2017 版 Journal Citation Reports (JCR), 下划线部分为 Linguistics 类别中的排序排名前十的来源出版物。

²期刊 *Discourse Studies* 在 JCR 的类别为 Communication。

³在 Web of Science 中未查询到期刊 *English* 的影响因子数据。

不难看出, 国外基于语料库的话语分析兴起于二十世纪九十年代, 此后呈快速发展之势, 2013 年至今更是该领域成果激增的一个时期(图 1)。该领域发展至今, 研究成果广泛见于国际语言学界主流期刊(表 1), 也说明基于语料库的话语分析已经受到语言学界的认可, 成为语言学的一个重要方向。

3. 结果与分析

图 2 是 CiteSpace 5.3.R2 的主页面截屏, 主要的参数选取如图所示。本研究对 Time Slicing 和 Node Types 中参数进行了相应的设置, 其余参数选择默认值。Node Types, 也就是节点类型, 不同的节点类型对应着不同的图谱绘制方法。界面提供了 4 种颜色, 13 种节点。其中, 4 种颜色对应前文介绍的 CiteSpace 提供的 4 种类型的图谱。本文选取的节点是 Reference, 即文献共被引分析。这是 CiteSpace 被最广泛应用的功能。我们可以通过文献共被引图谱中的关键节点、聚类 and 色彩来分析某个研究主题的发展演变[27]。

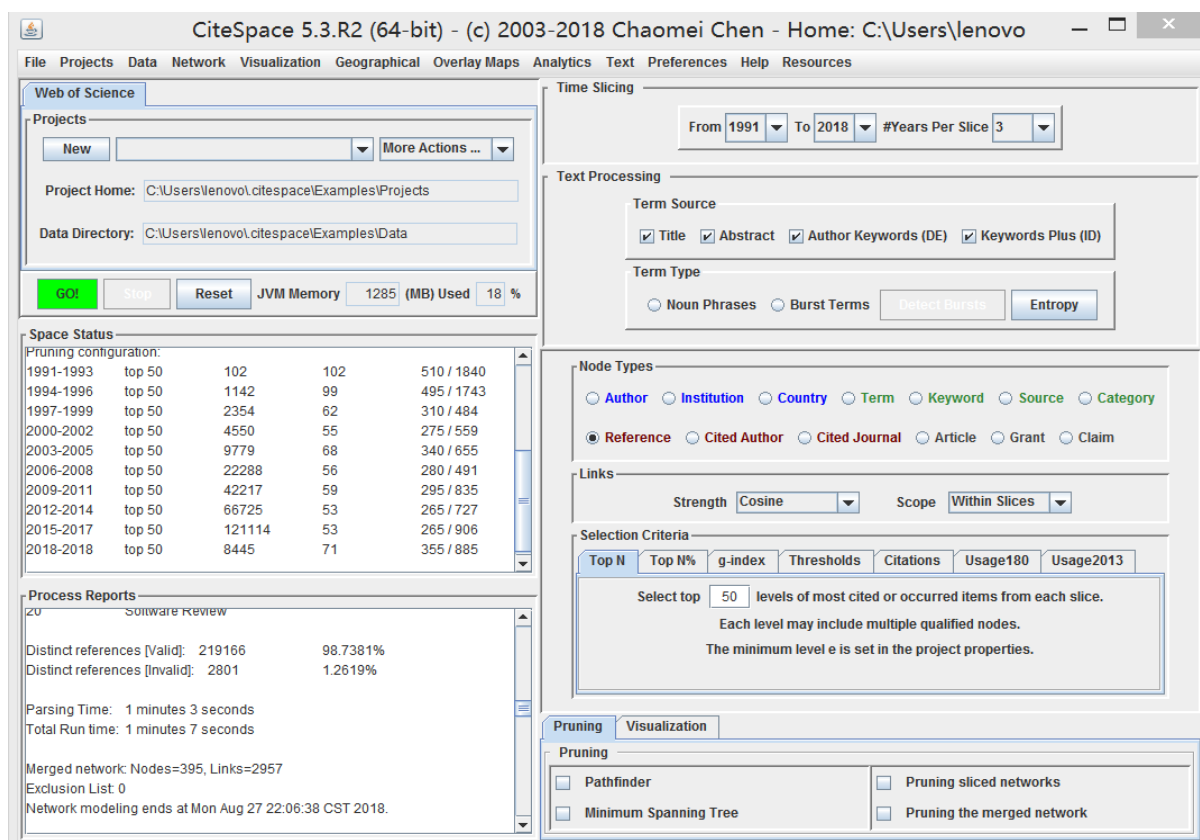


Figure 2. The interface of CiteSpace

图 2. CiteSpace 主界面

根据图 1, 本研究数据集中所采集到的文献年份跨度为 1991~2018 年, 因此我们在 Time Slicing 中对年份跨度进行了相应的设置。可以看到, 在 Time Slicing 中还有一个 Years Per Slice 的参数。顾名思义, CiteSpace 会对 1991~2018 年间的数据库按时间段进行切片分析, 而切片的时间跨度就是 Years Per Slice。如陈超美教授所言, “CiteSpace 的功能类似一架照相机, 只是它拍摄的对象是科学文献而不是自然景色。设置 CiteSpace 的各项参数大致相当于取景、调焦、对光圈。” [28] Years Per Slice 就像是相机的曝光时间, 想要获得效果较好的图谱, 就必须对这一参数进行调试。经过反复尝试, 在本研究中, 将这一参数设置为 3 时可以获得较好的效果。

3.1. 文献聚类分析

图3为利用CiteSpace可视化功能绘制的基于语料库话语分析的文献共被引知识图谱。图3顶部的色块为1991~2018以3年为单位切分的时间段，由左往右颜色由冷色调向暖色调转变，越偏向冷色调的时间越早，反之则越近。图中每个节点代表一篇文献，节点之间的连线代表文献之间存在共被引关系，连线的颜色与图顶色块对应，表示引用发生的时间。节点周围的圆圈实际是文献的被引年轮，可以反映文献的被引频次，圆圈越大被引频次越高，反之越低；圆圈内的颜色与图顶色块对应，表示文献被引的时间，深色在内圈，亮色在外圈。有些圆圈外围带有紫圈，这可以反映该节点在图谱中的中介中心性值(Betweenness Centrality)的大小。文献的聚类之间往往有一些起到桥梁作用的文献，这样的文献通常具有较高的中介中心性值，通过对这种桥梁文献的分析，我们可以管窥领域研究的发展延伸脉络。还有些圈内有鲜红色区域，这表示该文献的引用突变度(Citation Burstness)。引用突变度反应的是某文献在某个时段内突然被大量引用的程度，这样的文献可以反映出某一时段的领域研究热点。鲜红色区域的大小反映的是引用突变度的大小，鲜红色区域在引用年轮内的位置则反映出文献被引激增的时段。每个节点旁的黑色字体标签显示的是该文献的作者和发表年份。为了使图谱清晰可辨，图3只显示了被引频次高于200次的文献标签。

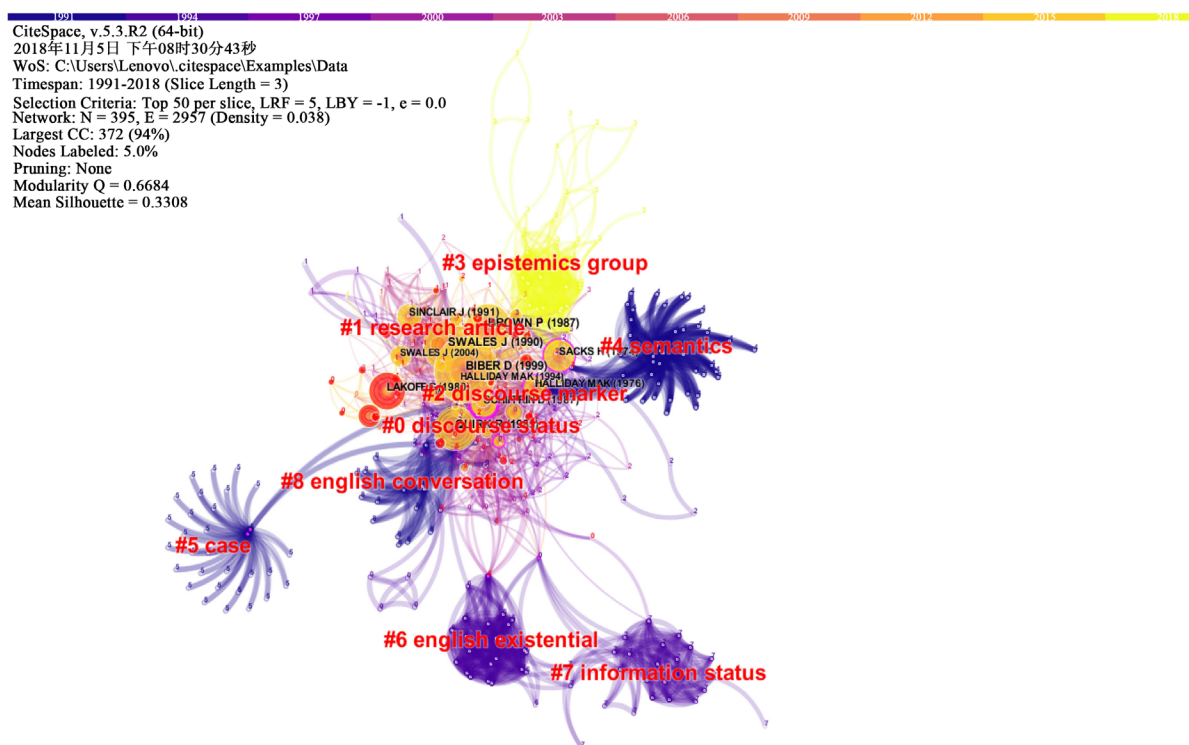


Figure 3. Co-citation knowledge graph

图3. 文献共被引分析图谱

利用CiteSpace的聚类分析功能，我们还可以对文献共被引分析图谱中的文献进行聚类。同一个聚类中的文献引用关系较为密切，可以被看作具有相近的研究主题。图3展示了通过聚类分析得到的9个较大的聚类，分别被标记为聚类#0~#8，聚类0到8所包含的文献数量依次递减。聚类分析后每个节点也会被赋予一个数字，代表该节点所属的聚类编号。每个聚类都被自动赋予了一个红色字体的标签。CiteSpace

提供了多种标签生成的算法和方式。算法包括 LSI、LLD、MI 三种，其中 LLD 为软件的默认算法；分析的对象可以是文献的标题、关键词和摘要。文献的标题往往是该文献研究内容、方向、方法的高度浓缩，较之关键词信息更全，较之摘要更简洁。综合考量之下，本研究利用软件默认的 LLD 算法对文献标题进行分析，得到图 3 所示的聚类标签。聚类标签可以在一定程度上反映聚类的研究方向和特点，但要深入了解各个聚类，还需根据各个聚类的实际情况作进一步分析。表 2 是 9 大聚类的主要参数结果，表中数据与图 3 对应。聚类大小表示该聚类所包含的文献数量；Silhouette 值为聚类的轮廓值，反映聚类效果，该值越接近于 1 则聚类效果越好；文献平均年份反映该聚类文献发表年份的均值。

Table 2. Literature clustering result

表 2. 文献聚类分析结果

聚类编号	聚类大小	Silhouette 值	文献平均年份	聚类标签
0	62	0.585	1993	discourse status
1	60	0.796	1997	research article
2	59	0.671	1990	discourse marker
3	47	0.959	1996	epistemics group
4	44	0.994	1982	semantics
5	31	0.998	1972	case
6	26	0.984	1983	English existential
7	25	0.958	1984	information status
8	18	0.977	1981	English conversation

聚类 0 的标签是“话语状态”(discourse status)。这是 9 个聚类中规模最大的一个，包含 60 篇文献，许多高被引文献都出现在这一聚类中。表 3 列出了聚类 0 中被引频次⁴超过 100 的文献，备注一栏是对各篇文献的简要描述。不难看出，语料库的研究方法在多篇文献中都有直接的体现。最具代表性的当属 Quirk *et al.* (1985) [29] 和 Huddleston *et al.* (2002) [30]，这两部英语语法词典均是基于语料库进行词典编纂的典范，在学界具有广泛影响。Chafe (1994)、Goldberg (2006)、Arnold *et al.* (2000) [31] [32] [33] 将语料库语言学与意识、构式语法等结合起来，对话语中的信息单位和信息结构进行了定量与定性相结合的研究。另外，不难发现，认知学派的文献在聚类 0 的高被引文献中占较高比重。如 Lakeoff (1980)、Lakeoff (1987)、Langacker (1987) [34] [35] [36] 都是认知语言学的奠基之作。构式语法虽然已是一门独立的研究流派，但其来源于认知语法，本质上还是可以归入认知语言学范畴。表 3 中的 Goldberg (1995) [37]、Goldberg (2006) [32] 都是构式语法的重要著作。此外，该聚类被引频次第 3 位的 Pragglejaz Group (2007) [38] 聚焦隐喻研究，提出了根据词汇进行隐喻识别的步骤；第 7 位的 Chafe (1994) [31] 探究心理意识与语言中信息单位的关系；第 14 位的 Gundel *et al.* (1993) [39] 研究话语指代的认知状态和形式；这些也都可以归属到认知语言学的研究范畴。除了认知语言学范畴的文献外，表 3 中也不乏系统功能语言学。例如被引频次第 11 的第一版《功能语法导论》[40] 和第 12 位的 Lambrecht (1994) [41]。前者在功能语言学派的地位不言而喻，后者则是将功能语法框架下进行的语言信息结构研究总结起来。我们可以大致推测聚类 0 的特点：1) 利用语料库方法与认知语言学、功能语言学等理论相结合进行定量与定性相结合的话语研究；2) 话语中的隐喻、话语中的信息结构及状态等是研究的热点方向。

⁴需要指出的是，表中文献的被引频次为该文献在本研究收集的数据集中的被引频次，而并非该文献在学术共同体中的实际被引频次，后同。

Table 3. Literature with high citation frequency in cluster 0**表 3.** 聚类 0 中的高被引文献

序号	被引频次	文献
1	417	Quirk, R., Sidney, G., and Leech, G., and Svartvik, J. (1985) <i>A Comprehensive Grammar of the English Language</i> [M]. Longman, London and New York, 1985.
2	323	Lakoff, G. and Johnson, M. (1980) <i>Metaphors We Live By</i> . University of Chicago Press, Chicago.
3	194	Pragglejaz Group ⁵ . (2007) MIP: A Method for Identifying Metaphorically Used Words in Discourse. <i>Metaphor & Symbol</i> , 22, 1-39.
4	192	Goldberg, A. E. (1995) <i>Constructions: A Construction Grammar Approach to Argument Structure</i> . University of Chicago Press, Chicago.
5	184	Lakoff, G. (1987) <i>Women, Fire, and Dangerous Things: What Categories Reveal About the Mind</i> . Chicago University Press, Chicago.
6	175	Huddleston, R. and Pullum, G. K. (2002) <i>The Cambridge Grammar of the English Language</i> . Cambridge University Press, Cambridge.
7	166	Chafe, W. L. (1994) <i>Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing</i> . University of Chicago Press, Chicago, IL, US.
8	141	Langacker, R. W. (1987) <i>Foundations of Cognitive Grammar: Theoretical Prerequisites</i> . Stanford University Press, Stanford, California.
9	141	Tomassello, M. (2013) <i>Constructing a Language: A Usage-based Theory of Language Acquisition</i> [M]. Harvard University Press, Cambridge, Massachusetts, and London, England.
10	138	Goldberg, A. (2006) <i>Constructions at Work: The Nature of Generalization in Language</i> [M]. Oxford University Press, Oxford.
11	133	Halliday, M. A. K. (1985) <i>An Introduction to Functional Grammar</i> (1st ed.). Arnold, London.
12	129	Lambrecht, K. (1994) <i>Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents</i> . Cambridge University Press, Cambridge.
13	120	Arnold, J. E., Losongco, A., Wasow, T., and Ginstrom, R. (2000) Heaviness vs. Newness: The effects of Structural Complexity and Discourse Status on Constituent Ordering. <i>Language</i> , 76, 28-55.
14	107	Gundel, J. K., Hedberg, N., and Zacharski, R. (1993) Cognitive Status and the Form of Referring Expressions in Discourse. <i>Language</i> , 69, 274-307.

聚类 1 的标签是“研究论文”(research articles)。该聚类包含 60 篇文献, 规模仅次于聚类 0, 同样聚集了大量的被引文献。表 4 列出了聚类 1 中被引频次高于 100 的文献。从表 4 可以看出, 聚类 1 侧重于学术话语的研究, 这与该聚类的标签十分契合。表 4 中 Swales (1990), Swales (2004), Hyland (2000), Bhatia (1993), Coxhead (2000), Hyland (2005a) [42]-[47] 都是直接对学术文本进行的研究。体裁分析是一种重要的话语分析方法, 近年来在原有的体裁分析基础上, 更是发展出了批评性体裁分析(Critical Genre Analysis, CGA)。Swales 及其学生 Bhatia 都是体裁分析的代表人物, Swales 更是有“体裁分析之父”的美誉。他们侧重于特殊用途英语(English for Specific Purposes, ESP), 尤其是学术用途英语(English for Academic Purposes, EAP)方面的体裁研究。Swales (1990) [42] 一书中提出的 CARS (Create a Research Space) 分析模型不仅为学术英语写作教学提供了理论支撑, 还为学术英语的话语分析提供了重要的方法论。Swales (2004) [43] 是对 Swales (1990) [42] 的发展和补充, 对体裁分析理论重新进行了系统地阐述和梳理。Bhatia (1993) [45] 对商务、学术、法律等体裁进行了较为全面的探讨, 是利用体裁分析理论开展话语分析的重要文献。Hyland (2000) [44] 和 Hyland (2005) [47] 关注的是学术话语中作者和读者的互动。Hyland 利用语料库的方法对 8 个学科的学术文本进行了量化考察, 并对业内人士进行了访谈, 他提出的互动分析

⁵Pragglejaz Group 为团体作者, 具体包括以下学者: Crisp, P (Crisp, Peter); Gibbs, R (Gibbs, Raymond); Deignan, A (Deignan, Alice); Low, G (Low, Graham); Steen, G (Steen, Gerard); Cameron, L (Cameron, Lynne); Semino, E (Semino, Elena); Grady, J (Grady, Joe); Cienki, A (Cienki, Alan); Kovecses, Z (Kovecses, Zoltan)。其中, Crisp, P 为通讯作者。

模型为学术话语的研究添加了一个全新维度。Coxhead (2000) [46]利用语料库的手段从词汇层面对学术文本进行了研究, 建立了一个包含 570 个词族的新学术词表。该研究在建立语料库、建立词表、评估词表等方面为同类研究树立了方法论典范, 同时其研究成果对于学术英语教学和研究也有重要的参考意义。除了以上对学术话语进行直接研究的文献外, 表 4 还涉及到了语料库语言学、系统功能语言学和语用学的语言学理论。表 4 中还涉及到 Biber 和 Sinclair 两位语料库语言学大家的经典著作(Sinclair 1991; Biber et al, 1999; Biber, 1988; Biber, 2004) [48] [49] [50] [51], 系统功能语言学派的经典著作(Halliday & Hassan, 1976; Halliday, 1994; Halliday & Matthiessen, 2004) [52] [53] [54], 评价理论(Martin & White, 2005) [55], 以及语用学中的礼貌原则和面子理论(Brown & Levinson, 1987) [56]等。这些文献所包含的理论早已被广泛运用于话语分析之中。就学术话语而言, 它们虽未针对性地对学术话语进行研究, 但在学术话语研究领域均有极高的理论指导意义, 使得学术话语分析有法可守、有章可循。因此, 我们可以推断聚类 1 的主要研究方向是基于语料库的学术话语分析, 依托的主要理论框架有体裁分析、礼貌原则、语篇衔接、评价理论、词块分析等等。

Table 4. Literature with high citation frequency in Cluster 1

表 4. 聚类 1 中的高被引文献

序号	被引频次	文献
1	532	Biber, D., Johansson, S. and Leech, G., <i>et al.</i> (1999) Longman Grammar of Spoken and Written English. Pearson Education, Essex, England.
2	461	Swales, J. (1990) <i>Genre Analysis: English in Academic Settings</i> . Cambridge University Press, Cambridge.
3	436	Brown, P. and Levinson, S. C. (1987) <i>Politeness: Some Universals in Language Usage</i> . Cambridge University Press, Cambridge, UK.
4	275	Halliday, M. A. K. and Hasan, R. (1976) <i>Cohesion in English</i> . Longman Group, London.
5	246	Sinclair, J. (1991) <i>Corpus, Concordance, Collocation</i> . Oxford University Press, Oxford.
6	230	Halliday, M. A. K. (1994) <i>An Introduction to Functional Grammar</i> (2nd ed.). Arnold, London.
7	210	Swales J. (2004) <i>Research Genres: Explorations and Applications</i> . Cambridge University Press, Cambridge.
8	186	Hyland, K. (2000) <i>Disciplinary Discourses: Social Interactions in Academic Writing</i> . Pearson Education Limited, Harlow, UK.
9	174	Biber, D. (1998) <i>Variation Across Speech and Writing</i> . Cambridge University Press, Cambridge.
10	161	Halliday, M. A. K. and C.M.I.M. Matthiessen. (2004) <i>An Introduction to Functional Grammar Third Edition</i> . Arnold, London.
11	145	Bhatia, V. (1993) <i>Analysing Genre: Language Use in Professional Settings</i> . Longman, London.
12	142	Coxhead, A. (2000) A New Academic Word List. <i>TESOL Quarterly</i> , 34, 213-238.
13	141	Martin, J.R. and White, P. R. (2005) <i>The Language of Evaluation: Appraisal in English</i> . Palgrave Macmillan, Basingstoke.
14	129	Biber, D., Conrad, S., and Cortes, V. (2004) <i>If you look at...: Lexical Bundles in University Teaching and Textbooks</i> . <i>Applied linguistics</i> , 25, 371-405.
15	102	Hyland, K. (2005) <i>Stance and Engagement: A Model of Interaction in Academic Discourse</i> . <i>Discourse Studies</i> , 7, 173-192.

聚类 2 的标签是“话语标记”(discourse marker)。该聚类共 59 篇文献, 被引频次大于 100 的文献如表 5 所示。从高被引文献来看, 与标签“话语标记”相关的只有被引频次最高的 Schiffrin (1987) [57]。这是一部话语标记语的专门研究著作, 通过访谈的方式对中低级英语水平的美籍犹太人如何使用 oh、well、or、but、so、because、now、then、y'know、I mean 等话语标记语进行了研究, 意在探究这些标记语在实

现语篇的连贯性和互动性中的作用。被引频次第二的是 Goffman (1981) [58]。其实 Goffman 最负盛名的应该是其 1974 年所著的《框架分析》一书, 但该著作中的框架分析理论更多地被应用于新闻传播领域的分析。Goffman (1981) [58] 将框架分析理论运用到语言研究中, 对口语会话交际过程背后的社会原则和框架进行了深入的讨论, 为后来的会话分析提供了理论支撑。被引频次第三的是 Grice (1975) [59], 该论文是会话逻辑的重要文献, Grice 在其中提出的合作原则和会话含义理论受到学界的广泛关注和认可, 合作原则已经成为了语用学领域最重要的基本原则之一。被引频次第四的是 Clark (1996) [60], 该著作聚焦于语言的使用, 利用口语语料库中即席对话的真实案例作为支撑, 其中心思想就是语言具有社会性, 人们在使用语言的时候总是处在一种相互联合的社会活动之中(joint activity), 语言的使用是一种联合行动(joint action)。被引频次第五和第六的分别是 Leech (1983) [61] 和 Levinson (1983) [62]。这两部著作都是语用学领域的经典教科书, 它们的出版是语用学成为语言学独立分支的重要标志之一, 往后虽也有语用学的教科书陆续出版, 但所涉及的基本原理和内容也难逃这两部经典的窠臼。综合来看, 聚类 2 带有浓厚的语用学色彩, 且重点关注真实的口语会话。

Table 5. Literature with high citation frequency in Cluster 2

表 5. 聚类 2 中的高被引文献

序号	被引频次	文献
1	285	Schiffrin, D. (1987) <i>Discourse Markers</i> . Cambridge University Press, New York.
2	168	Goffman, E. (1981) <i>Forms of Talk</i> . University of Pennsylvania Press, Philadelphia.
3	160	Grice, H. P. (1975) <i>Logic and Conversation</i> . In: Cole, P. and Morgan, J. (eds.) <i>Syntax and Semantics Volume 3: Speech Acts</i> . Academic Press, New York, 41-58.
4	128	Clark, H. H. (1996) <i>Using Language</i> . Cambridge University Press, Cambridge.
5	123	Leech, G. N. (1983) <i>Principles of Pragmatics</i> . Longman, London and New York.
6	121	Levinson, S. C. (1983) <i>Pragmatics (Cambridge Textbooks in Linguistics)</i> . Cambridge University Press, Cambridge/New York/Sidney.

聚类 3 的标签是 epistemics group, 包含 47 篇文献。仅看软件给出的标签, 较难判定该聚类的特点。从该聚类的高被引文献来看, 被引频次超过 100 的只有两篇, 即 Sacks *et al.* (1947) [63] 和 Sacks (1992) [64]。为了更好地反映聚类 3 的研究方向, 表 6 列出了聚类 3 中被引频次超过 50 的文献。Harvey Sacks, Emmanuel Schegloff 和 Gail Jefferson 三位学者被视作会话分析(conversation analysis)的先驱。聚类 3 中两篇被引频次过百的文献均为他们的代表作。Sacks *et al.* (1974) [63] 被认为是会话分析的奠基之作, 其中提出的话轮转换模型是口语会话研究的重要理论模型。不幸的是, 在该文章发表后的第二年, Sacks 便因车祸去世。之后, 由 Gail Jefferson 主持, 将 1964 年秋至 1972 年春期间 Sacks 的讲课内容整理成册并出版, 并由 Emmanuel Schegloff 主笔绪论。Sacks (1992) [64] 是整理出版的上卷, 收录了 Sacks 1964 年秋至 1968 年春期间的讲课内容, 下卷于 1995 年出版。Sacks 的两卷讲课集在会话分析领域有着举足轻重的地位, 对后世影响深远。表 6 中被引频次第三的 Pomernantz (1984) [65] 被收录于 Atkinson & Heritage (1984) [66] 中, 该书编者在绪论中便指出该书所收录的研究均从会话分析的视角展开, 而这一研究视角源自 Sacks 的讲课内容以及 Emmanuel Schegloff 和 Gail Jefferson 对会话分析理论的发展。事实上, 聚类 3 中被引频次第五的 Heritage (1984) [67], 虽然未被列入表 6, 但也被收录在 Atkinson & Heritage (1984) [66] 一书中。被引频次第四的 Jefferson (2004) [68] 是一篇专门讲述会话研究中转录工作的文章, 对转录所要考虑的因素、用到的符号等作了概要介绍, 这对于构建口语语料库在操作层面具有很强的指导意义。综上所述, 将聚类 3 的标签判定为会话分析应该更为妥当。

Table 6. Literature with high citation frequency in Cluster 3**表 6.** 聚类 3 中的高被引文献

序号	被引频次	文献
1	319	Sacks, H., Schegloff, E. A. and Jefferson, G. (1974) A Simplest Systematics for the Organization of Turn-Taking for Conversation. <i>Language</i> , 50, 696-735.
2	160	Sacks, H. (1992) <i>Lectures on Conversation</i> (Vol. 1). Edited by Gail Jefferson with Introductions by Emanuel A. Schegloff. Basil Blackwell, Oxford.
3	76	Pomerantz, A. (1984) Agreeing and Disagreeing with Assessments: Some Features of Preferred/Dispreferred Turn Shapes. In: Atkinson, J. M. and Heritage, J., Eds. <i>Structures of Social Action: Studies in Conversation Analysis</i> . Cambridge University Press, Cambridge, 57-101.
4	61	Jefferson, G. (2004) Glossary of Transcripts Symbols with an Introduction. In: Lerner, G.H., Ed. <i>Conversation Analysis: Studies from the First Generation</i> . John Benjamins Publishing, Amsterdam/Philadelphia, 13-31.

聚类 4 的标签是“语义学”(semantics)。该聚类包含了 44 篇文献,但是除了 Labov (1972) [69] 的被引频次达到了 119 次,其余文献的被引频次均为 1 次。Labov 是美国著名的语言学家,被视为社会语言学的创始人,在语言变化和方言研究方面有突出贡献。*Sociolinguistic Pattern* (Labov, 1972) [69] 一书是 Labov 在语言变化研究领域的代表著作,书中运用了 Labov 首创的田野调查模式,对不同阶层、不同年龄的语言使用进行了实证研究,向世人揭示了社会不同阶层的语言变体是可以被调查和分析的,开创了同类研究的先河。根据表 2,聚类 4 的 Silhouette 值高达 0.994,说明该聚类网络中的节点具有很高的同质性。因此,我们可以推测,聚类 4 的主要研究方向与社会语言学,尤其与语言变体和方言研究紧密相关。为了印证这一推测,笔者对该聚类的其他文献进行了考察。笔者发现,该聚类中比较活跃的学者是 William B. McGregor,他有 5 篇文献出现在该聚类,分别是 McGregor (1987a, 1987b, 1988, 1989, 1990) [70] [71] [72] [73] [74]。这 5 篇文献都是关于 Gooniyandi 语(澳大利亚西部南金伯利地区的方言)的研究,属于典型的方言研究,印证了上文的推测。社会语言学强调语言 and 实际社会的联系,这与批评话语分析将话语视作社会实践的理念不谋而合。社会语言学的研究常常把阶层、年龄、性别、地域等诸多因素纳入考虑,这对基于语料库的话语分析而言,在操作层面具有极强的指导意义。

聚类 5 的标签是“案例”(case)。该聚类总共包含 31 篇文献,被引频次普遍较低,被引频次最高的 *A Grammar of Contemporary English* (Quirk et al., 1972) [75] 也仅有 3 次,其余均为 1 次。Quirk et al. (1972) [75] 是 Quirk 等人编纂的第一部英语语法著作,该书在编纂理念上既继承传统语法的精神,又兼容并包地吸收了结构主义、生成语法等当代主要语法流派的思想,可谓是一部承上启下之作,对后世的语法著作有很强的启发意义。根据表 2,聚类 5 的 Silhouette 值高达 0.998,因此我们可以根据 Quirk et al. (1972) [75] 推测,该聚类的主要方向是语法研究。此外,该聚类还包含了 Jespersen 的两部著名语法著作, *Language: Its Nature, Development, and Origin* [76] 和 *The Philosophy of Language* [77], 也印证了该聚类方向是语法研究的推测。这类研究属于基础性的理论研究,可以为基于语料库的话语分析提供理论支撑。

聚类 6 的标签是“英语存在句”(English existential)。该聚类总共包含 26 篇文献,被引频次也普遍较低。被引频次最高的为(Hawkins, 1978) [78], 被引 9 次。这是一部专门研究定冠词与不定冠词的著作, Hawkins 从语义和语用两方面入手,希望建立一套统一的冠词指称理论。实际上,指称问题也是英语存在句研究中的重要内容。因此, CiteSpace 给聚类 6 贴上“英语存在句”的标签也是可以理解的。笔者对该聚类中其他文献的标题也做了考察,发现了较多关于英语存在句以及指称问题的研究,如: Lumsden (1988), Hawkins (1991), Freeze (1992), Erdmann (1976), Birner (1994) [79] [80] [81] [82] [83] 等等。再结合聚类 6 较高的 Silhouette 值(0.984),我们可以推测,该聚类的主要研究方向是英语存在句,以及英语中的指称问题。

聚类 7 的标签是“信息状态”(information status)。该聚类的 Silhouette 值为 0.958, 共包含 25 篇文献中, 被引频次最高的为 Ward (1988) [84] 和 Hoekstra & Mulder (1990) [85], 均为 5 次。Ward (1998) [84] 根据真实语料对英语前置结构的语用能力进行了系统的研究。Ward (1998) [84] 认为前人对前置结构的研究没有抓住其背后的语用学原理, 他指出要通过语义和语用特征来区分前置结构的类型, 并阐述了前置结构的两大语篇功能, 提出了一套前置结构的分类体系。Hoekstra & Mulder (1990) [85] 研究了非作格动词作系动词的现象, 以及这种现象与方位地点指示和存在意义表达之间的关系。比较有特色的是, 该研究的案例涉及到英语、法语、西班牙语、意大利语、荷兰语等多个语种, 理论成果具有一定的普遍性。结合该聚类的标签来看, 这两篇文献实际上都是对不同句子结构或成分所开展的研究, 而研究的落脚点正是这些语言单位承载信息的状态。再考虑到该聚类较高的 Silhouette 值, “信息状态”的标签是可信的。

聚类 8 的标签是“英语会话”(English conversation), 聚类大小为 18, Silhouette 值为 0.977。该聚类中被引频次相对较高的文献是 Brown & Yule (1983) [86] 和 Svartvik & Quirk (1980) [87], 被引频次分别为 15 次和 9 次, 其他文献的频次均为 1 次。Brown & Yule (1983) [86] 是一部系统阐述话语分析著作, 该书以传统语言学理论为根基, 博采社会语言学、心理语言学、计算语言学、语言哲学等众家之长, 重点围绕人类如何运用语言交流的问题展开讨论; 书中的核心思想就是要揭示说者(或作者)是如何为听者(或读者)构建语言信息, 而听者(或读者)又是如何解读这些信息的。Svartvik & Quirk (1980) [87] 其实是英语会话语料库的转录文本。该语料库是著名的 LLC 口语语料库(London-Lund Corpus of Spoken English)的一部分, 包含 34 篇面对面的对话转录文本, 共约 17 万词, 语料采集对象均为受教育程度在以上的英国人。LLC 语料库对学界有着重要影响, Quirk *et al.* (1972) [75] 中的例句就主要来自于该语料库。综上所述, 该聚类“英语会话”的标签是合理的。这与聚类 3 的“会话分析”有所重叠。造成这一问题的原因, 可能需要从该聚类的高引用文献上去寻找。(高引用文献指的是积极引用高被引文献的文献。)通过 CiteSpace 的 Cluster Explorer 功能可以发现, 该聚类的高引用文献只有 Geluykens (1991) [88], 该聚类的 18 篇文献在这篇文章的参考文献中均可找到, 也就是说该聚类中文献的共被引关系主要来自 Geluykens (1991) [88]。这篇文献是对英语会话中 it 分裂句语篇功能的研究, 与聚类 3 主要聚焦话轮转换的会话分析还是有区别的, 所以聚类 8 被单独列出也不足为怪。

3.2. 高中介中心性文献分析

图 4~7 为中介中心性最高的五篇文献在时间线图的高亮展示, 图中作者名前圆括号内的数值为对应文献的被引频次, 图标题方括号中为该文献的中介中心性值。拥有高中介中心性值的文献, 在文献共被引网络图中一般具有重要的桥梁作用, 在学科发展中往往具有引起研究范式转移的开拓性意义。图 4 反映的是中介中心性最高的文献 Schiffrin (1987) [57]。前面已经提到, 这是一部研究话语标记的著作, 而且是一部承前启后之作。话语标记的研究始于上世纪 50 年代, 但直到 80 年代才逐渐成为一个独立的语言学研究方向, 而推动这一转变的正是 Schiffrin 的这部 *Discourse Markers*。话语标记的研究发展至今, 已经成为语用学、话语分析等多个领域关注的重点。图 5 中的文献为 Halliday 与 Hasan 合著的 *Cohesion in English* 一书, 该书的出版标志着衔接理论的创立。衔接理论自问世以来就广受学界关注, 因为它对语篇分析、翻译理论、人机对话和文体研究都有重要作用, 尤其是为篇章结构的研究开创了新局面, 对语言学的发展具有卓越贡献[89]。图 6 为 Sacks 等 1974 年发表的论文“*A Simplest Systematics for the Organization of Turn-Taking for Conversation*” [64]。上文中也已提到, 这篇文章被认为是会话分析的开山之作, 文中提出的话轮转换模型被广泛应用于各类会话研究中。会话分析经过不断发展, 研究内容不断丰富、领域不断拓展, 其研究成果还被广泛应用于对法庭、课堂教学和访谈等机构话语的研究中[90]。图 7 为 Hawkins (1978) [78], 相对而言, 虽然该书的被引频次较低(9 次), 但其对于聚类 6 的研究而言却具有重要意义。

该书是较早对英语冠词的指称功能进行系统的理论梳理的著作。聚类 6 聚焦于存在句的指称问题，Hawkins (1978) [78]为其提供了重要的理论框架。图 8 为 Chafe (1994) [31]，该著作在一定程度上可以看作是 Chafe 对其一生学术思想的总结[91]。Chafe (1994) [31]在书中结合真实语料，重点探讨了语言信息与意识和时间之间的相互作用。书中对意识的阐释，对语言信息三分法(新信息、旧信息、可及信息)的思想，对口语和书面语中意识和时间作用模式的论述，都对后世该领域研究产生了深远影响(表 7)。

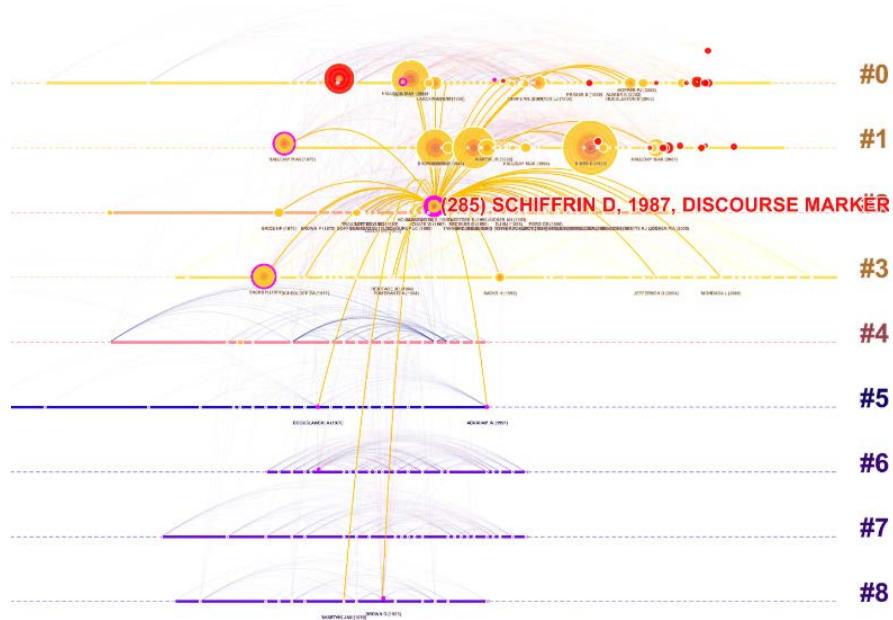


Figure 4. Schiffrin (1987) [0.29]
图 4. Schiffrin (1987) [0.29]

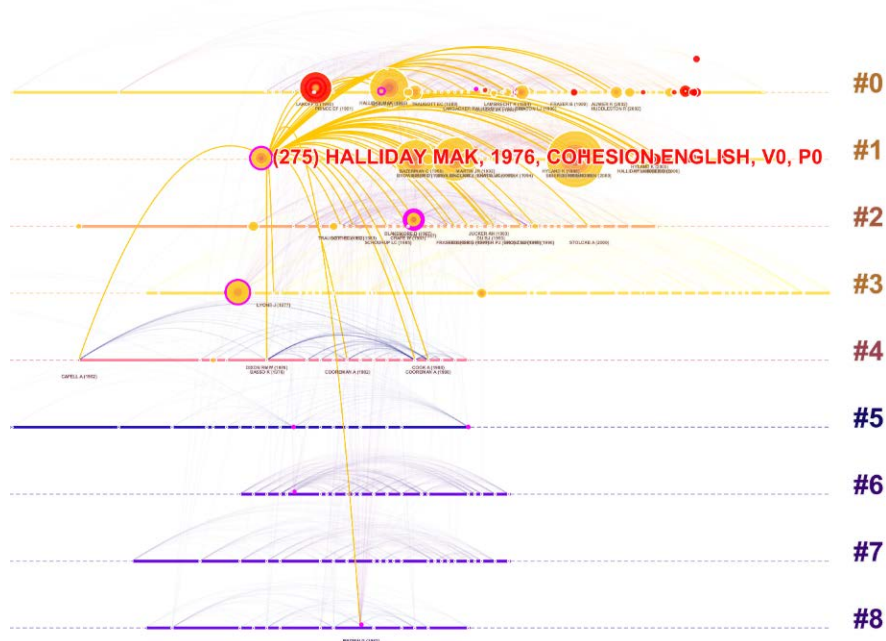


Figure 5. Halliday (1976) [0.18]
图 5. Halliday (1976) [0.18]

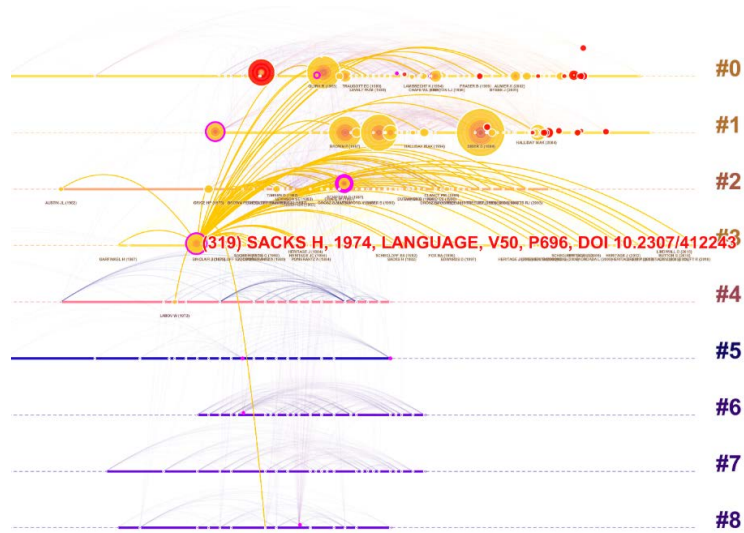


Figure 6. Sacks (1974) [0.17]

图 6. Sacks (1974) [0.17]

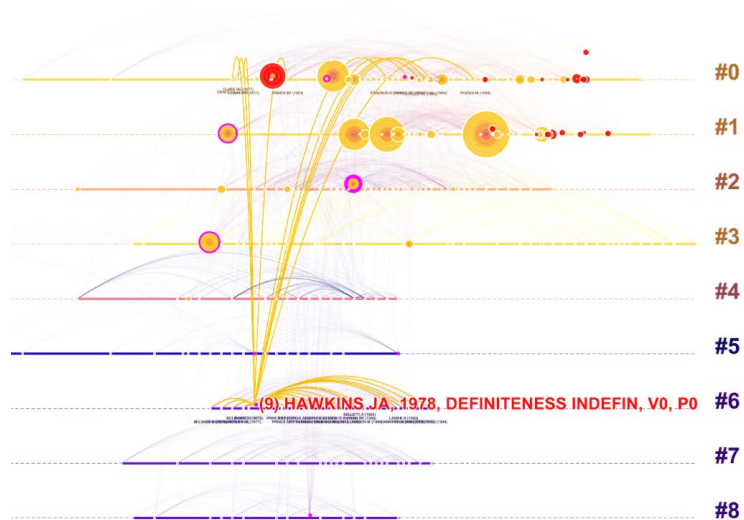


Figure 7. Hawkins (1978) [0.16]

图 7. Hawkins (1978) [0.16]

Table 7. Literature with high centrality value

表 7. 高中介中心性文献

序号	中介中心性	被引频次	所在聚类	文献
1	0.29	285	2	Schiffrin, D. (1987) Discourse Markers. Cambridge University Press, New York.
2	0.18	275	1	Halliday, M. A. K. and Hasan, R. (1976) Cohesion in English. Longman Group, London.
3	0.17	319	3	Sacks, H., Schegloff, E.A. and Jefferson, G. (1974) A Simplest Systematics for the Organization of Turn-Taking for Conversation. Language, 50, 696-735.
4	0.16	9	6	Hawkins, J. A. (1978) Definiteness and Indefiniteness. Humanities Press, Atlantic Highland, NJ.
5	0.15	166	0	Chafe, W. L. (1994) Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing. University of Chicago Press, Chicago, IL, US.

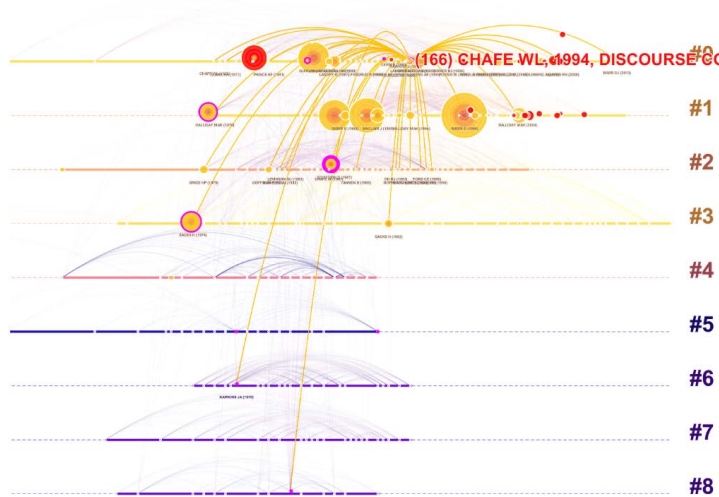


Figure 8. Chafe (1994) [0.15]

图 8. Chafe (1994) [0.15]

3.3. 潜在研究热点分析

利用 CiteSpace 的 Timeline View 功能, 可以得到如下的聚类时间线图, 该图可以较为直观地呈现各个聚类随时间的演变过程。可以看到, 关键节点文献主要集中于前四个聚类(#0~#3), 尤其是聚类 0 和聚类 1 中, 这说明这几个聚类在基于语料库的话语分析中居主导地位。从时间线演变来看, 前四个聚类还在不断发展当中。尤其是聚类 3, 其节点连线多为黄色, 说明该聚类中文献共被引发生的时间较近, 是一个新兴的研究方向。后五个聚类(#4~#8)在上世纪 90 年代后没有发生显著变化。结合上文总结的各聚类特点可以作如下小结, 在基于语料库的话语分析中, 利用语料库方法与认知语言学、系统功能语言学、语用学等理论相结合的研究范式较为普遍, 学术语篇和口语会话是两大比较受关注的话语类型, 口语会话分析是可能是新兴的热点方向。

由上观之, 图 9 中有不少文献的引用年轮是深红色的, 这表明这些文献具有较高的引用突变度。上文已经提到引用突变度反应的是某文献在某个时段内突然被大量引用的程度, 这样的文献可以反映出某一时段的领域研究热点。为了更直观地展示, 可以利用 CiteSpace 的 Burstness (突变度) 功能, 查看具有较高引用突变度的文献。在本文所涉及的所有文献中, 共检测到 126 篇高突变度文献, 突变值最高为 23.6161, 最低为 3.4704。图 10 是将这些文献按照引用突变时段排列后得到的, 图中的 21 篇文献均为引用突变时段持续到 2018 年的文献。

笔者对这 21 篇文献进行了整理, 按突变值大小排序在表 8 中列出。结合表 8, 我们可以大致推测, 在基于语料库的话语分析领域, 近年来 COCA 语料库(Davies, 2008) [92]被较多地利用; 在进行话语分析时, 涉及的热门理论, 或者说较热门的关注点包括元话语(Hyland, 2004, 2005a, 2005b) [47] [93] [94]和隐喻(Lakoff, 1980, 1993, 1999; Charteris-Black, 2004; Pragglejaz Group, 2007; Semino, 2008) [34] [38] [95] [96] [97] [98], 当然也还涉及到会话分析(Jefferson, 2004) [68]、体裁分析(Bhatia, 2004) [99]、评价理论(Martin & White, 2005) [55]等经典理论, 此外还有不少从语料库语言学本身出发探索话语分析新路径的文献(Hyland, 2008; Baker, 2006; Baker *et al.*, 2008 等) [15] [16] [100]。从被研究的语篇类型来看, 学术语篇较为热门(Coxhead, 2000; Hyland, 2004, 2005a, 2005b, 2008; Bhatia, 2004; Simpson-Vlach & Ellis, 2010 等) [46] [47] [93] [94] [99] [100] [101]; 此外突变值排在第三的 Jefferson (2004) [68]也表明, 口语语篇的分析也是近年来的热点。

Table 8. Latest literature with high citation burstness value
表 8. 近期的引用高突变值文献

序号	文献	突变值	突变 起始 年份	突变 结束 年份
1	Davies, M. (2008) <i>The Corpus of Contemporary American English: 425 Million Words, 1990-Present</i> . URL (last checked 18 March 2012). http://corpus.byu.edu/coca/ , 2019-5-12.	23.6161	2015	2018
2	Hyland, K. (2005a) <i>Metadiscourse: Exploring Interaction in Writing</i> . Continuum, London.	21.9755	2015	2018
3	Jefferson, G. (2004) <i>Glossary of Transcripts Symbols with an Introduction</i> . In: Lerner, G. H., Ed. <i>Conversation Analysis: Studies from the First Generation</i> . John Benjamins Publishing, Amsterdam/Philadelphia, 13-31.	19.7056	2015	2018
4	Hyland, K. (2008) <i>As Can be Seen: Lexical Bundles and Disciplinary Variation. English for Specific Purposes</i> , 27: 4-21.	19.6128	2015	2018
5	Lakoff, G. (1993) <i>The Contemporary Theory of Metaphor</i> . In: Ortony, A., Ed. <i>Metaphor and Thought</i> , 2nd Edition. Cambridge University Press, Cambridge, 202-251.	17.7546	2015	2018
6	Barr, D. J., Levy, R., and Scheepers, C., <i>et al.</i> (2013) <i>Random Effects Structure for Confirmatory Hypothesis Testing: Keep it Maximal</i> . <i>Journal of Memory and Language</i> , 68, 255-278.	17.7546	2015	2018
7	Semino, E. (2008) <i>Metaphor in Discourse [M]</i> . Cambridge University Press, Cambridge.	17.7546	2015	2018
8	Hyland, K., and Tse, P. (2004) <i>Metadiscourse in Academic Writing: A Reappraisal</i> . <i>Applied Linguistics</i> , 25, 156-177.	16.4555	2015	2018
9	Bresnan, J., Anna, C., Tatiana, N., and Baayen, R.H. (2007) <i>Predicting the Dative Alternation</i> . In: Bouma, G., Kraemer, I. and Zwarts, J., Eds. <i>Cognitive Foundations of Interpretation</i> . Royal Netherlands Academy of Science, 69-94.	16.4555	2015	2018
10	Bhatia, V. (2004) <i>Worlds of Written Discourse: A Genre-based View</i> . Bloomsbury Publishing, London/New Delhi/New York/Sydney.	16.4555	2015	2018
11	Charteris-Black, J. (2004) <i>Corpus Approaches to Critical Metaphor Analysis</i> . Palgrave Macmillan, Hampshire and New York.	15.5473	2015	2018
12	Baker, P. (2006) <i>Using Corpora in Discourse Analysis</i> . Continuum, London & New York.	15.1059	2015	2018
13	Hyland, K. (2005b) <i>Stance and Engagement: A Model of Interaction in Academic Discourse</i> . <i>Discourse Studies</i> , 7, 173-192.	14.4336	2014	2018
14	Martin, J. R. and White, P. R. (2005) <i>The Language of Evaluation: Appraisal in English</i> . Palgrave Macmillan, Basingstoke.	14.2403	2015	2018
15	Lakoff, G. and Johnson, M. (1980) <i>Metaphors We Live By</i> . University of Chicago Press, Chicago.	13.8345	2015	2018
16	Lakoff, G. and Johnson, M. (1999) <i>Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought [M]</i> . Basic Books, New York.	13.8026	2015	2018
17	Simpson-Vlach, R., and Ellis, N. C. (2010) <i>An Academic Formulas List: New Methods in Phraseology Research</i> . <i>Applied linguistics</i> , 31, 487-512.	11.5688	2014	2018
18	Baayen, R. H. (2008) <i>Analyzing Linguistic Data: A Practical Introduction to Statistics Using R</i> . Cambridge University Press, Cambridge.	11.4231	2012	2018
19	Baker, P., Gabrielatos, C., KhosraviNik, M., and Krzyżanowski, M., <i>et al.</i> (2008) <i>A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK</i> . <i>Discourse & Society</i> , 19, 273-306.	11.3066	2012	2018
20	Pragglejaz Group. (2007) <i>MIP: A Method for Identifying Metaphorically Used Words in Discourse</i> . <i>Metaphor & Symbol</i> , 22, 1-39.	8.8991	2016	2018
21	Coxhead, A. (2000) <i>A New Academic Word List</i> . <i>TESOL Quarterly</i> , 34, 213-238.	3.9836	2015	2018

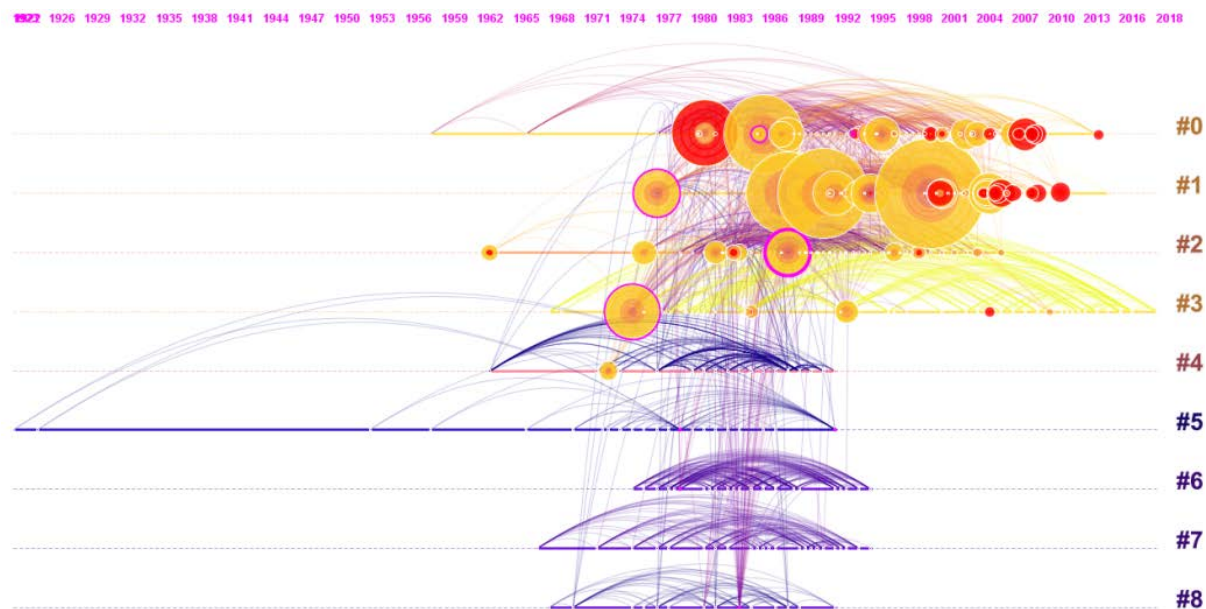


Figure 9. The citation knowledge graph presented in timeline
图 9. 文献共被引聚类时间线图

BAAYEN RH, 2008, ANAL LINGUISTIC DATA, V0, P0	2008	11.4231	2012	2018
BAKER P, 2008, DISCOURSE SOC, V19, P273, DOI	2008	11.3066	2012	2018
HYLAND K, 2005, DISCOURSE STUD, V7, P173, DOI	2005	14.4336	2014	2018
SIMPSON-VLACH R, 2010, APPL LINGUIST, V31, P487, DOI	2010	11.5688	2014	2018
DAVIES M, 2008, CORPUS CONT AM ENGLI, V0, P0	2008	23.6161	2015	2018
HYLAND K, 2005, METADISOURSE EXPLOR, V0, P0	2005	21.9755	2015	2018
JEFFERSON G, 2004, CONVERSATION ANAL ST, V0, P13, DOI	2004	19.7056	2015	2018
HYLAND K, 2008, ENGL SPECIF PURP, V27, P4, DOI	2008	19.6128	2015	2018
LAKOFF G, 1993, METAPHOR THOUGHT, V2, P202, DOI	1993	17.7546	2015	2018
BARR DJ, 2013, J MEM LANG, V68, P255, DOI	2013	17.7546	2015	2018
SEMINO E, 2008, METAPHOR IN DISCOURSE, V0, P1	2008	17.7546	2015	2018
HYLAND K, 2004, APPL LINGUIST, V25, P156, DOI	2004	16.4555	2015	2018
CUENI A, 2007, COGNITIVE FDN INTERP, V0, P69	2007	16.4555	2015	2018
BHATIA V, 2004, WORLDS WRITTEN DISCO, V0, P0	2004	16.4555	2015	2018
CHARTERISBLACK J, 2004, CORPUS APPROACHES TO CRITICAL METAPHOR ANALYSIS, V0, P1, DOI	2004	15.5473	2015	2018
BAKER P, 2006, USING CORPORA DISCOU, V0, P0	2006	15.1059	2015	2018
MARTIN JR, 2005, LANGUAGE EVALUATION, V0, P0	2005	14.2403	2015	2018
LAKOFF G, 1980, METAPHORS WE LIVE BY, V0, P0	1980	13.8345	2015	2018
LAKOFF G, 1999, PHILOS FLESH EMBODIE, V0, P0	1999	13.8026	2015	2018
COXHEAD A, 2000, TESOL QUART, V34, P213, DOI	2000	3.9836	2015	2018
CRISP P, 2007, METAPHOR SYMBOL, V22, P1	2007	8.8991	2016	2018

Figure 10. Latest literature with high citation burstness value
图 10. 近期高突变度文献

4. 结语

本文运用 CiteSpace 软件和 Web of Science 引文数据对基于语料库的话语分析领域进行了可视化分析。借助 CiteSpace 的文献聚类功能，我们生成了该领域中 9 个较大的聚类，涵盖了基于语料库进行话语分析的诸多方向；通过对知识图谱中关键节点文献的分析，我们定位了基于语料库的话语分析的重点文献和热点方向。语料库语言学具有定性定量相结合的特征，可以为话语分析提供可靠的统计数据和客观真实的例证，正在打破传统的话语分析模式，被广泛运用于话语分析领域，扮演着越来越重要的工具角色。可以说，语料库语言学正在引领话语分析的范式转移，传统的话语分析理论，正搭载着语料库

的快车焕发新的生机和活力。

基金项目

2018 年度河南省哲学社会科学规划项目(青年项目)“语料库视角下的中国崛起海外学术话语研究”, 批准号 2018CY030。

参考文献

- [1] 陈汝东. 论话语研究的现状与趋势[J]. 浙江大学学报(人文社会科学版), 2008, 38(6): 130-137.
- [2] Fowler, R. (1996) On Critical Linguistics. In: Caldas-Coulthard, R. and Coulthard, M., Eds., *Texts and Practices: Readings in Critical Discourse Analysis*, Routledge, London, 3-13.
- [3] Stubbs, M. (1994) Grammar, Text, and Ideology: Computer-Assisted Methods in the Linguistics of Representation. *Applied Linguistics*, **15**, 201-223. <https://doi.org/10.1093/applin/15.2.201>
- [4] Stubbs, M. (1998) Whorf's Children: Critical Comments on Critical Discourse Analysis. In: Ryan, A. and Wray, A., Eds., *Evolving Models of Language*, Multilingual Matters, Bristol, 100-116.
- [5] Widdowson, H.G. (1995) Discourse Analysis: A Critical View. *Language and Literature*, **4**, 157-172. <https://doi.org/10.1177/096394709500400301>
- [6] Widdowson, H.G. (1995) Norman Fairclough: Discourse and Social Change (Book Review). *Applied Linguistics*, **16**, 510-516. <https://doi.org/10.1093/applin/16.4.510>
- [7] Fairclough, N. (1996) A Reply to Henry Widdowson's 'Discourse Analysis: A Critical View'. *Language & Literature*, **5**, 49-56. <https://doi.org/10.1177/096394709600500105>
- [8] 辛斌. 批评性语篇分析: 问题与讨论[J]. 外国语(上海外国语大学学报), 2004(5): 64-69.
- [9] 辛斌. 批评话语分析: 批评与反思[J]. 外语学刊, 2008(6): 63-70.
- [10] 辛斌, 高小丽. 批评话语分析: 目标、方法与动态[J]. 外语与外语教学, 2013(4): 1-5+16.
- [11] 梁茂成, 李文中, 许家金. 语料库应用教程[M]. 北京: 外语教学与研究出版社, 2010: 3.
- [12] Hardt-Mautner, G. (1995) 'Only Connect.' Critical Discourse Analysis and Corpus Linguistics. Lancaster: UCREL. <http://stg.lancs.ac.uk/papers/techpaper/vol6.pdf>
- [13] Stubbs, M. (1996) Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture. Blackwell Publishers Ltd., Oxford.
- [14] Baker, P. (2005) Public Discourses of Gay Men. Routledge, London. <https://doi.org/10.4324/9780203643532>
- [15] Baker, P. (2006) Using Corpora in Discourse Analysis. Continuum, London.
- [16] Baker, P., Gabriellatos, C., Khosravi Nik, M., Krzyżanowski, M., et al. (2008) A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press. *Discourse & Society*, **19**, 273-306. <https://doi.org/10.1177/0957926508088962>
- [17] 马博森. 语料库与基于语料库的话语研究[J]. 外国语(上海外国语大学学报), 2009, 32(3): 28-35.
- [18] 许家金. 从词语到话语: 通过语料库开展话语研究[J]. 中国英语教育, 2010, 27(1): 11-19.
- [19] 钱毓芳. 语料库与批判话语分析[J]. 外语教学与研究, 2010, 42(3): 198-202+241.
- [20] 唐丽萍. 语料库语言学在批评话语分析中的作为空间[J]. 外国语(上海外国语大学学报), 2011, 34(4): 43-49.
- [21] 胡江. 意义单位与批评话语分析——基于语料库的西方媒体涉华军事报道意识形态分析[J]. 解放军外国语学院学报, 2016, 39(5): 73-81.
- [22] 陈悦, 刘则渊. 悄然兴起的科学知识图谱[J]. 科学学研究, 2005, 23(2): 149-154.
- [23] 孙晓宁, 闫励, 张强. 科学知识图谱在学科可视化研究中的应用[J]. 图书馆, 2014(5): 87-91.
- [24] Chen, C. (2017) Science Mapping: A Systematic Review of the Literature. *Journal of Data and Information Science*, **2**, 1-40. <https://doi.org/10.1515/jdis-2017-0006>
- [25] 李杰. CiteSpace 中文指南[EB/OL]. <http://blog.sciencenet.cn/blog-554179-1066981.html>, 2019-05-12.
- [26] Chen, C., Hu, Z., Liu, S. and Tseng, H. (2012) Emerging Trends in Regenerative Medicine: A Scientometric Analysis in CiteSpace. *Expert Opinion on Biological Therapy*, **12**, 593-608. <https://doi.org/10.1517/14712598.2012.674507>
- [27] 陈悦, 陈超美, 刘则渊, 等. CiteSpace 知识图谱的方法论功能[J]. 科学学研究, 2015, 33(2): 242-253.

- [28] 陈超美. 如何选取 CiteSpace 中的各项参数[EB/OL]. 科学网陈超美博客. <http://blog.sciencenet.cn/blog-496649-378974.html>, 2019-05-12.
- [29] Quirk, R., Sidney, G., Leech, G. and Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. Longman, London.
- [30] Huddleston, R. and Pullum, G.K. (2002) *The Cambridge Grammar of the English Language*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/9781316423530>
- [31] Chafe, W.L. (1994) *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. University of Chicago Press, Chicago, IL.
- [32] Goldberg, A. (2006) *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press, Oxford.
- [33] Arnold, J.E., Losongco, A., Wasow, T. and Ginstrom, R. (2000) Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering. *Language*, **76**, 28-55. <https://doi.org/10.1353/lan.2000.0045>
- [34] Lakoff, G. and Johnson, M. (1980) *Metaphors We Live By*. University of Chicago Press, Chicago, IL.
- [35] Lakoff, G. (1987) *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago University Press, Chicago, IL. <https://doi.org/10.7208/chicago/9780226471013.001.0001>
- [36] Langacker, R.W. (1987) *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford University Press, Redwood City, CA.
- [37] Goldberg, A.E. (1995) *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press, Chicago, IL.
- [38] Praggeljaz Group (2007) MIP: A Method for Identifying Metaphorically Used Words in Discourse. *Metaphor & Symbol*, **22**, 1-39. <https://doi.org/10.1080/10926480709336752>
- [39] Gundel, J.K., Hedberg, N. and Zacharski, R. (1993) Cognitive Status and the Form of Referring Expressions in Discourse. *Language*, **69**, 274-307. <https://doi.org/10.2307/416535>
- [40] Halliday, M.A.K. (1985) *An Introduction to Functional Grammar*. Arnold, London.
- [41] Lambrecht, K. (1994) Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511620607>
- [42] Swales, J. (1990) *Genre Analysis: English in Academic Settings*. Cambridge University Press, Cambridge.
- [43] Swales, J. (2004) *Research Genres: Explorations and Applications*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9781139524827>
- [44] Hyland, K. (2000) *Disciplinary Discourses: Social Interactions in Academic Writing*. Pearson Education Limited, Harlow, Essex.
- [45] Bhatia, V. (1993) *Analyzing Genre: Language Use in Professional Settings*. Longman, London.
- [46] Coxhead, A. (2000) A New Academic Word List. *TESOL Quarterly*, **34**, 213-238. <https://doi.org/10.2307/3587951>
- [47] Hyland, K. (2005) Stance and Engagement: A Model of Interaction in Academic Discourse. *Discourse Studies*, **7**, 173-192. <https://doi.org/10.1177/1461445605050365>
- [48] Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford University Press, Oxford.
- [49] Biber, D., Johansson, S., Leech, G., *et al.* (1999) *Longman Grammar of Spoken and Written English*. Pearson Education, Harlow, Essex.
- [50] Biber, D. (1998) *Variation across Speech and Writing*. Cambridge University Press, Cambridge.
- [51] Biber, D., Conrad, S. and Cortes, V. (2004) If You Look at...: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, **25**, 371-405. <https://doi.org/10.1093/applin/25.3.371>
- [52] Halliday, M.A.K. and Hasan, R. (1976) *Cohesion in English*. Longman Group, London.
- [53] Halliday, M.A.K. (1994) *An Introduction to Functional Grammar*. 2nd Edition, Arnold, London.
- [54] Halliday, M.A.K. and Matthiessen, C.M.I.M. (2004) *An Introduction to Functional Grammar*. 3rd Edition, Arnold, London.
- [55] Martin, J.R. and White, P.R. (2005) *The Language of Evaluation: Appraisal in English*. Palgrave Macmillan, London.
- [56] Brown, P. and Levinson, S.C. (1987) *Politeness: Some Universals in Language Usage*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511813085>
- [57] Schiffrin, D. (1987) *Discourse Markers*. Cambridge University Press, New York. <https://doi.org/10.1017/CBO9780511611841>

- [58] Goffman, E. (1981) *Forms of Talk*. University of Pennsylvania Press, Philadelphia, PA.
- [59] Grice, H.P. (1975) *Logic and Conversation*. In: Cole, P. and Morgan, J., Eds., *Syntax and Semantics, Volume 3: Speech Acts*, Academic Press, New York, 41-58.
- [60] Clark, H.H. (1996) *Using Language*. Cambridge University Press, Cambridge.
- [61] Leech, G.N. (1983) *Principles of Pragmatics*. Longman, London.
- [62] Levinson, S.C. (1983) *Pragmatics (Cambridge Textbooks in Linguistics)*. Cambridge University Press, Cambridge.
- [63] Sacks, H., Schegloff, E.A. and Jefferson, G. (1974) A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, **50**, 696-735. <https://doi.org/10.1353/lan.1974.0010>
- [64] Sacks, H. (1992) *Lectures on Conversation, Volume 1*. Blackwell, Oxford.
- [65] Pomerantz, A. (1984) Agreeing and Disagreeing with Assessments: Some Features of Preferred/Dispreferred Turn Shapes. In: Atkinson, J.M. and Heritage, J., Eds., *Structures of Social Action: Studies in Conversation Analysis*, Cambridge University Press, Cambridge, 57-101. <https://doi.org/10.1017/CBO9780511665868.008>
- [66] Atkinson, J.M. and Heritage, J. (1984) *Structures of Social Action: Studies in Conversation Analysis*. Cambridge University Press, Cambridge.
- [67] Heritage, J. (1984) A Change-of-State Token and Aspects of Its Sequential Placement. In: Atkinson, J.M. and Heritage, J., Eds., *Structures of Social Action: Studies in Conversation Analysis*, Cambridge University Press, Cambridge, 299-345. <https://doi.org/10.1017/CBO9780511665868.020>
- [68] Jefferson, G. (2004) Glossary of Transcripts Symbols with an Introduction. In: Lerner, G.H., Ed., *Conversation Analysis: Studies from the First Generation*, John Benjamins Publishing, Amsterdam, 13-31. <https://doi.org/10.1075/pbns.125.02jef>
- [69] Labov, W. (1972) *Sociolinguistic Patterns*. University of Pennsylvania Press, Philadelphia, PA.
- [70] McGregor, W.B. (1987) The Structure of Gooniyandi Narratives. *Australian Aboriginal Studies*, **2**, 20-28.
- [71] McGregor, W.B. (1987) Event Progression in Gooniyandi Texts. *Australian Review of Applied Linguistics*, **10**, 158-170. <https://doi.org/10.1075/ara1.10.1.09mcg>
- [72] McGregor, W.B. (1988) Jack Bohemia and the Banjo Affair. *Meridian*, **7**, 34-58.
- [73] McGregor, W.B. (1989) The Discourse Basis of Ergative Marking in Gooniyandi. *La Trobe Working Papers in Linguistics*, **2**, 127-158.
- [74] McGregor, W.B. (1990) *A Functional Grammar of Gooniyandi*. Benjamins, Amsterdam. <https://doi.org/10.1075/slcs.22>
- [75] Quirk, R., Sidney, G., Leech, G. and Svartvik, J. (1972) *A Grammar of Contemporary English*. Longman, London.
- [76] Jespersen, O. (1922) *Language: Its Nature, Development, and Origin*. Routledge, London, New York.
- [77] Jespersen, O. (1924) *The Philosophy of Language*. G. Allen & Unwin Ltd., London.
- [78] Hawkins, J.A. (1978) *Definiteness and Indefiniteness*. Humanities Press, Atlantic Highland, NJ.
- [79] Lumsden, M. (1988) *Existential Sentences: Their Structure and Meaning*. Croom Helm, London.
- [80] Hawkins, J.A. (1991) On (In)definite Articles: Implicatures and (Un)grammaticality Prediction. *Journal of Linguistics*, **27**, 405-442. <https://doi.org/10.1017/S0022226700012731>
- [81] Freeze, R. (1992) Existentials and Other Locatives. *Language*, **68**, 553-595. <https://doi.org/10.2307/415794>
- [82] Erdmann, P. (1976) *There: There Sentences in English: A Relational Study Based on a Corpus of Written Texts*. Tübingen-Verlagsgesellschaft, München.
- [83] Birner, B., and Ward, G. (1994) Uniqueness, Familiarity, and the Definite Article in English. *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society*, Berkeley, CA, 18-21 February 1994, 93-102. <https://doi.org/10.3765/bls.v20i1.1479>
- [84] Ward, G.L. (1988) *The Semantics and Pragmatics of Preposing*. Garland, New York.
- [85] Hoekstra, T. and Mulder, R. (1990) Unergatives as Copular Verbs; Locational and Existential Predication. *The Linguistic Review*, **7**, 1-79. <https://doi.org/10.1515/tlir.1990.7.1.1>
- [86] Brown, G. and Yule, G. (1983) *Discourse Analysis*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511805226>
- [87] Svartvik, J. and Quirk, R. (1980) *A Corpus of English Conversation*. Lund University Press, Lund, Scania.
- [88] Geluykens, R. (1991) Discourse Functions of it-clefts in English Conversation. *Communication and Cognition: An Interdisciplinary Quarterly Journal*, **24**, 343-357.

- [89] 秦明利, 史兴松. 对韩礼德-哈桑衔接理论的再思索[J]. 外语学刊, 1996, 87(3): 24-27.
- [90] 代树兰. 会话分析的缘起与进展[J]. 外语学刊, 2015(6): 25-32.
- [91] 胡壮麟. “Wallace Chafe”文参考文献[J]. 国外语言学, 1992(3): 44-48.
- [92] Davies, M. (2008) The Corpus of Contemporary American English: 425 Million Words, 1990-Present. <http://corpus.byu.edu/coca/>
- [93] Hyland, K. and Tse, P. (2004) Metadiscourse in Academic Writing: A Reappraisal. *Applied Linguistics*, **25**, 156-177. <https://doi.org/10.1093/applin/25.2.156>
- [94] Hyland, K. (2005) Metadiscourse: Exploring Interaction in Writing. Continuum, London.
- [95] Lakoff, G. (1993) The Contemporary Theory of Metaphor. In: Ortony, A., Ed., *Metaphor and Thought*, 2nd Edition, Cambridge University Press, Cambridge, 202-251. <https://doi.org/10.1017/CBO9781139173865.013>
- [96] Lakoff, G. and Johnson, M. (1999) Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought. Basic Books, New York.
- [97] Charteris-Black, J. (2004) Corpus Approaches to Critical Metaphor Analysis. Palgrave Macmillan, Hampshire, New York. <https://doi.org/10.1057/9780230000612>
- [98] Semino, E. (2008) Metaphor in Discourse. Cambridge University Press, Cambridge.
- [99] Bhatia, V. (2004) Worlds of Written Discourse: A Genre-Based View. Bloomsbury Publishing, London, New Delhi, New York, Sydney.
- [100] Hyland, K. (2008) As Can Be Seen: Lexical Bundles and Disciplinary Variation. *English for Specific Purposes*, **27**, 4-21. <https://doi.org/10.1016/j.esp.2007.06.001>
- [101] Simpson-Vlach, R. and Ellis, N.C. (2010) An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics*, **31**, 487-512. <https://doi.org/10.1093/applin/amp058>

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2330-1708, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: ml@hanspub.org