INNOSCIENCE PRESS

Scientific and Social Research, 2021, Volume 3, Issue 6
http://journals.innosciencepress.com/index.php/ssr
ISSN Online: 2661-4332

Knowledge Mapping of Text Linguistics: Discourse Analysis, Text Analysis and Social Media

Hongli Song[1]*, Yuqian Song[2]

[1]International Education College, Zhengzhou University of Light Industry, Zhengzhou 450002, Henan Province, China

[2]School of Journalism and Communication, Henan University, Kaifeng 475001, Henan Province, China

Corresponding author: Hongli Song, kittysong2004@hotmail.com

Abstract: Text linguistics is becoming a significant reference source for theorist as well as practitioners in information fusion, decision making and operations, given its research focus ranging from discourse analysis, text analysis and social media-inspired analysis. This article reviews and maps academic literature in the subject of text linguistics using a cutting-edge scientometrics approach. It recognizes major research outputs, themes, and authors in this way. It compares data harvests to discuss the field's future trends. The work is the first systematic mapping of the subject of text linguistics, and it has important implications for using the scientometrics method to review academic publications.

Keywords: Text linguistics; Discourse analysis; Text analysis; Social media; Literature metrology; Knowledge mapping

Publication date: December 2021; ***Online publication:*** December 29, 2021

1. Introduction

Text linguistics (TL) has been developed rapidly over the years and is becoming increasingly important as the great social discourse power of literature in the past has been transferred to mass media culture in recent years. In the 1970s and 1980s, there were only a few scholars engaging in discourse analysis (DA) research in related disciplines, for example, Widdowson, de Beaugrande and Dressler, and Brown & Yule [1-3]. In pursuit of DA study, critical discourse analysis (CDA) has been rising as a theoretical movement which is focusing on social issues rather than academic inquiries [4-7]. In other words, DA study, in its own academic way, has actively participated in social debate and academic research. The research in TL subsequently stepped beyond its interests in the capacity of DA to the study of text analysis (TA) with the aid of technology. Meanwhile, social media has become a measurable platform within or across which mixing voices and genres of communication shift [8]. Social media sites, such as Twitter, Facebook and YouTube, have emerged to be an important analytical tool in TL research through data mining. Studies on these dynamic and multimodal texts have been important for a more social cognitive approach in conducting research ranging from critical and socio-political discourse to cognitive psychology of discourse processing.

In reality, for the last 40 years of modern TL study, scholars have been attempting to bridge the gap between language and literature, sentence grammar and discourse, action and language, discourse and cognition, and finally cognition and society. This has resulted in a proliferation of viewpoints and analytical rigor across a wide range of discourse genres (politics, media, education, law, etc.) and social spheres. Simultaneously, the sophistication of TL research has resulted in countless papers in its various sub-disciplines. The increasing importance of TL research as a multidisciplinary area necessitates a more holistic view of its evolution from the perspective of its ever-growing literature. Existing assessments of major TL literature are mostly based on the knowledge of a diverse group of researchers and practitioners [9]. While these provide useful information about this discipline and the study agenda in general, they are limited, individualistic, and biased. Furthermore, in the age of social media, assessments of linked TL literatures from multidisciplinary viewpoints - linguistics and technology in particular - are not adequately combined. As a result, their coverage of literature has been limited and contextualized.

In this sense, recent advances in literature metrology have been critical in bridging this gap by evaluating (knowledge mapping) academic publications using big data. Information mapping is defined as the procedures, methods, and tools used to uncover features, display concepts, and link knowledge in a certain research subject so that viewers can look at a huge corpus and gain deeper insights based on a high-level, multidisciplinary view of maps [10]. Although various network modelling tools for visualization had been performed considerately for mining huge implicit domains through social network analysis of citations [11], all researchers indicate that Citespace by Chen Chaomei is a useful tool for discovering trends and emerging topics in the development of a field or domain [12]. The scholars in different specialties have attempted visualization works with the help of Citespace. For instance, Li et al. visualized and analyzed the three top hospitality research journals to identify the hotspot issues and the influential researchers [13].

This work uses knowledge mapping to provide a complete analysis of the field's evolution in the order of DA, TA, and social media by highlighting major research agenda and publications. TL researchers have been attempting to incorporate the orthodox study of text (discourse) under social or cultural environment into the niche study of feasibility of world-wide topics through natural language interface, as evidenced by a comparison of literatures before and after the emergence of social media. It also discovers that popular themes at various stages of this sector are most likely to be related to data volume availability, media qualities, and analytical methodologies. The ultimate goal of TL research is to use data mining techniques to produce a steady and long-term development between discourse analysis, text analysis, and their applications to social media.

The following is a breakdown of the paper's structure. The research approach and data collection are discussed in the following section. The primary findings are then presented in three sections: DA, TA, and social media-inspired research. This is followed by a discussion of the findings in relation to each of the TL study's scenarios. Finally, the study's broader ramifications are discussed in the conclusion section.

2. Methodology

2.1. Web of Science

Related literature data are retrieved from Web of Science (WoS) because it is one of the largest and the most comprehensive database for academic publications. Another reason is that the supported formats of Citespace were a set of bibliographic data files in the field tagged from Institute for Scientific Information (ISI) Export Format.

2.2. Citespace

In this study, Citespace (5.4. R2 version) has been employed as a tool to obtain a visual result of trends and topics in the field of TL study. The visualized analysis began with some basic parameter settings:

(1) Time Slicing: The entire time interval of research was chosen according to data collections.

(2) Pruning: This research concentrates on minimum panning tree pruning.

(3) Links strength between nodes and clusters was processed by a cosine function.

The knowledge map can be processed and the parameters can be figure out through some evaluation indicators such as clustering coefficient, centrality calculation and frequency (or burst) detections.

2.3. Databases

There are two steps in building datasets:

2.3.1. Step 1

Key words, discourse analysis and text analysis, are used separately to search entries in WoS. Timespan is set by “all years” and “Science Citation Index Expanded (SCI-E): 1900-present,” “Social Science Citation Index (SSCI): 1990-present” and “Arts& Humanities Citation Index (A&HCI): 1990-present” are selected for citation database. Book reviews and editorial letters are excluded to eliminate “noise” in the database. 9,423 articles (Year: 1952-2019) are obtained for mapping discourse-analysis research (hereinafter as Database T) and 2,491 articles (Year: 1964-2019) for mapping text-analysis research (hereinafter as Database D).

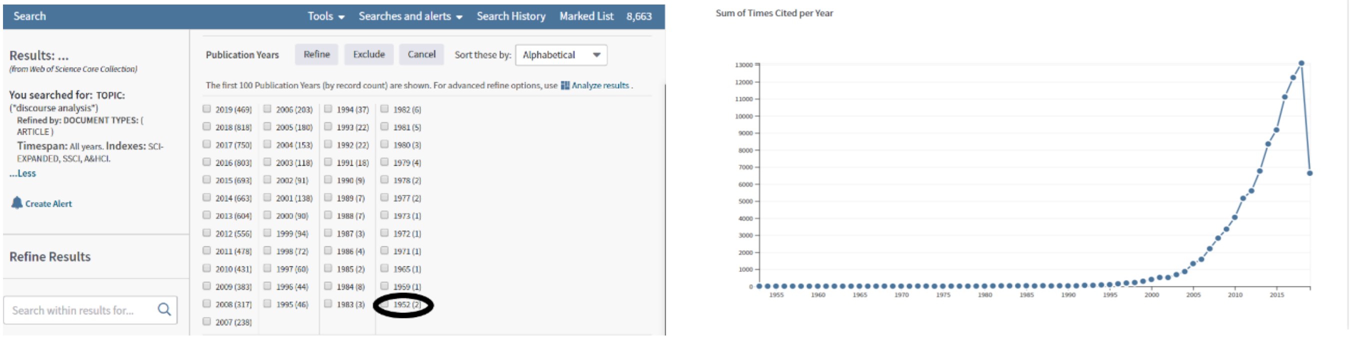
2.3.2. Step 2

The above database is refined by selecting “social media” as a filter. 252 articles between 2011and 2019 are collected as a new database (hereinafter as Database S) for mapping TL study when social media involved.

3. Knowledge mapping of TL study from perspective of DA

3.1. The trend

Within the data source of WoS, the papers in 1952 firstly presented a formal method for DA study (**Figure 1a**). Further review of the paper shows that the analysis of discourse depended only on the occurrence of morphemes as distinguishable elements instead of the analyst’s knowledge of the meaning of each morph [14]. According to the trend of citations (**Figure 1b)**, however, not so many researchers scrutinized on DA study until 2000. The burst at the beginning of 21st century indicates that the work done before is fundamental for DA study, especially in the areas of humanities, linguistics and other social sciences indicated in **Figure 2**.



Results: ...
(from Web of Science Core Collection)

You searched for: TOPIC: ("discourse analysis")
Refined by: DOCUMENT TYPES: (ARTICLE)
Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI.

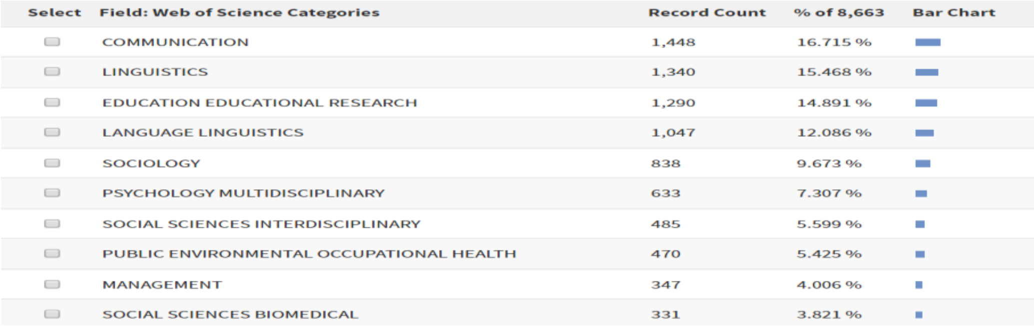
Sort these by: Alphabetical

The first 100 Publication Years (by record count) are shown. For advanced refine options, use Analyze results .

□ 2019 (469)
□ 2018 (818)
□ 2017 (750)
□ 2016 (803)
□ 2015 (693)
□ 2014 (663)
□ 2013 (604)
□ 2012 (556)
□ 2011 (478)
□ 2010 (431)
□ 2009 (383)
□ 2008 (317)
□ 2007 (238)
□ 2006 (203)
□ 2005 (180)
□ 2004 (153)
□ 2003 (118)
□ 2002 (91)
□ 2001 (138)
□ 2000 (90)
□ 1999 (94)
□ 1998 (72)
□ 1997 (60)
□ 1996 (44)
□ 1995 (46)
□ 1994 (37)
□ 1993 (22)
□ 1992 (22)
□ 1991 (18)
□ 1990 (9)
□ 1989 (7)
□ 1988 (7)
□ 1987 (3)
□ 1986 (4)
□ 1985 (2)
□ 1984 (8)
□ 1983 (3)
□ 1982 (6)
□ 1981 (5)
□ 1980 (3)
□ 1979 (4)
□ 1978 (2)
□ 1977 (2)
□ 1973 (1)
□ 1972 (1)
□ 1971 (1)
□ 1965 (1)
□ 1959 (1)
□ 1952 (2)

(a) Number of published papers on DA study (b) Citations of publications on DA study

Figure 1. Numbers and citations of publications on DA study each year



Select	Field: Web of Science Categories	Record Count	% of 8,663	Bar Chart
□	COMMUNICATION	1,448	16.715 %	▬
□	LINGUISTICS	1,340	15.468 %	▬
□	EDUCATION EDUCATIONAL RESEARCH	1,290	14.891 %	▬
□	LANGUAGE LINGUISTICS	1,047	12.086 %	▬
□	SOCIOLOGY	838	9.673 %	▬
□	PSYCHOLOGY MULTIDISCIPLINARY	633	7.307 %	▬
□	SOCIAL SCIENCES INTERDISCIPLINARY	485	5.599 %	▪
□	PUBLIC ENVIRONMENTAL OCCUPATIONAL HEALTH	470	5.425 %	▪
□	MANAGEMENT	347	4.006 %	▪
□	SOCIAL SCIENCES BIOMEDICAL	331	3.821 %	▪

Figure 2. Numbers and bar chart of publications of top 10 disciplines in DA study

3.2. Key networks

Parameter settings: Years Per Slice: 8; Node Type: reference; Top N per slice: 50

By inputting Dataset D in Citespace, the visualized map (**Figure 3**) can help identify the key elements networks in DA study domain.

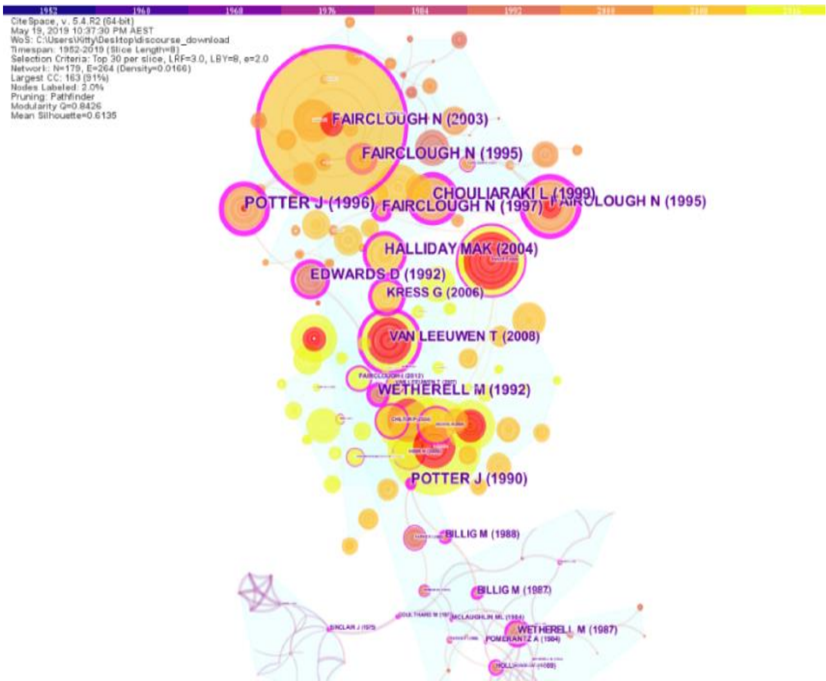


Figure 3. Structure of the most influential references in Dataset D

It can be clearly found that the citations written by Norman Fairclough are the most prominent contribution in DA research. The main reason is that modern entertainment properties have been cumulatively important both for media studies and for sociolinguistics since 1990s. The DA research conformed to shift conventional media language from a single dimensional utterance to multi-dimensional platforms discourse. Fairclough N, is a book by applying his framework to wider processes of social and cultural changes through the traditional media such as TV, radio, and newspapers [15]. The visualization indicates that Fairclough N, can be recognized as another important reference. By detailed study of this article, we can found that DA study were becoming more distinct and classified by providing a method of critical discourse analysis (CDA) against theoretical background [16].

The centrality of Dataset D verifies that, since then, progressive theorists began to vigorously pursue the combination of conventional media language and CDA practice (**Figure 3**). In addition, the greatest values of citations of DA study are both Potter J, (Centrality=0.87) and Chouliaraki L, (Centrality=0.87), indicating that they are greatly acknowledged by researchers of this field. Through detailed examination of each citation, they are the faithful followers of Fairclough, and Fairclough. For example, Potter J, was engaging in overviewing traditional data from newspaper stories to political arguments, even accounts of paranormal events [17]. However, Chouliaraki L improved CDA practice by connecting critical social scientific research [18].

The other significant nodes can be recognized as well in **Figure 3**. For instance, both authoritative references and classical literatures in various social issue niches are represented by Potter J, Edwards D, Wetherell M. The authors at that time drew on a wide range of examples from written to spoken discourses under social environment. Further examination on these publications specifies, we can find that: (1) Wetherell's book was the first systematic introduction of the application of DA to social psychology early in 1987 [19]; (2) Potter J and Edwards D emphasized the role of language (discourse) in everyday life or in public conversations [20,21]; (3) Wetherell M extended to cover broader social issues such as racism, social structure and power ideology [22].

In **Figure 3**, Halliday MAK, Van Leeuwen T and Kress G formed a group of references. Through the detailed investigations, it can be found that they joined together to have a great impact on the topics of the current volume of DA study from semantic analysis aspect. Halliday MAK is a book that scientifically explored discourse on the semantic characters from grammatical metaphor [23]. Van Leeuwen T, presented an analytical framework of semantic constructions from the concept of social dialect, distinct from the concept of phonology and lexicon-grammar [24]. Kress G, provided a toolkit for reading images through enormous data ranging from children's drawings, photo-journalism to three-dimensional forms of sculpture and architecture [25].

3.3. Emerging issues

As shown in **Table 1**, Fairclough N (Burst=88.32; Frequency=169) and Fairclough N (Burst=40.52; Frequency=112) took the first and the second place according to the ranking. This indicates that the two citations have been leading the research trend of DA study in the first decade of 21st century. Actually, through further examination on the two articles, TL study at this stage can be characterized by an integration of TA study into DA study. Fairclough N regards TA study as an essential part of DA study and thus inquired into specific texts and grammatical language in an oscillating way [26], while Fairclough N, puts forward an alternative conception of discourse research towards integrationist methodology [27].

What needs to be pointed out in particular is that Fairclough N emphasized CDA on critical policy studies which laid a foundation for discursive turn of TL practice. It (Frequency=42) (**Table 1**) also suggests that the scientific policy research would be more popular in the near future according to higher value of frequency but the latest year of 2013 [28]. Besides, a group of scholars have been pursuing DA study from the perspectives of inter-discipline, multimedia and technology, according to the other citations in **Table 1**. Baker P examined discourse through combining corpus linguistics [29], yet Wodak R shared an interdisciplinary approach through international examples extracted from online data [30]. After Gee JP introduced a variety of disciplines of DA study such as applied linguistics, psychology, anthropology and communication, Gee JP further provided a toolkit, which is an ideal companion [31].

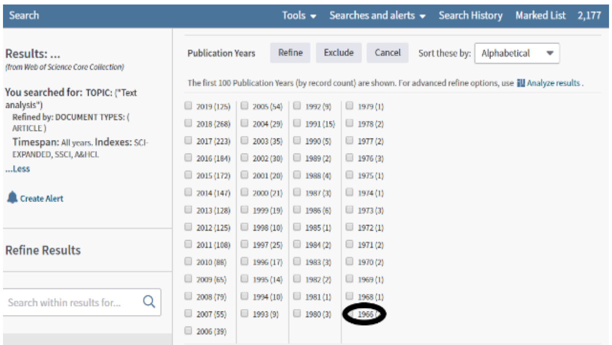
Table 1. Ranking and frequency of top 10 references in burst of DA study

Rank	Burst	Freq.	Author	Year	Title
1	88.32	169	Fairclough N	2003	Analyzing Discourse: Textual analysis for social research
2	40.52	112	Fairclough N	2010	Critical Discourse Analysis in Organizational Studies: Towards an Integrationist Methodology
3	37.41	63	Fairclough N	1995b	Critical discourse analysis: The critical study of language
4	29.35	53	Potter J	1996	Representing reality: discourse, rhetoric and social construction
5	25.3	52	Chouliaraki L	1999	Discourse in late modernity: rethinking critical discourse Analysis
6	23.59	81	Baker P	2008	A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press
7	23.22	56	Gee JP	2005	An introduction to discourse analysis: theory and Method (2nd edition)
8	23.17	78	Wodak R	2009	Methods of Critical Analysis (2nd edition)
9	21.68	42	Fairclough N	2013	Critical discourse analysis and critical policy studies
10	17.53	57	Gee JP	2011	How to do discourse analysis: A Toolkit (1st edition)

4. Knowledge mapping of TL study from perspective of TA

4.1. The trend

Again, the term “text analysis” was originally presented since 1960s based on WoS data source (**Figure 4a**). From the detailed study of literature in this decade, it can be founded that TA study were mostly related to German works at the very beginning of this field regarding the discipline of Informatics [32-34]. In addition, the source from WoS suggests that the frequency of citations in TA study increased rapidly since 2004 (**Figure 4b**). Different from the inclination of DA research, the disciplines of Computer Science and Linguistics took the advantage of the rapid development of technology according to yearly-citation of WoS ranking (**Figure 5**).



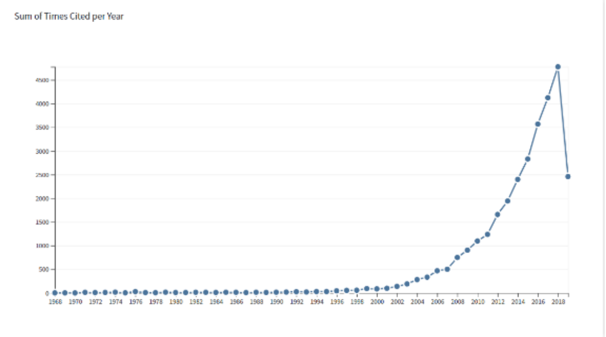
Results: ...
(from Web of Science Core Collection)

You searched for: TOPIC: ("Text analysis")
Refined by: DOCUMENT TYPES: (ARTICLE)
Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI.
...Less

Sort these by: Alphabetical

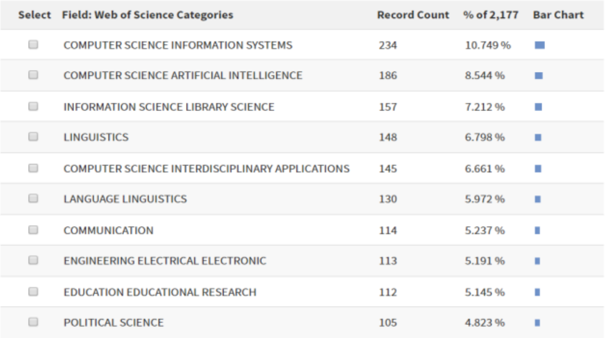
The first 100 Publication Years (by record count) are shown. For advanced refine options, use Analyze results .

□ 2019 (125)
□ 2018 (268)
□ 2017 (223)
□ 2016 (184)
□ 2015 (172)
□ 2014 (147)
□ 2013 (128)
□ 2012 (125)
□ 2011 (108)
□ 2010 (88)
□ 2009 (65)
□ 2008 (79)
□ 2007 (55)
□ 2006 (39)
□ 2005 (54)
□ 2004 (29)
□ 2003 (35)
□ 2002 (30)
□ 2001 (20)
□ 2000 (21)
□ 1999 (19)
□ 1998 (10)
□ 1997 (25)
□ 1996 (17)
□ 1995 (14)
□ 1994 (10)
□ 1993 (9)
□ 1992 (9)
□ 1991 (15)
□ 1990 (5)
□ 1989 (2)
□ 1988 (4)
□ 1987 (3)
□ 1986 (6)
□ 1985 (1)
□ 1984 (2)
□ 1983 (3)
□ 1982 (2)
□ 1981 (1)
□ 1980 (3)
□ 1979 (1)
□ 1978 (2)
□ 1977 (2)
□ 1976 (3)
□ 1975 (1)
□ 1974 (1)
□ 1973 (3)
□ 1972 (1)
□ 1971 (2)
□ 1970 (2)
□ 1969 (1)
□ 1968 (1)
□ 1966



(a) Number of published papers on TA study (b) Citations of publications on TA study

Figure 4. Numbers and citations of publications on TA study each year



Select	Field: Web of Science Categories	Record Count	% of 2,177	Bar Chart
□	COMPUTER SCIENCE INFORMATION SYSTEMS	234	10.749 %	■
□	COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE	186	8.544 %	■
□	INFORMATION SCIENCE LIBRARY SCIENCE	157	7.212 %	■
□	LINGUISTICS	148	6.798 %	■
□	COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS	145	6.661 %	■
□	LANGUAGE LINGUISTICS	130	5.972 %	■
□	COMMUNICATION	114	5.237 %	■
□	ENGINEERING ELECTRICAL ELECTRONIC	113	5.191 %	■
□	EDUCATION EDUCATIONAL RESEARCH	112	5.145 %	■
□	POLITICAL SCIENCE	105	4.823 %	■

Figure 5. Numbers and bar chart of publications of top 10 disciplines in TA study

4.2. Key networks

Parameter settings: Years Per Slice: 8; Node Type: reference; Top N per slice: 50

By inputting Dataset T in Citespace, the visualized map can help identify the key element networks in TA study domain **(Figure 6)**.

Figure 6 shows that Pennebaker JW and Tausczik YR are the most frequently cited publications. The research of TA were not influential until the end of 1990s because of the development of techniques. By means of a word-coding or computerized style, TL study was different from DA study. It was ever a kind of textual analysis instead of discursive analysis. However, TA study did examine written language as an independent and meaningful way from exploring personality or daily word usage to a broad array of real-world behaviors. From this point of view, it was consistent with the integration-orientation of DA study at this period of time.

In **Figure 6**, Weber RP, Mctavish DG, Pennebaker JW, Pennebaker JW, Pennebaker JW and Pennebaker JW were linked closely to each other. Through the careful individual literature investigation, this can be explained that interpretive modes (such as content analysis, contextual data examine, quantitative cognition and software technology) were the popular research issues in TA study field. Pennebaker JW is more admitted textbook at the beginning of 21st century [35]. It updated the original application of TL study (emotional, cognitive and structural components) with a more modern software design of an expanded dictionary - Linguistic Inquiry and Word Count (LIWC).

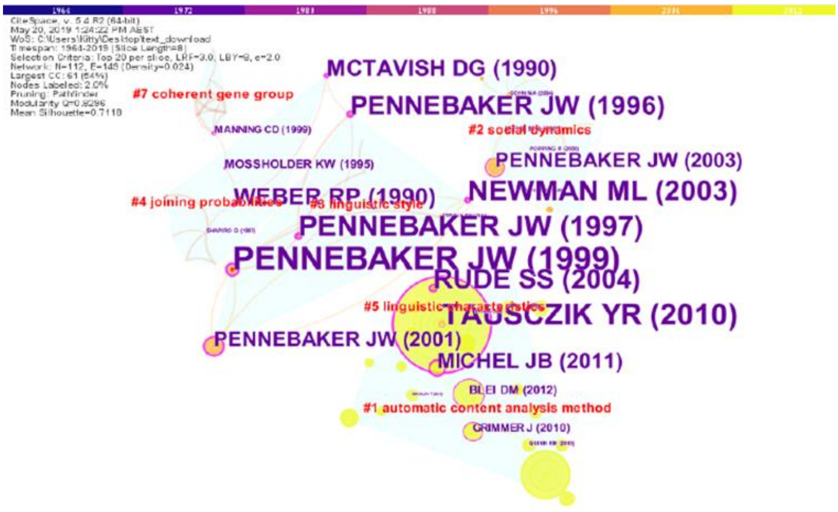


Figure 6. Structure of the most influential references from Dataset T

Mossholder KW, Manning CD, Rude SS, Newman ML, Michel JB, Gimmer J and Blei DM, are formed into the other group. It shows that various computer-based techniques were so popular on analyzing cognitive processing at that time. For example, (1) Mossholder KW, illustrated a qualitative research technique of Dictionary of Affect in Language (DAL) to measure emotion [36]; (2) Manning CD, moved the research technique to the statistical natural language processing [37]; (3) Rude SS, computed the incidence of words in predesignated categories of depressions [38]; (4) Newman ML, correctly classify liars-truth-tellers by a computer-based text analysis program [39]; (5) Gimmer J, estimates the priorities of political actors by Bayesian Hierarchical Topic Model [40]; (6) Blei DM, examined the topic models by surveying a suite of algorithms [41].

4.3. Emerging issues

Similarly, the emerging issues in TA study can be identified according to the values of burst and frequency. Except for Tausczik YR, which is clearly visualized in the map (**Figure 6**), Pennebaker JW (burst=9.93) and Pennebaker JW (burst=9.05) are becoming more and more accepted in recent years because both burst valued and frequency value are the higher (**Table 2**). Through the detailed study, it can be found that the structured analytical methods such as LIWC, a renovated software designed in 2001, continue to be acknowledged by most scholars in the field of TA stud y [42-44]. Moreover, Kahn presents three experimental studies to demonstrate that LIWC is a valid method for measuring verbal expression [45].

Grimmer J reflects another research focus in TA study in the near future because it ranks second according to burst and frequency values (Burst=15.14; Frequency=60) (**Table 2**). Further examination on the citation sheds light on the presentation of automated textual analytical methods, regarded as a standard tool for political scientists [46]. In fact, the research tendency on political science can be traced back to Slapin JB [47]. He proposed a scaling algorithm called WORDFISH to allow researchers to locate parties in one or multiple elections based on word frequencies. This trend of TA study coincided with the most popular research topic and method of DA study of "critical policy studies" indicated in the citation of Fairclough N.

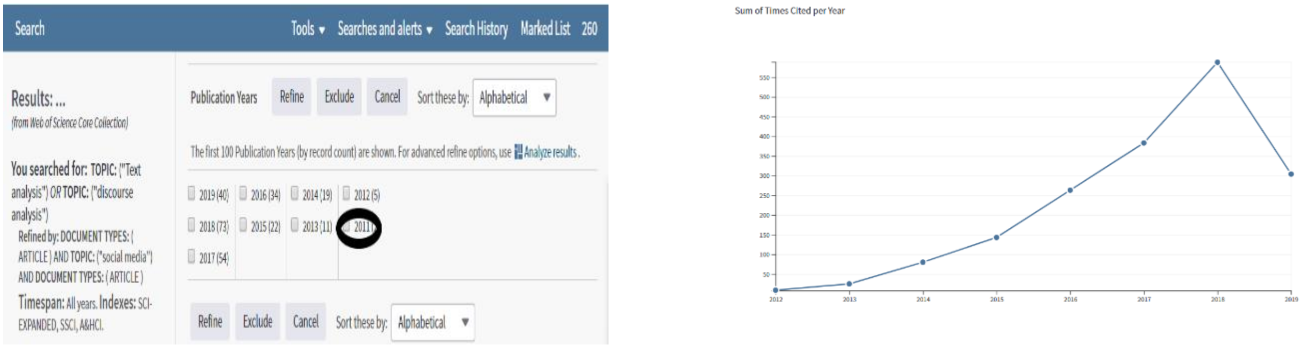
Table 2. Frequency and ranking of top 10 references in burst of TA study

Rank	Burst	Freq.	Author	Year	Title
1	16.18	111	Tausczik YR	2010	The psychological meaning of words: LIWC and computerized text analysis methods
2	15.14	60	Grimmer J	2013	Text as data: The promise and pitfalls of automatic content analysis methods for political texts
3	12.74	36	Pennebaker JW	2001	Linguistic Inquiry and Word Count (LIWC): LIWC2001
4	11.82	26	Pennebaker JW	2003	Psychological aspects of natural language use: Our words, our selves
5	9.93	26	Pennebaker JW	2007a	Linguistic Inquiry and Word Count: LIWC2007
6	9.45	25	Pennebaker JW	2007b	The development and psychometric properties of LIWC2007
7	9.05	24	Blei DM	2012	Probabilistic topic models
8	8.72	23	Pennebaker JW	1999	Linguistic style: Language use as an individual difference
9	7.51	21	Kahn JH	2007	Measuring emotional expression with the Linguistic Inquiry and Word Count
10	7.28	21	Slapin JB	2008	A scaling model for estimating time-series party positions from texts

5. Knowledge mapping of TL study in the era of social media

5.1. The trend

By selecting "social media" as filter to refine the whole dataset of "discourse analysis" and "text analysis" in WoS, social media involved in TL study since the first decade of 21st century (**Figure 7a**). The detailed reviewing on the paper shows that Terras et al. firstly considered the utilization of Twitter, in which the archive of "tweets" was analyzed with a qualitative categorization through open coded [48]. The study in this field advanced quickly since 2013 (**Figure 7b**). The top one discipline in TL study involving in social media has been Communications which is also predominant in traditional DA study (**Figure 8**). This is followed by Sociology, Computer Sciences and Linguistics which totally account for the rest half of the percentage.



Results: ...
(from Web of Science Core Collection)

You searched for: TOPIC: ("Text analysis") OR TOPIC: ("discourse analysis")
Refined by: DOCUMENT TYPES: (ARTICLE) AND TOPIC: ("social media") AND DOCUMENT TYPES: (ARTICLE)
Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI.

Publication Years

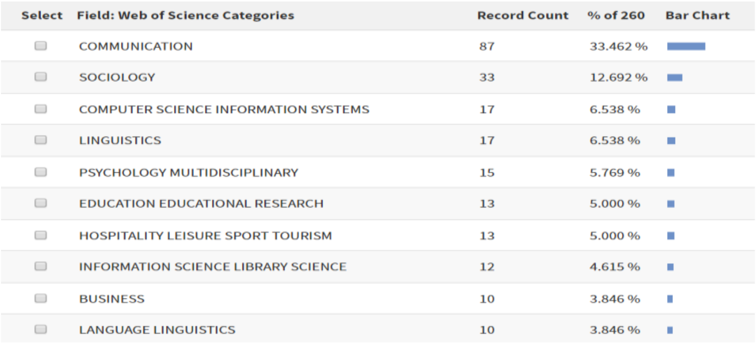
The first 100 Publication Years (by record count) are shown. For advanced refine options, use Analyze results.

□ 2019 (40)	□ 2016 (34)	□ 2014 (19)	□ 2012 (5)
□ 2018 (73)	□ 2015 (22)	□ 2013 (11)	□ 2011
□ 2017 (54)			

(a) Number of published papers

(b) Citations of publications

Figure 7. Numbers and citations of publications on TL study after social media involvement



Select	Field: Web of Science Categories	Record Count	% of 260	Bar Chart
□	COMMUNICATION	87	33.462 %	
□	SOCIOLOGY	33	12.692 %	
□	COMPUTER SCIENCE INFORMATION SYSTEMS	17	6.538 %	
□	LINGUISTICS	17	6.538 %	
□	PSYCHOLOGY MULTIDISCIPLINARY	15	5.769 %	
□	EDUCATION EDUCATIONAL RESEARCH	13	5.000 %	
□	HOSPITALITY LEISURE SPORT TOURISM	13	5.000 %	
□	INFORMATION SCIENCE LIBRARY SCIENCE	12	4.615 %	
□	BUSINESS	10	3.846 %	
□	LANGUAGE LINGUISTICS	10	3.846 %	

Figure. 8 Numbers of publications of top 10 disciplines in TL study after social media involvement

5.2. Influential references

Parameter settings: Years Per Slice: 2; Node Type: reference; Top N per slice: 20

Due to social media-inspired TL study are quiet young, knowledge map was not able to generate the result when Dataset S was input in Citespace. However, there are two most important citations should be noted because of the higher value of centrality or burst. Primarily, critical policy studies would continue to be the most concerned topic in the future study of TL, since Fairclough N is the only one significant publication (centrality=0.19). Secondly, Tausczik YR (Burst=2.47) is the only one current reference generated by Citespace shows that LIWC would be more and more acknowledged in social media-inspired TL study.

The frequency values of citations of social media-inspired TL study can also be ranked with the help of Citespace. Baker P, Blei DM, and Grimmer J (**Table 3)**, used to be the predominant ones in DA study or TA study, still have the significant impact on TL study after the involvement of social media. This suggests that the elements of corpus linguistics, content analysis and topic model have become the key point to which lots of researchers is paying attention in the field of social media-inspired TL study. From the examination on the other citations in **Table 3** such as Bollen J, Lasorsa DL and Kozinets RV, the most likely research elements in the future work would be related to sentiment analysis, social media data and data mining tools in various practices such as stock market prediction, presidential election and word-of-mouth marketing [49-51].

Table 3. Top 5 references in frequency in Dataset S

Rank	Freq.	Author	Year	Title
1	10	Tausczik YR	2010	The psychological meaning of words: LIWC and computerized text analysis methods
2	6	Bollen J	2011	Twitter mood predicts the stock market
3	5	Blei DM	2012	Probabilistic topic models
4	4	Baker P	2008	A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press
4	4	Lasorsa DL	2012	Normalizing Twitter: Journalism Practice in an Emerging Communication Space
4	4	Kozinets RV	2010	Networked narratives: Understanding Word-of-Mouth Marketing in Online Communities
5	3	Grimmer J	2013	Text as data: The promise and pitfalls of automatic content analysis methods for political texts

5.3. Frequently discussed topics

Parameter settings: Years Per Slice: 5; Node Type: keywords; Top N per slice: 50

By inputting Dataset D and Dataset T respectively, the hot topics within the whole process of TL study can be analyzed according to the frequency of keywords. According to the comparisons of frequency values in **Table 4**, the most frequently discussed issues for DA study are “critical discourse analysis,” “identity,” “gender,” “power” and “policy” etc. While the TA researchers are mostly concerned with “system” and “management.” Meanwhile, the hottest method for DA study is “critical discourse analysis,” whereas the most popular methodologies for TA study are “model,” “text mining” and “natural language processing.” Both DA study and TA study have been examining “language,” however, the research in TA has been focusing on “word” based on discourse or text.

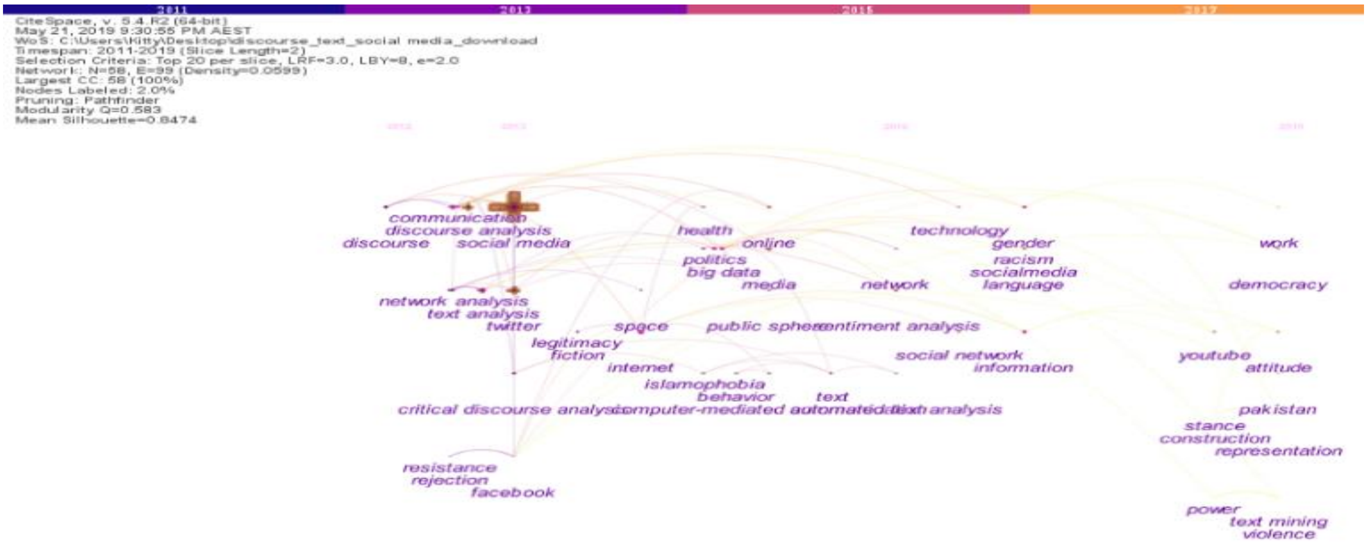
Table 4. Top 10 Keywords in frequency of DA and TA respectively

Rank	Discourse analysis (1952-2019)		Rank	Text analysis (1964-2019)	
	Freq.	**Keyword**		**Freq.**	**Keyword**
1	2815	discourse analysis	1	405	text analysis
2	1200	discourse	2	109	model
3	861	critical discourse analysis	3	96	language
4	520	identity	4	79	system
5	446	gender	5	73	text
6	373	language	6	72	information
7	317	politics	7	64	word
8	333	media	8	58	text mining
9	317	power	9	57	Natural language processing
10	297	policy	10	51	management

5.4. Future issues

Parameter settings: Years Per Slice: 2; Node Type: keywords; Top N per slice: 20

By inputting Data S in Citespace, the hot topics of social-media-related TL study can be visualized (**Figure 9**).



CiteSpace, v. 5.4.R2 (64-bit)
May 21, 2019 9:30:55 PM AEST
WoS: C:\Users\Kitty\Desktop\discourse_text_social media_download
Timespan: 2011-2019 (Slice Length=2)
Selection Criteria: Top 20 per slice, LRF=3.0, LBY=8, e=2.0
Network: N=58, E=99 (Density=0.0599)
Largest CC: 58 (100%)
Nodes Labeled: 2.0%
Pruning: Pathfinder
Modularity Q=0.583
Mean Silhouette=0.8474

Figure 9. Keywords timeline by Dataset S

According to the timeline map, the hot topics of “discourse analysis,” “text analysis,” “communication,” “network analysis,” “critical discourse analysis” “resistance” are closely related to “social media,” “twitter” and “facebook” during the period of 2011 to 2013. With the support of technologies such as “big data,” “online” and “computer-mediated auto text analysis,” TL research topics had been transferred to “politics,” “legitimacy,” “public sphere,” “islamophobia” and “health” from 2013 to 2015.

After that, the hotspots have been focusing on more detailed issues such as “sentiment analysis,” “gender,” “racism” and “language,” associated closely with “social media” and “technology.” With the development of text mining technology, public opinions like “work,” “democracy,” “attitude,” “stance” and “power” would be studied competitively through data mining from “YouTube” since 2017.

6. Discussion on the findings

Firstly, DA laid the fundamental theories in the areas of humanities, linguistics and social sciences, while TA developed rapidly in computer science area with the conversion of data nature according to **Figure 1, Figure 2, Figure 4** and **Figure 5**. Both number of publications and citations (**Figure 7** and **Figure 8**) of TA study keep soaring in the era of social media with the blossom of information technology utilized in the disciplines of Communication, Sociology and Computer Science Information systems.

Secondly, DA research was originally conducted by a disparate group of scholars (Norman Fairclough, Margaret Whetherell and Derek Edwards), which emphasizes the role of language in the practice of linguistics under the traditional forms of communication. Fairclough’s theories of discourse analysis have had a tremendous impact on the social sciences and media discourse. Another group of scholars (Norman Fairclough, Jonathan Potter and Lilie Chouliaraki), who came from more or less diverse academic backgrounds, combined social theories with linguistic research. In fact, led by the researchers such as Norman Fairclough, James Paul Gee, Theo van Leeuwen, Gunther Kress, Paul Baker and Ruth Wodak) in the early 1990s, DA study is made further to be more explicit and systematic by focusing on CDA practices. CDA reflects the interdisciplinary approach of conventional social media including mono- and three-dimensional forms. The integrationists view TA study as an essential part of DA study focusing on specific texts and order of discourse as well through variety of approaches and disciplines since the new century.

Thirdly, the evolution of TA study is clearly different from DA study. TA is initially regarded as a kind of textual processing in Informatics concerning much about computerization in information and library science. And then it has been developed into a kind of content analysis by focusing on the patterns of emphasized ideas in or underlying the real-world behaviors through the word-coding style since 1990s. The emergence of the interest in relating the study of text to social events are concurrent instead. From one point of view, Pennebaker has been remarkably turning text analysis as a cognitive process of natural language in a computer-based techniques in which people make meaning in psychological contexts. The psychometric tool of LIWC, which is developed by Pennebaker, modified by Tausczik and demonstrated by Kahn, is acknowledged greatly by the other related researchers. From another point of view, the group of citations such as Weber, Gimmer, Michel, Mctavish, Blei, Mossholder, Rude and Newman were paying their attention on the changing nature of the textual data to statistically examine emotion, culture and policy under social contexts. In recent years, structured methods or automated textual analytical methods are concerned significantly in TA study.

Fourthly, the two-side features can be identified in the study of TL supported by social media. On the one hand, the stage of this field is focusing chiefly on fundamental theories or techniques such as Tausczik’s LIWC in the psychological area in 2010, Blei’s probabilistic topic models in 2012, Baker’s critical analysis through corpus linguistics in 2008 and Gimmer’s automatic content analysis in political texts in 2013. Nevertheless, Fairclough N of DA study and Gimmer J of TA study should be the most significant and competitive citations because they are acknowledged greatly not only before but after social media involvement. On the other hand, some references are rising to be more influential by trying to capitalize on social media tools in DA study and TA study, taking Bollen's stock market analysis by Twitter, Lasorsa's Twitter's normalization in communication space and Kozinets's comprehending of networked narratives in marketing investigation as examples.

Finally, according to detailed analysis of the literature, the formation and transformation of TL study can be summarized based on different research perspectives, methods and emphases:

(1) Comparison of keywords between DA study and TA study validates the hotspots of TL study before the social media involved in it. Political Science, Communication and Social Science are highlighted in DA study, while the words related to information systems are underlined in TA study. Furthermore, the study of TA paid much attention to the techniques and methodologies such as "model," "text mining" and "natural language processing," etc.

(2) Social media has been involving in DA study and TA study as well gradually and continuously according to the timeline map. General topics linked with localized communication were mostly discussed at the beginning of social media involvement, but public opinions are examined through multimodal social media in recent years.

(3) Twitter was one of the most popular social media to survey the individual personalities like sentiments, gender and racism under the context of social network. YouTube, however, has been becoming more and more important to analyze the political discourse like democracy, attitude and power with the help of big data and computer-mediated text analysis.

7. Conclusion

In this paper, we present a visual and scientometrics survey of text linguistics covering all relevant articles from three main databases of WoS. The approach in this article is based on knowledge mapping tool (Citespace).

Our research yields certain results based on statistical data from various things (e.g. reference, keyword and author). By evaluating various relational variables such as centrality, burst, and frequency, we discovered some common findings. The parallels and contrasts between discourse analysis and text analysis literature are also discussed. The main idea is that when studying text linguistics in relation to social media, it is important to pay close attention. Finally, in order to build the framework for a follow-up study, we generated predictions about research trends and hotspots in this subject.

The results of this study show that the scientometrics method can be used to discover important research publications. Future research in this area will focus on fine-tuning the data extraction and analysis process, as well as delving further into various sub-sectors of text linguistics to investigate their unique dynamics.

Acknowledgments

We are especially grateful to Dr. Bela Stantic (Professor and Head of School of Information and Communication Technology, Griffith University, Australia), Dr. Hui Feng (Professor of Asia Research Institute, Griffith University, Australia), Dr. Yuping Wang (Associate Professor of Humanities, Language and Social Sciences, Griffith University, Australia) and Dr. Sebastian Binnewies (Lecturer of School of Information and Communication Technology, Griffith University, Australia).

Funding

This work was supported by Science and Technology Department of Henan Province (China) under Grant 172102410063; the project of Research on the Policy Evolution and Development Strategy of Sino-foreign

Cooperative Education in the New Era (2021YB0119) supported by Henan Education Department; the Project of Study on Strategies of Western Spread of Confucian Classics from the Perspective of "Community of Shared Future for Mankind" under 2020 Training Young Talented Teachers in Colleges and Universities in Henan Province (2020GGJS132).

Disclosure statement

The author declares no conflict of interest.

References

[1] Widdowson HG, 1979, Explorations in Applied Linguistics. Oxford: Oxford University Press.

[2] de Beaugrande RA, Dressler W, 1981, Introduction to Text Linguistics. Longman's Linguistic Library.

[3] Brown G, Yule G, 1983, Discourse Analysis. Cambridge: Cambridge University Press.

[4] Titscher S, Mayer M, Wodak R, Vetter E, 2000, Methods of Text and Discourse Analysis, Thousand Oaks, CA, Sage.

[5] Johnstone B, 2002, Discourse Analysis. Massachusetts and Oxford: Blackwell Publishers.

[6] van Dijk TA, 2002. Ideology: Political Discourse and Cognition. In P. Chilton, & Ch. Schaffner (Eds.), Politics as Text and Talk, Amsterdam: Benjamins, 33–57.

[7] Wodak R, Meyer M, 2008, Critical Discourse Analysis: History, Agenda, Theory, and Methodology in Methods for Critical Discourse Analysis (2nd ed.), CA, Sage, 1–33.

[8] Unger J, Wodak R, KhosraviNik M, 2016, Critical Discourse Studies and Social Media Data. In: David Silverman, ed. Qualitative Research (4th edition), London, SAGE.

[9] Bouvier G, Machin D, 2018, Critical Discourse Analysis and the Challenges and Opportunities of Social Media. Review of Communications, 18(3): 178–192.

[10] Shiffrin RM, Bornet K, 2004, Mapping Knowledge Domains. Proceedings of the National Academy of Sciences of the United States of America, 101(1): 5183-5185.

[11] Sedighi M, Jalalimanesh A, 2014, Mapping Research Trends in the Field of Knowledge Management. Malaysian Journal of Library & Information Science, 19(1): 71-85.

[12] Chen CM, 2006, Citespace: A Practical Guide for Mapping Scientific Literature. Nova Publishers.

[13] Li XJ, Ma E, Qu HL, 2017, Knowledge Mapping of Hospitality Research: A Visual Analysis Using Citespace. International Journal of Hospitality Management, 60: 77-93.

[14] Wittgenstein L, 1953, Philosophical Investigations. Basil Blackwell Ltd.

[15] Fairclough N, 1995a, Media Discourse. London and New York, E. Arnold.

[16] Fairclough N, 1995b, Critical Discourse Analysis: The Critical Study of Language. Longman Publishing, New York.

[17] Potter J, 1996, Representing Reality: Discourse, Rhetoric and Social Construction. SAGE.

[18] Chouliaraki L, Fairclough N, 1999, Discourse in Late Modernity, Rethinking Critical Discourse Analysis, Edinburgh, UK, Edinburgh University Press.

[19] Potter J, Wetherell M, 1987, Discourse and Social Psychology: Beyond Attitudes and Behavior. London, U.K, Sage Publications Ltd.

[20] Potter J, Edwards D, 1990, Discourse Analysis. European Journal of Social Psychology, 20: 405-24.

[21] Edwards D, Potter J, 1992, The Chancellor's Memory: Rhetoric and Truth in Discursive Remembering. Applied Cognitive Psychology, 6: 187-215.

[22] Wetherell M, Potter J, 1992, Mapping the Language of Racism: Discourse and the Legitimation of Exploitation. London and New York, Harvester Wheatsheaf and Columbia University Press.

[23] Halliday MAK, 2004, The Language of Science. London: Continuum.

[24] van Leeuwen T, 2008, Discourse and Practice: New Tools for Critical Discourse Analysis in Oxford Studies in Sociolinguistics. Oxford University Press.

[25] Kress G, van Leeuwen T, 2006, Reading Images: The Grammar of Visual Design. London and New York, Routledge.

[26] Fairclough N, 2003, Analyzing Discourse: Textual Analysis for Social Research. London and New York, Routledge.

[27] Fairclough N, 2010, Critical Discourse Analysis in Organizational Studies: Towards an Integrationist Methodology. Journal of Management Studies, 47(6): 1213-1218.

[28] Fairclough N, 2013, Critical Discourse Analysis and Critical Policy Studies. Critical Policy Studies, 7(2): 177-197.

[29] Baker P, Gabrielatos C, Khosravinik 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(3): 273-306.

[30] Wodak R, Meyer M, 2009, Methods of Critical Analysis. SAGE.

[31] Gee JP, 2005, An Introduction to Discourse Analysis: Theory and Method (2nd edition). New York and London, Routledge, Taylor & Francis Group.

[32] Koch WA, 1966, Problems of Text Analysis. Lingua, 16: 383-398.

[33] Vonrothk KC, 1966, Methodical Text Analysis: Reporting Technique in Light of Automatic Documentation. Nachrichten fur Dokumentation, 17(5): 169-169.

[34] Batori I, 1969, Automatic Text Analysis for Machine Processing. Nachrichten fur Dokumentation, 20(2): 92.

[35] Pennebaker JW, Francis ME, Booth RJ, 2001, Linguistic Inquiry and Word Count (LIWC): LIWC2001. Mahwah: Lawrence Erlbaum Associates.

[36] Mossholder KW, Setton RP, Harris SG, 1995, Measuring Emotion in Open-ended Survey Responses: An Application of Textual Data Analysis. Journal of Management, 21(2): 335-355.

[37] Manning CD, Schutze H, 1999, Foundations of Statistical Natural Language Processing. Cambridge, Massachusetts, London and England, The MIT Press.

[38] Rude SS, Gortner EM, Pennebaker JW, 2004, Language Use of Depressed and Depression-Vulnerable College Students, Cognition and Emotion, 18(8): 1121-1133.

[39] Newman ML, Berry DS, Richard JM, 2003, Lying Words: Predicting Deception from Linguistic Styles. Available in Personality and Social Psychology Bulletin, 29(5): 665-675.

[40] Gimmer J, 2010, A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Release. Political Analysis, 18: 1-35.

[41] Blei DM, 2012, Probabilistic Topic Models. Communications of the ACM, 55(4): 77-84.

[42] Tausczik YR, Pennebaker JW, 2010, The Psychological Meaning of Words: LIWC and Computerized

Text Analysis Methods. Journal of Language and Social Psychology, 29(1): 24-54.

[43] Pennebaker JW, Booth RJ, Francis M, Linguistic Inquiry and Word Count: LIWC2007, USA: LIWC.net.

[44] Pennebaker JW, Cindy K, Chung CK, et al., 2007, The Development and Psychometric Properties of LIWC2007. USA: LIWC.net.

[45] Kahn JH, Tobin RM, Massey AE, 2007, Measuring Emotional Expression with the Linguistic Inquiry and Word Count. The American Journal of Psychology, 120(2): 263-286.

[46] Gimmer J, 2013, Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis, 21: 267–297.

[47] Slapin JB, Proksch SO, 2008, A Scaling Model for Estimating Time-Series Party Positions from Texts. American Journal of Political Science, 52(3): 705-722.

[48] Terras RC, Warwick MC, Welsh A, 2011, Enabled Backchannel: Conference Twitter Use by Digital Humanists. Journal of Documentation, 67(2): 214-237.

[49] Bollen J, Mao H, Zeng X, 2011, Twitter Mood Predicts the Stock Market. Journal of Computational Science, 2: 1-8.

[50] Lasorsa DL, Lewis SC, Holton AE, 2012, Normalizing Twitter: Journalism Practice in an Emerging Communication Space. Journalism Studies, 13(1): 19-36.

[51] Kozinets RV, de Valck K, Wojnicki AC, et al., 2010, Networked Narratives: Understanding Word-of-Mouth Marketing in Online Communities. Journal of Marketing, 74: 71–89.