Linguistic Studies on Social Media: A Bibliometric Analysis

Ya Sun¹, Gongyuan Wang¹, and Haiying Feng¹

Abstract
This study aimed to present the status quo of linguistic studies on social media in the past decade. In particular, it conducted a bibliometric analysis of articles from the field of linguistics of the database of Web of Science Core Collection with the aid of the tool CiteSpace to identify the general characteristics, major strands of linguistics, main research methods, and important research themes in the area of linguistic studies on social media. The main findings are summarized as follows. First, the study reported the publication trend, main publication venues, researched social media platforms, and languages used in researched social media. Second, sociolinguistics and pragmatics were found to be major strands of linguistics used in relevant studies. Third, the study identified seven main research methods: discourse analysis, critical discourse analysis, conversation analysis, multimodal analysis, narrative analysis, ethnographic analysis, and corpus analysis. Fourth, important research themes were extracted and classified based on four dimensions of the genre framework of social media studies. They were the participation nature and technology affordances of social media in the dimension of compositional level, the researched topics of education, (language) policy and politics in the dimension of thematic orientations, the researched discursive practices of (im)politeness, humor, indexicality and multilingualism in the dimension of stylistic traits, and the researched communicative functions of constructing identity, communicating (language) ideology, and expressing attitude in the pragmatic dimension. Moreover, linguistic studies on social media tended to be characterized by cross-disciplinary and mixed-method approaches.

Keywords
social media, literature review, linguistic studies, genre framework

Introduction
Social media “usually refers to any application or technology through which users participate in, create, and share media resources and practices with other users by means of digital networking” (Reinhardt, 2019, p. 3). Though social media is believed to be conceptually related to social networking sites (SNSs), and the two terms are considered to be used interchangeably, it is widely acknowledged that social media is a broader term, and SNSs, “specially associated with the use of sites such as Facebook and MySpace,” is regarded as a type of social media (McCay-Peet & Quan-Haase, 2017, p. 15). Other forms of social media include message boards and discussion forums, blogging platforms, microblogging services (e.g., Twitter), media-sharing sites (e.g., YouTube), and instant messaging services (e.g., SMS). Social media has become a platform for social interaction among individuals to communicate, entertain, and share, as well as to help promote social bonds or ambient affiliation (Zappavigna, 2012). Moreover, it has also become a communication tool for corporations for engaging with the public, allowing organizations “to contribute to the creation and maintenance of both their identities and their reputations” (Huang-Horowitz & Freberg, 2016, p. 196).

Social media has become one of the most discussed topics in various fields, such as communications, economics, and information technology (Hjorth & Hinton, 2019). Communication studies on social media have focused on impression and relation management (Bazarova et al., 2013; Benthaus et al., 2016), and privacy protection (Lankton et al., 2017). In the area of economics, studies have been conducted on the use of social media to report financial results (Alexander & Gentry, 2014), and on the impact of information disclosure on corporate performances (Cade, 2018; Schniederjans et al., 2013). Research from the perspective of information technology has approached information and emotion detection using computer-mediated technologies (Gründer-Fahrer et al., 2018; Misopoulos et al., 2014). Since “activities on social media primarily consist of
Table 1. Retrieval Queries.
(from Web of Science Core Collection)
(TS = social media) AND LANGUAGE = (English) AND DOCUMENT TYPES = (Article)
Refined by: WEB OF SCIENCE CATEGORIES = (Linguistics OR Language Linguistics)
Timespan: 2009 to 2020. Indexes: SSCI.

language use” (Gnach, 2018, p. 195), social media has also become the research focus of linguistics (Page et al., 2014).

Though some focused reviews have been conducted, such as socioculturally informed studies on language and new media (Akkaya, 2014), and the application of conversation analysis in online talks through social media (Paulus et al., 2016), it seems that a comprehensive review of linguistic studies on social media (henceforth abbreviated as LSSM) is in need. Therefore, the article intends to conduct a bibliometric analysis of articles published in the field of linguistics in the database of Web of Science Core Collection with the aid of CiteSpace to present the status quo in the area of LSSM. More specifically, the following research questions are to be addressed in the present study.

1. What are the general characteristics (the publication trend, publication venues, researched social media platforms, and languages used in researched social media) in the area of LSSM?
2. What are the major strands of linguistics in the area of LSSM?
3. What are the main research methods in the area of LSSM?
4. What are the important research themes in the area of LSSM?

Methodology

The bibliometric analysis “provides a quantitative method for reviewing and investigating extant literature in a given field” (Mou et al., 2019, p. 221). CiteSpace used in the study, one of the most popular bibliometric tools, is “a Java application for analyzing and visualizing co-citation networks” (Chen, 2004, p. 363). It offers various analyses represented by reference journal analysis and keyword analysis that help researchers identify current and future research trends in a field (Mou et al., 2019). For example, Zhang et al. (2015) generated keywords using CiteSpace to identify research foci of social media studies with the support of visualization of references and topics.

First, the database of Web of Science Core Collection was retrieved on December 11, 2020 with queries presented in Table 1 to extract the bibliometric information. To be specific, the purpose of the retrieval was to search the articles related to social media studies in the area of linguistics that were published in English between 2009 and 2020. A total of 794 articles were identified by the queries and their bibliometric information, including article titles, journal titles, publishing years, keywords, abstracts, citation, etc., was downloaded for the follow-up analyses.

Second, the general characteristics in the area of LSSM were detected based on the bibliometric information after using the “analyze results” function of the database of Web of Science Core Collection. It includes the publication trend, main publication venues, researched social media platforms, and languages used in researched social media.

Third, the co-citation network and co-citation frequency list of journals were generated by using CiteSpace to process the imported bibliometric information to identify the major strands of linguistics for researching social media.

Fourth, the keyword list of 394 keywords (Table 2 listing those occurring at least 10 times) was produced also using CiteSpace to pinpoint main research methods of LSSM. The keyword list was searched based on terms of research methods in linguistics and applied linguistics listed in Litosseliti (2010) and McKinley and Rose (2020), such as discourse analysis and corpus analysis, to identify main research methods used in LSSM.

Fifth, the keyword list was further investigated to reveal important research themes. Keywords that occurred at least 10 times (freq ≥ 10) were retained (40 of 394 keywords) as they may best index important research themes. Next, we excluded three types of keywords: (1) meanings of which are too general to be considered as research themes, such as “communication” and “people”; (2) meanings of which are related to research targets of this study, such as “language” and “social media”; (3) meanings of which are related to other research questions, such as “Twitter,” “English,” and “critical discourse analysis.” Then, 18 keywords (words in bold fonts in Table 2) were left for further analyses, together with their semantically relevant keywords (freq < 10), such as “humor” (freq = 11) and “sexist humor” (freq = 2).

The 18 keywords were then categorized based on four dimensions of the genre framework of social media studies (Lomborg, 2014). The framework is adopted because it entails a relatively comprehensive linguistic analysis of social media as emerging communicative genres, “a focus on studying the communicative purpose as manifest in genre conventions of form, style, and content, in recurrent communicative situations or texts” (Lomborg, 2014, p. 26). The first dimension of compositional level examines “the network composition, structures of participation, and activity levels in concrete instances of social media,” and analyze “the social organization of communicative practices on social media” (Lomborg, 2014, p. 31). The second dimension of thematic orientations explores “the predominant topics of communication and the
associated relevance structures as negotiated and regulated by patterns of responsiveness” (Lomborg, 2014, p. 31). The third dimension of stylistic traits addresses “the specific tone that participants must master in order to enact the genre competently, and be recognized and validated by fellow participants as relevant peers and members, in the ongoing negotiations of genre” (Lomborg, 2014, p. 31). The tones are indicated by “choices about a set of related but distinct conversational practices” (Tracy & Robles, 2013, p. 173), such as speech acts, politeness, and humor. The fourth pragmatic dimension identifies “the communicative functionalities and social practices that characterize the uses of a specific genre” (Lomborg, 2014, p. 31).

Results and Discussion

General Characteristics

This section presents the publication trend, main publication venues, researched social media platforms, and languages used in researched social media. Table 3 displays the number of peer-reviewed journal publications by year between 2009 and 2020. A linear regression model was fit and results showed a significant increase in the number of articles published in the examined decades ($F=163.711, p=.000$, $R^2=.942$, Adjusted $R^2=.937$). The publication trend (Figure 1) shows that studies of social media have been receiving increasing academic attention in linguistics, especially after the year of 2017. Table 4 lists the titles of journals that published at least ten relevant articles. Most journals are leading ones in the area of linguistics, such as *Journal of Pragmatics* and *Journal of Sociolinguistics*. Some involve cross-disciplinary studies of linguistics and politics, education, psychology, and anthropology among others, as journal titles suggest, such as *Journal of Language and Politics* and *Journal of Language and Social Psychology*.

Table 3. Number of Publications Published Per Year.

| Year | Number of publications |
|------|------------------------|
| 2009 | 22                     |
| 2010 | 38                     |
| 2011 | 36                     |
| 2012 | 37                     |
| 2013 | 42                     |
| 2014 | 55                     |
| 2015 | 54                     |
| 2016 | 73                     |
| 2017 | 70                     |
| 2018 | 92                     |
| 2019 | 138                    |
| 2020 | 137                    |
| Total| 794                    |

Table 2. Keyword List (Freq >= 10).

| Rank | Keyword          | Freq | Rank | Keyword        | Freq |
|------|------------------|------|------|----------------|------|
| 1    | Language         | 108  | 21   | People         | 15   |
| 2    | Social media     | 99   | 22   | Construction   | 15   |
| 3    | Identity         | 75   | 23   | Model          | 14   |
| 4    | Media            | 68   | 24   | Policy         | 14   |
| 5    | Discourse        | 57   | 25   | Leaner         | 14   |
| 6    | English          | 46   | 26   | Perspective    | 13   |
| 7    | Twitter          | 39   | 27   | Perception     | 13   |
| 8    | Facebook         | 30   | 28   | Ideology       | 12   |
| 9    | Gender           | 29   | 29   | Indexicality   | 12   |
| 10   | Internet         | 24   | 30   | Computer mediated communication | 12 |
| 11   | Literacy         | 23   | 31   | Adult          | 12   |
| 12   | Multimodality    | 22   | 32   | Participation  | 12   |
| 13   | Student          | 22   | 33   | Technology     | 12   |
| 14   | Language ideology| 22   | 34   | Discourse analysis | 11 |
| 15   | Attitude         | 21   | 35   | Humor          | 11   |
| 16   | Education        | 21   | 36   | Media discourse | 10 |
| 17   | Communication    | 20   | 37   | Language policy | 10 |
| 18   | Online           | 18   | 38   | Sociolinguistics | 10 |
| 19   | Critical discourse analysis | 17 | 39   | French         | 10   |
| 20   | Children         | 16   | 40   | Multilingualism | 10   |

Table 2. Keyword List (Freq >= 10).

| Rank | Keyword          | Freq | Rank | Keyword        | Freq |
|------|------------------|------|------|----------------|------|
| 1    | Language         | 108  | 21   | People         | 15   |
| 2    | Social media     | 99   | 22   | Construction   | 15   |
| 3    | Identity         | 75   | 23   | Model          | 14   |
| 4    | Media            | 68   | 24   | Policy         | 14   |
| 5    | Discourse        | 57   | 25   | Leaner         | 14   |
| 6    | English          | 46   | 26   | Perspective    | 13   |
| 7    | Twitter          | 39   | 27   | Perception     | 13   |
| 8    | Facebook         | 30   | 28   | Ideology       | 12   |
| 9    | Gender           | 29   | 29   | Indexicality   | 12   |
| 10   | Internet         | 24   | 30   | Computer mediated communication | 12 |
| 11   | Literacy         | 23   | 31   | Adult          | 12   |
| 12   | Multimodality    | 22   | 32   | Participation  | 12   |
| 13   | Student          | 22   | 33   | Technology     | 12   |
| 14   | Language ideology| 22   | 34   | Discourse analysis | 11 |
| 15   | Attitude         | 21   | 35   | Humor          | 11   |
| 16   | Education        | 21   | 36   | Media discourse | 10 |
| 17   | Communication    | 20   | 37   | Language policy | 10 |
| 18   | Online           | 18   | 38   | Sociolinguistics | 10 |
| 19   | Critical discourse analysis | 17 | 39   | French         | 10   |
| 20   | Children         | 16   | 40   | Multilingualism | 10   |
not only because it is the lingua franca in social media, particularly in US-based platforms such as Twitter and Facebook, but also because nearly half of 794 articles were published in English-speaking countries including US, Great Britain, and Australia.

**Figure 1.** Publication trend in the examined decades.

**Table 4.** Publications venues (number of publications ≥ 10).

| Rank | Journals                                           | Number of publications | Rank | Journals                                           | Number of publications |
|------|----------------------------------------------------|------------------------|------|----------------------------------------------------|------------------------|
| 1    | Social Semiotics                                   | 61                     | 15   | Language and Education                            | 15                     |
| 2    | Journal of Language and Politics                  | 44                     | 16   | System                                             | 15                     |
| 3    | Journal of Pragmatics                             | 42                     | 17   | Computer Assisted Language Learning               | 13                     |
| 4    | Journal of Multilingual and Multicultural Development | 28                  | 18   | Humor: International Journal of Humor Research    | 13                     |
| 5    | Gender and Language                               | 23                     | 19   | Digital Scholarship in the Humanities            | 12                     |
| 6    | Language & Communication                          | 23                     | 20   | English Teaching Practice and Critique            | 12                     |
| 7    | World English                                     | 21                     | 21   | Pragmatics and Society                            | 12                     |
| 8    | International Journal of Bilingual Education and Bilingualism | 20                  | 22   | Applied Linguistics Review                        | 11                     |
| 9    | Journal of Language and Social Psychology         | 17                     | 23   | Aphasiology                                        | 10                     |
| 10   | Language in Society                               | 17                     | 24   | International Journal of Multilingualism          | 10                     |
| 11   | Text & Talk                                       | 17                     | 25   | Language Policy                                   | 10                     |
| 12   | Journal of Linguistic Anthropology                | 16                     | 26   | Literacy                                           | 10                     |
| 13   | Pragmatics                                         | 16                     | 27   | Multilingual Journal of Cross Cultural and Interlanguage Communication | 10                     |
| 14   | Journal of Sociolinguistics                        | 15                     | 28   | Natural Language Engineering                      | 10                     |

**Major Strands of Linguistics**

“Journal co-citation analysis can be used as an operational indicator for the discipline organization of the sciences” (Hu et al., 2011, p. 658). In visual analysis of the research area of social media and government trust, for instance, three
Sun et al.

sub-disciplines (management, environment science, and political science) were identified using journal co-citation analysis (Tong & Song, 2021). The journal co-citation network and journal co-citation frequency list were generated by CiteSpace as shown in Figure 2 and Table 6. Except for the journal Thesis (see Figure 2), the articles of which cannot fall into the domain of linguistic research, the other top six co-cited journals (see Table 6) can be generally categorized into two strands: sociolinguistics and pragmatics. This general division can be confirmed not only by the titles of the journals, such as sociolinguistics-oriented journals as Journal of Sociolinguistics, Language in Society, and Discourse & Society and the pragmatics-oriented journal as Journal of Pragmatics, but also by the keywords shown in Table 2, such as “identity,” “gender,” “ideology,” and “humor,” which are topics frequently explored in sociolinguistics and pragmatics.

A more in-depth analysis based on this categorization reveals that sociolinguistic studies have mainly investigated language changes induced by social media, and the relationship between language variation and society in the digital world. For example, De Decker and Vandekerckhove (2017) investigated the extent to which social variables of gender and age correlate with four linguistic features in social media communication, that is, flooding, leetspeak, abbreviations, and cluster reduction.

Pragmatics, as our analysis reveals, has also played an increasingly important role in social media studies. For instance, Carr et al. (2012) has examined the use of speech acts in the status messages of Facebook and found that expressive speech acts were most frequently used to construct the status messages, followed by assertives, and humor was integrated into almost 20% of the status messages.

**Main Research Methods**

Table 7 provides an overview of the main research methods used in LSSM and their classification based on Litosseliti

---

**Table 5. Most Frequently Researched Social Media Platforms and Languages Used.**

| Rank | Social media platform | Freq | Rank | Language      | Freq |
|------|-----------------------|------|------|---------------|------|
| 1    | Twitter               | 39   | 1    | English       | 48   |
| 2    | Facebook              | 30   | 2    | French        | 10   |
| 3    | Blog                  | 10   | 3    | Chinese       | 9    |
| 4    | Instagram             | 6    | 4    | Spanish       | 5    |
| 5    | YouTube               | 5    | 5    | African language | 1   |

---

**Figure 2. Journal co-citation network.**
Among them, qualitative methods include discourse analysis, multimodal analysis, critical discourse analysis, conversation analysis, narrative analysis, and ethnographic analysis. Discourse analysis, for example, was adopted by Zappavigna (2015) to investigate the use of hashtags on Twitter, and findings showed that they can serve experiential and interpersonal linguistic functions at the level of lexicogrammar, and enact metacommentary at the level of discourse semantics. Critical discourse analysis (discourse-historical approach) was used by Kreis (2017) to explore the meaning and function of Trump’s discursive strategies on Twitter to find that Trump used an informal, direct, and provoking communication style, and employed positive self-presentation and negative other-presentation to construct right-wing populist discourses. Some studies go beyond language symbols and examine non-textual modes as well with multimodal analysis. Texts and photos, for instance, were examined by Hunt (2015) on the Facebook page of a diabetes organization. The results showed that the textual mode, such as personal pronouns, enhanced social relationships among the organization, community members, and the audience, and the visual mode of photos provided representations of successful self-management, charitable volunteering, and need of emotional support of the diabetic individuals.

Themistocleous (2015) applied conversation analysis to explore the structure of digital code-switching between Cypriot and Standard Greek and how it was influenced by the medium- and social-specific characteristics of internet relay chat. Narrative analysis and ethnographic analysis tend to investigate the broader issues in social media studies.

Page’s (2012) exploration of storytelling styles in the Facebook status updates with longitudinal narrative analysis revealed that self-reports were a dominant type of activity, accounting for approximately 70%, followed by projections and shared stories. Back’s (2013) ethnographic analysis of the use of Portuguese on Facebook and the Portuguese level of Brazil students before and after studying abroad found that social media tools, such as Facebook, are playing an important role in improving their Portuguese and helping the acquisition of social media terms in Portuguese.

The quantitative method, corpus analysis, was used by Hardaker and McGlashan (2016) for example, to investigate how sexual aggression is enacted and spread on Twitter by focusing on the frequency, collocation, and keywords.

### Important Research Themes

Table 8 presents results of grouping keywords based on four dimensions of the genre framework of social media studies (Lomborg, 2014) as mentioned above, which reflect important research themes in LSSM.

#### Compositional level

In the dimension of compositional level, two important themes are the participation nature and technology affordances of social media. “The potential for users to participate in interaction and to contribute content is perhaps the main defining characteristic of social media” (Landert, 2017, p. 31). In the circulation of visual small stories (iconic image of Alan Kurdi) on Twitter, users are found to participate in sharing and spreading the story by commenting, resharing, and replying (Giaxoglou & Spilioti, 2020).
The technology of social media provides affordances for users, such as Twitter’s 140-character limit (Bucher & Helmond, 2018), hashtag #, addressing @, emoji, and emoticons, among others. For instance, Matley (2018) examined how users of Instagram strategically employed hashtags as a non-apology marker in a balancing act of (im)politeness that allowed for a level of sanctioned face attack in potentially inappropriate posts. Politicians are found to use @username to refer to themselves and others for identity construction (Coesemans & De Cock, 2017). Matulewska and Gwiazdowicz (2020) investigated the use of emojis in Facebook and blog comments to emphasize emotions toward the community of hunters of ruthless and bloodthirsty animal killers. They found that emojis usually accompany verbal messages to express both positive (support, happiness, and cheerfulness) and negative (anger, hatred, and disgust) emotions.

**Thematic orientations.** In the dimension of thematic orientations, the frequently explored topics are education and (language) policy. The theme of education often involves the use of social media in language teaching and learning. Social media can afford various developments in language teaching and learning, including digital literacies and self-directed learning among others (Reinhardt, 2019). For example, Prichard (2013) used Facebook as the platform to design several activities (e.g., creating profiles, joining groups, and making posts) to help Japanese EFL learners reach the three goals set by the TESOL technology standards. Results suggested that the training was successful in helping the learners’ digital literacies. In another example, Aloraini and Cardoso (2020) investigated students’ use of social media platforms in self-directed EFL learning. Findings revealed the groups’ strategical choices of SM according to different language purposes and the skills to be learned (e.g., they preferred WhatsApp for communication with family and friends, Twitter for reading, and Snapchat for learning aural skills).

Language policy, often triggered by the tension between an official language and a lesser-used language (Lee, 2017), concerns the production and enforcement of linguistic norms (Vessey, 2018) in social media. Despite the top-down language policy of standardization required by the government, for instance, the indigenous youth endeavored to revitalize Maya, an indigenous language of Mexico, by actively using the language on Facebook (Cru, 2015).

Another central topic emerging in LSSM is politics, indicated by keyword “politics” (freq = 6) and other relevant keywords (e.g., political discourse, ideology, Trump, Europe). For example, McDonnell (2020) examined presidential candidates’ political discourse in the 2016 US election on Twitter, and found that both Donald Trump and Hilary Clinton conformed to, and defied, gendered linguistic stereotypes.

**Stylistic traits.** In the dimension of stylistic traits, the repeatedly examined discursive practices are (im)politeness, humor, indexicality, and multilingualism. (Im)politeness often involves the use of speech acts (Falomir, 2015; Taguchi, 2012), such as the politeness-sensitive speech acts of apologies and directives (Leech, 2014). For example, Page (2014) adopted Spencer-Oatey’s description of an apology as a “post-event speech act” (Spencer-Oatey, 2008, p. 19) to examine corporate apologies on Twitter in response to customer complaints, and found that the speech act of apology is combined with additional components, such as explanations and offers of repair, to serve the purpose of saving the face of the corporate and re-establish rapport with customers. Zjakic et al. (2017) explored the use of requests in the form of imperatives in Facebook advertisements, and concluded that...
request, a typical face-threatening speech act when consid-
ered on its own, can ultimately benefit customers when the
surrounding discourse like the visual elements (such as
image and video) used by the advertiser are taken into
account. In addition, Maíz-Arévalo and García-Gómez
(2013) argued that Facebook users made compliments to per-
form politeness and strengthen social rapport by using such
grammatical and lexical choices as exclamative sentences,
declarative sentences, and ellipticals. In the area of business
communication, Juita et al. (2019) examined how executive
leaders on Facebook employed representative, directive,
expressive, commissive, and declarative speech acts to per-
form different politeness principles respectively based on the
context of the speaker’s communicative purpose. Li and Wu
(2018) re-defined speech acts in terms of relational acts, and
argued that multinational companies on Twitter and Weibo
used speech acts of greeting, directing, expressing, sharing,
and retweeting as linguistic strategies to perform facework
and build solidarity with their followers.

Similarly, humor is frequently used to demonstrate rela-
tionships and reinforce group solidarity on social media
(Carr et al., 2012). For example, political humors, in the
forms of captions, caricatures, graffiti, posters, and slogans,
were used to create solidarity among protesters and support-
ers, and as a means for criticism of social reality (Görkem,
2015).

Next, indexicality links (non)linguistic signs with the
sociocultural contexts and meaning in online communication
(Varis, 2016). Indexical devices, such as first person pro-
nouns and third person self-references, were found in politi-
cians’ Twitter posts to construct politician identities
(Coesemans & De Cock, 2017).

Last, multilingualism on social media focuses on lan-
guage diversity and multilingual practice. Studies of lan-
guage diversity often concern the protection or revitalization
of minority languages by users’ speaking indigenous lan-
guages on social media (Leppänen & Peuronen, 2012). For
example, Stern (2017) investigated a Balinese language
Facebook group and found the members persist in the use of
Balinese for its revitalization regardless of the invasion of
the official state language and international language.
Multilingual practice refers to users’ multilingual resources
and repertoires in social media communications. For
instance, Themistocleous (2015) explored how users in inter-
et relay chat switched between Cypriot and Standard Greek,
the two varieties of Greek spoken on the island of Cyprus.
Users were found to bring into play the various languages
available in their linguistic repertoire and, consequently,
switch between them.

**Pragmatic dimension.** In the pragmatic dimension, the fre-
cently researched communicative functions are those of
constructing identity, communicating ideology, and
expressing attitudes. The most important communicative
function of language use in social media is to construct
identities at both the individual and collective levels. Indi-
vidual identities in social media studies include gender,
age, politician identities, among others. For example, Jing-
Schmidt and Peng (2018) found that the male used the mor-
pheme biào (slut) as a gendered personal suffix in the
Chinese cyber lexicon on Weibo to construct masculine
identity of power. By focusing on instances of explicit and
implicit references to age and aging of a Greek female user
of Facebook, Georgalou (2015) argued that age identity is
an interactive and collaborative process both facilitated and
hindered by certain Facebook configurations. In Coese-
mans and De Cock’s (2017) work, politicians were found to
use self-reference to construct politician identity on Twitter
during the election.

Collective identities constructed in social media include
national, group, and corporate identities. Mexican immi-
grants living in the US, for instance, are found to co-con-
struct an imagined experience on Twitter to forge their ethnic
identity and display their sense of belonging to Mexican cul-
ture by participating in specific cultural practices using
hashtags, memes, and multimodal resources (Christiansen,
2019). Tagg and Seargeant (2012) observed that Thai-English
bilinguals use letter repetitions on Facebook and MSN to
index a group identity. Bloggers use different forms of comp-
ounds to construct their identities as members of a distinct-
ive and cohesive social community (Crawford Camiciottoli,
2016). Companies tend to interact with the stakeholders on
social media, in the process of which corporate identities are
constructed. For example, Feng and Wu (2016) examined the
construction of corporate identities on Weibo across differ-
ential ownership in China and concluded with five corporate
identities: authentic identity, specialist identity, companion
identity, journalist identity, and CSR identity. Facebook and
Twitter accounts of Turkish banks were investigated by
Ozdora-Aksak and Atakan-Duman (2014) to find that they
tend to emphasize the softer, especially socially responsible
side of their organizational identities.

The second communicative function of social media is to
communicate ideologies in both explicit and implicit ways.
Politicized text types seem to be the most explicit ideologi-
cally loaded discourse on social media. For example, Trump
used informal, direct, and provoking communication style,
and employed positive self-representation, negative other-
representation, and top-down use of Twitter to communicate
his right-wing populism ideology (Kreis, 2017). Ideologies
are more implicit in the form of users’ responses to online
interlocutors on social media (Pihlaja & Musolff, 2017). For
instance, the way sexual aggression is enacted and spread on
Twitter was investigated as online communities respond to
and participate in forms of extreme online misogyny
(Hardaker & McGlashan, 2016). Besides, Zhao and Liu
(2020) examined the language ideologies surrounding two
regional Putonghua (Chinese standard language) varieties,
and how they are perceived by laypeople through the analy-
sis of Weibo posts.
Similar to the ideological function of social media use is its attitudinal function. Users express attitudes toward social and political life among others on social media as it has become the front of public opinions. For example, Zhang and Zhao (2020) analyzed how Chinese diaspora Youtube vloggers relate their experience during the COVID-19 pandemic drawing on theories of narrative and stance-taking. Results show that vloggers display both universal (e.g., fears) and culturally specific (e.g., mask-wearing) feelings, and invite their viewers to co-construe the emotional experience through narrative devices (e.g., shared stories) and linguistic choices, such as the pronoun ni (you) and address term dajia (everyone).

Cross-Disciplinary and Mixed-Method Study

The above findings further show that LSSM have become increasingly cross-disciplinary in that they are integrated with business, politics, computer science, among others. For instance, Page (2014) combines pragmatics studies on speech acts with the theory of image repair to examine corporate apologies posted on Twitter. Kreis (2017) integrates critical discourse analysis with researches on right-wing populism to explore the meaning and function of Trump’s discursive strategies on Twitter. Ceron and D’Adda (2016)’s use of sentiment analysis suggests a cross-disciplinary approach of linguistics and computer science when they investigated information on Twitter accounts of Italian parties and voting intentions on Twitter of voters.

Second, mixed methods are not uncommon in LSSM despite the classification shown in Table 7. In mixed methods, quantitative approaches can provide contextual information for large-scale use of language and qualitative approaches can focus on specific details (Georgakopoulou & Spiro, 2016; Page et al., 2014; Zeller, 2017). For example, Page (2014) used the corpus-linguistic method to quantify the frequency of corporate apologies posted on Twitter and then identified strategies adopted for image repair through qualitative analysis of example posts from the dataset. In addition, the qualitative survey of students’ opinions on the Skype-assisted learning was combined with the quantitative analysis of types of lessons chosen by students, lessons taken per week by students, CMC (computer-mediated communication) lessons per student in a university classroom to reveal the important role of social media in English learning (Terhune, 2016).

In addition, studies of social media in other fields other than in that of linguistics tend to apply research methods based on non-text data generated on social media, such as geospatial analysis and network centrality analysis. Geospatial analysis uses geographical and spatial data on social media to get better insights about phenomena under investigation (Buchel & Pennington, 2017). For instance, Sobkowicz et al. (2012) argued that geospatial analysis could be used to understand the topics and opinions of the stakeholder from different regions in real-time to forecast political opinions on social media by modeling opinion formation. Network centrality analysis, structural analysis of social networks, aims to help understand the effect of networked influence (Gruzd & Wellman, 2014) on behaviors and connections of online users (Ghajar-Khosravi & Chignell, 2017). For example, Ghajar-Khosravi (2015) conducted the network centrality analysis of social networks on the web to investigate the topological features of social networks with different connections, such as the average distance between the nodes of collaboration-based networks much higher than friendship-based networks.

Conclusion

This paper attempted to present the status quo of linguistic studies on social media using bibliometric analysis. Our findings showed that such studies are increasing and being published in leading journals of linguistics as well as cross-disciplinary ones involving linguistics and politics, education, psychology, and anthropology, among others. Twitter ranked the first in the researched social media platforms, followed by Facebook, Blogs, Instagram, and YouTube. English was the most frequently used language in researched social media followed by French, Chinese, Spanish, and African language. Second, sociolinguistics and pragmatics are found to be the major strands of linguistics. Third, the study identified seven research methods: discourse analysis, critical discourse analysis, conversation analysis, multimodal analysis, narrative analysis, ethnographic analysis, and corpus analysis. Fourth, important research themes were extracted and classified based on four dimensions of the genre framework of social media studies. They were the participation nature and technology affordances of social media in the dimension of compositional level, the researched topics of education, (language) policy and politics in the dimension of thematic orientations, the researched discursive practices of (im)politeness, humor, indexicality and multilingualism in the dimension of stylistic traits, and the researched communicative functions of constructing identity, communicating (language) ideology and expressing attitude in the pragmatic dimension. Moreover, linguistic studies on social media tended to be characterized by cross-disciplinary and mixed-method approaches.

There are also limitations to the current study in some aspects. For instance, the study only selected articles published in the area of linguistics for the sake of focusing on linguistic perspectives, but articles in the area of communication studies may also adopt linguistic approaches. Future research may include more data types (articles published in edited books and proceedings, and monographs) in linguistics as well as other disciplines (communication, politics, education). In addition, the reviewed aspects of strands of linguistics, research methods, and research themes are worth exploring in more depth in future research.
Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Humanities and Social Sciences Project of the Ministry of Education of the People’s Republic of China (grant number: 21YJA740035).

ORCID iD
Gongyuan Wang https://orcid.org/0000-0002-2588-6202

References
Akkaya, A. (2014). Language, discourse, and new media: A linguistic anthropological perspective. Language and Linguistics Compass, 8(7), 285–300. https://doi.org/10.1111/lnc3.12082
Alexander, R. M., & Gentry, J. K. (2014). Using social media to report financial results. Business Horizons, 57(2), 161–167. https://doi.org/10.1016/j.bushor.2013.10.009
Aloraini, N., & Cardoso, W. (2020). Social media in language learning: A mixed-methods investigation of students’ perceptions. Computer Assisted Language Learning. Advance online publication. https://doi.org/10.1080/09588221.2020.1830804
Back, M. (2013). Using Facebook data to analyze learner interaction during study abroad. Foreign Language Annals, 46(3), 377–401. https://doi.org/10.1111/flan.12036
Bazarova, N. N., Taft, J. G., Choi, Y. H., & Cosley, D. (2013). Managing impressions and relationships on Facebook. Journal of Language and Social Psychology, 32(2), 121–141. https://doi.org/10.1177/0261927x12456384
Benthaus, J., Risius, M., & Beck, R. (2016). Social media management strategies for organizational impression management and their effect on public perception. The Journal of Strategic Information Systems, 25(2), 127–139. https://doi.org/10.1016/j.jsis.2015.12.001
Buchel, O., & Pennington, D. R. (2017). Geospatial analysis. In L. Sloan & A. Quan-Haase (Eds.), The SAGE handbook of social media research methods (pp. 285–308). SAGE.
Bucher, T., & Helmond, A. (2018). The affordances of social media platforms. In J. Burgess, A. Marwick, & T. Poell (Eds.), The SAGE handbook of social media (pp. 233–253). SAGE Publications.
Cade, N. L. (2018). Corporate social media: How two-way disclosure channels influence investors. Accounting, Organizations and Society, 68–69, 63–79. https://doi.org/10.1016/j.aos.2018.03.004
Carr, C. T., Schroek, D. B., & Dauterman, P. (2012). Speech acts within Facebook status messages. Journal of Language and Social Psychology, 31(2), 176–196. https://doi.org/10.1177/0261927x12438535
Ceron, A., & D’Adda, G. (2016). E-campaigning on Twitter: The effectiveness of distributive promises and negative campaign in the 2013 Italian election. New Media & Society, 18(9), 1935–1955. https://doi.org/10.1177/1461444815571915
Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. Proceedings of the National Academy of Sciences, 101(Suppl. 1), 5303–5310. https://doi.org/10.1073/pnas.0307513100
Christiansen, M. S. (2019). “Listimos para los #xvderubi”: Constructing a chronotope as a shared imagined experience in Twitter to enact Mexicanness outside of Mexico. Lingua, 225, 1–15. https://doi.org/10.1016/j.lingua.2019.05.002
Coosemans, R., & De Cock, B. (2017). Self-reference by politicians on Twitter: Strategies to adapt to 140 characters. Journal of Pragmatics, 116, 37–50. https://doi.org/10.1016/j.pragma.2016.12.005
Crawford Camiciottoli, B. (2016). “All those Elvis-meets-golfer looks”: A corpus-assisted analysis of creative compounds in fashion blogging. Discourse Context & Media, 12, 77–86. https://doi.org/10.1016/j.dcm.2015.10.002
Cru, J. (2015). Language revitalisation from the ground up: Promoting Yueatec Maya on Facebook. Journal of Multilingual and Multicultural Development, 36(3), 284–296. https://doi.org/10.1080/01434632.2014.921184
De Decker, B., & Vandekerckhove, R. (2017). Global features of online communication in local Flemish: Social and medium-related determinants. Folia Linguistica, 51(1), 253–281. https://doi.org/10.1515/flu-2017-0007
Falomir, L. P. (2015). Leyre Ruiz de Zarobe and Yolanda Ruiz de Zarobe: Speech acts and politeness across languages and cultures. Intercultural Pragmatics, 12(2), 283–288. https://doi.org/10.1515/ip-2015-0014
Feng, W., & Wu, D. D. (2016). State-owned or otherwise: Dialogic constructions of corporate identities by Chinese banks on Sina Weibo. Intercultural Communication Studies, 25(2), 63–81.
Georgakopoulou, A., & Spilioti, T. (2016). Introduction. In A. Georgakopoulou & T. Spilioti (Eds.), The Routledge handbook of language and digital communication (pp. 1–15). Routledge.
Georgalou, M. (2015). Beyond the timeline: Constructing time and age identities on Facebook. Discourse Context & Media, 9, 24–33. https://doi.org/10.1016/j.dcm.2015.07.001
Ghajar-Khosravi, S. (2015). Characteristics of social networks relevant to predictions of social influence [Doctoral dissertation, University of Toronto]. UT Campus Repository. https://tspace.library.utoronto.ca/handle/1807/79725
Ghajar-Khosravi, S., & Chignell, M. (2017). Pragmatics of network centrality. In L. Sloan & A. Quan-Haase (Eds.), The SAGE handbook of social media research methods (pp. 309–327). SAGE.
Giaxoglou, K., & Spilioti, T. (2020). The shared story of #JeSuisAylan on Twitter. Pragmatics, 30(2), 277–302. https://doi.org/10.1075/prag.18057.gia
Gnach, A. (2018). Social media and community building: Creating social realities through linguistic interaction. In C. Cotter & D. Perrin (Eds.), The Routledge handbook of language and media (pp. 190–205). Routledge.
Görgem, Ş. Y. (2015). The only thing not known how to be dealt with: Political humor as a weapon during Gezi Park Protests. Humor: International Journal of Humor Research, 28(4), 583–609. https://doi.org/10.1515/humor-2015-0094
Gründer-Fahrer, S., Schlaf, A., Wiedemann, G., & Heyer, G. (2018). Managing impressions and relationships on Facebook. Journal of Pragmatics, 125, 63–81. https://doi.org/10.1016/j.pragma.2016.12.005

Gruzd, A., & Wellman, B. (2014). Networked influence in social media: Introduction to the special issue. *American Behavioral Scientist, 58*(10), 1251–1259. https://doi.org/10.1177/000276421452087

Hardaker, C., & McGlashan, M. (2016). “Real men don’t hate women”: Twitter rape threats and group identity. *Journal of Pragmatics, 91*, 80–93. https://doi.org/10.1016/j.pragma.2015.11.005

Hjorth, L., & Hinton, S. (2019). *Understanding social media*. SAGE Publications Ltd.

Huang-Horowitz, N. C., & Freberg, K. (2016). Uncovering customer service experiences with Twitter: The case of airline industry. *Management Decision, 54*(4), 705–723. https://doi.org/10.1108/MD-03-2012-0235

Hu, C. P., Hu, J. M., Gao, Y., & Zhang, Y. K. (2011). A journal co-citation analysis of library and information science in China. *Scientometrics, 86*(3), 657–670. https://doi.org/10.1007/s11192-010-0313-6

Hunt, D. (2015). The many faces of diabetes: A critical multimodal analysis of diabetes pages on Facebook. *Language & Communication, 43*, 72–86. https://doi.org/10.1016/j.langcom.2015.05.003

Jing-Schmidt, Z., & Peng, X. (2018). The sluttified sex: Verbal misogyny reflects and reinforces gender order in wireless China. *Language in Society, 47*(3), 385–408. https://doi.org/10.1017/s0047404518000386

Juita, N., Ermanoto, E., & Manaf, N. A. (2019). Language politeness of executive leader in Facebook [Conference session]. Proceedings of the Seventh International Conference on Languages and Arts (ICLA 2018), Indonesia.

Kreis, R. (2017). The “Tweet Politics” of President Trump. *Journal of Language and Politics, 16*(4), 607–618. https://doi.org/10.1075/jlp.17032.kre

Landert, D. (2017). Participation as user involvement. In C. R. Hoffmann & W. Bublitz (Eds.), *Pragmatics of social media* (pp. 31–59). Walter de Gruyter GmbH.

Lankton, N. K., McKnight, D. H., & Tripp, J. F. (2017). Facebook privacy management strategies: A cluster analysis of user privacy behaviors. *Computers in Human Behavior, 76*, 149–163. https://doi.org/10.1016/j.chb.2017.07.015

Lee, C. (2017). *Multilingualism online*. Taylor & Francis Group.

Leech, G. (2014). *The pragmatics of politeness*. Oxford University Press.

Leppänen, S., & Peuronen, S. (2012). Multilingualism on the internet. In M. Martin-Jones, A. Blackledge, & A. Creese (Eds.), *The Routledge handbook of multilingualism* (pp. 347–365). Taylor & Francis Group.

Li, C., & Wu, D. D. (2018). Facework by global brands across Twitter and Weibo. *Discourse Context & Media, 26*, 32–42. https://doi.org/10.1016/j.dcm.2018.03.006

Litosseliti, L. (Ed.). (2010). *Research methods in linguistics*. Continuum.

Lomborg, S. (2014). *Social media, social genres*. Routledge.

Maiz-Arévalo, C., & García-Gómez, A. (2013). “You look terrific!” Social evaluation and relationships in online compliments. *Discourse Studies, 15*(6), 735–760. https://doi.org/10.1177/1464145613490011

Matley, D. (2018). “Let’s see how many of you motherfuckers unfollow me for this”: The pragmatic function of the hashtag #sorrynot-sorry in non-apologetic Instagram posts. *Journal of Pragmatics, 133*, 66–78. https://doi.org/10.1016/j.pragma.2018.06.003

Matulewska, A., & Gwiazdowicz, D. J. (2020). Cyberbullying in Poland: A case study of aggressive messages with emojis targeted at the community of hunters in urbanized society. *Social Semiotics, 30*(3), 379–395. https://doi.org/10.1080/10350330.2020.1731194

McCay-Peet, L., & Quan-Haase, A. (2017). What is social media and what questions can social media research help us answer? In L. Sloan & A. Quan-Haase (Eds.), *The SAGE handbook of social media research methods* (pp. 13–26). SAGE.

McDonnell, A. (2020). Clinton stated, Trump exclaimed! Gendered language on Twitter during the 2016 presidential debates. *Journal of Language and Politics, 19*(1), 71–88. https://doi.org/10.1075/jlp.19085.mcd

McKinley, J., & Rose, H. (Eds.). (2020). *The Routledge handbook of research methods in applied linguistics* (1st ed.). Routledge.

Misopoulos, F., Mitic, M., Kapoulas, A., & Karapiperis, C. (2014). Uncovering customer service experiences with Twitter: The case of airline industry. *Management Decision, 52*(4), 705–723. https://doi.org/10.1108/MD-03-2012-0235

Mou, J., Cui, Y., & Kurcz, K. (2019). Bibliometric and visualized analysis of research on major e-commerce journals using CiteSpace. *Journal of Electronic Commerce Research, 20*(4), 219–237.

Ozdora-Aksak, E., & Atakan-Duman, S. (2014). The online presence of Turkish banks: Communicating the softer side of corporate identity. *Public Relations Review, 41*(1), 119–128. https://doi.org/10.1016/j.prrev.2014.10.004

Page, R. (2012). *Story and social media: Identities and interaction*. Routledge.

Page, R. (2014). Saying “sorry”: Corporate apologies posted on Twitter. *Journal of Pragmatics, 62*, 30–45. https://doi.org/10.1016/j.pragma.2013.12.003

Page, R., Barton, D., Unger, J. W., & Zappavigna, M. (2014). *Researching language and social media: A student guide*. Routledge.

Paulus, T., Warren, A., & Lester, J. N. (2016). Applying conversation analysis methods to online talk: A literature review. *Discourse Context & Media, 12*, 1–10. https://doi.org/10.1016/j.dcm.2016.04.001

Pihlaja, S., & Musolff, A. (2017). Discourse and ideology. In C. R. Hoffmann & W. Bublitz (Eds.), *Pragmatics of social media* (pp. 381–403). Walter de Gruyter GmbH.

Prichard, C. (2013). Training L2 learners to use Facebook appropriately and effectively. *CALICO Journal, 30*(2), 204–225. https://doi.org/10.11139/cj.30.2.204-225

Reinhardt, J. (2019). Social media in second and foreign language teaching and learning: Blogs, wikis, and social networking. *Language Teaching, 52*(1), 1–39. https://doi.org/10.1017/S0261444818000356

Schniederjans, D., Cao, E. S., & Schniederjans, M. (2013). Applying conversion analysis methods to online talk: A literature review. *Computers in Human Behavior, 29*(4), 470–479. https://doi.org/10.1016/j.chb.2012.06.005
Spencer-Oatey, H. (2008). Face, politeness and rapport. In H. Spencer-Oatey (Ed.), *Culturally speaking: Culture, communication and politeness theory* (pp. 11–47). Continuum.

Stern, A. J. (2017). How Facebook can revitalise local languages: Lessons from Bali. *Journal of Multilingual and Multicultural Development, 38*(9), 788–796. https://doi.org/10.1080/01434632.2016.1267737

Tagg, C., & Seargeant, P. (2012). Writing systems at play in Thai-English online interactions. *Writing Systems Research, 4*(2), 195–213. https://doi.org/10.1080/17586801.2011.628583

Taguchi, N. (2012). *Context, individual differences, and pragmatic development*. Multilingual Matters.

Terhune, N. M. (2016). Language learning going global: Linking teachers and learners via commercial Skype-based CMC. *Computer Assisted Language Learning, 29*(6), 1071–1089. https://doi.org/10.1080/09588221.2015.1061020

Themistocleous, C. (2015). Digital code-switching between Cypriot and Standard Greek: Performance and identity play online. *International Journal of Bilingualism, 19*(3), 282–297. https://doi.org/10.1177/1367006913512727

Tong, P., & Song, Z. (2021). Knowledge mapping of government trust and social media research: A visual analysis using CiteSpace. *Journal of the Australian Library and Information Association, 70*(2), 139–156. https://doi.org/10.1080/24750158.2020.1821321

Tracy, K., & Robles, J. S. (2013). *Everyday talk: Building and reflecting identities*. The Guilford Press.

Varis, P. (2016). Digital ethnography. In A. Georgakopoulou & T. Spilioti (Eds.), *The Routledge handbook of language and digital communication* (pp. 55–68). Francis Group.

Vessey, R. (2018). Language policy and the media. In C. Cotter & D. Perrin (Eds.), *The Routledge handbook of language and media* (pp. 329–342). Francis Group.

Zappavigna, M. (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web*. A&C Black.

Zappavigna, M. (2015). Searchable talk: The linguistic functions of hashtags. *Social Semiotics, 25*(3), 274–291. https://doi.org/10.1080/10350330.2014.996948

Zeller, F. (2017). Analyzing social media data and other data sources: A methodological overview. In L. Sloan & A. Quan-Haase (Eds.), *The SAGE handbook of social media research methods* (pp. 386–404). SAGE.

Zhang, L. T., & Zhao, S. (2020). Diaspora micro-influencers and COVID-19 communication on social media: The case of Chinese-speaking YouTube vloggers. *Multilingua, 39*(5), 553–563. https://doi.org/10.1515/multi-2020-0099

Zhang, X., Wang, W., de Pablos, P. O., Tang, J., & Yan, X. (2015). Mapping development of social media research through different disciplines: Collaborative learning in management and computer science. *Computers in Human Behavior, 51*, 1142–1153. https://doi.org/10.1016/j.chb.2015.02.034

Zhao, H., & Liu, H. (2020). (Standard) language ideology and regional Putonghua in Chinese social media: A view from Weibo. *Journal of Multilingual and Multicultural Development*. Advance online publication. https://doi.org/10.1080/01434632.2020.1814310

Zjakic, H., Han, C., & Liu, X. (2017). “Get fit!” – The use of imperatives in Australian English gym advertisements on Facebook. *Discourse Context & Media, 16*, 12–21. https://doi.org/10.1016/j.dcm.2017.01.002