How does a research topic evolve into a research field? A bibliometric analysis of metadiscourse research

Jihua Dong[1], Shuai Dong[1] & Louisa Buckingham[2]
[1] Shandong University (China) & [2] University of Auckland (New Zealand)
dongjihua@sdu.edu.cn, 202120217@mail.sdu.edu.cn,
l.buckingham@auckland.ac.nz

Abstract:

This study adopted a bibliometric approach to trace the diachronic changes in metadiscourse research, based on the research articles and their unique references retrieved from the Web of Science (WoS) core collections from 1980-2020. CiteSpace software was employed to conduct a co-citation analysis to investigate the common themes, the developmental stages, as well as the landmark publications of this domain over the last four decades. Twelve major themes (represented by clusters) were identified and most figure a strong focus in ESP. Studies in these clusters progressed through three developmental stages, namely the conceptualizing stage, the maturing stage, and the flourishing stage. Labeled transformative studies, three key studies were identified as playing a key role in the developmental stage. The use of structural variation analysis contributed to address the otherwise over-reliance on co-citation analysis. By identifying notable features in the development of metadiscourse research, this study provides insight into the evolution of the scientific field.

Keywords: bibliometric, metadiscourse studies, co-citation analysis, structural variation analysis, developmental stages

Resumen:

¿Cómo un tema de investigación termina convirtiéndose en un ámbito de investigación? Análisis bibliométrico de la investigación sobre el metadiscurso

Este estudio adopta un enfoque bibliométrico para trazar los cambios diacrónicos en la investigación del metadiscurso con base en los artículos de investigación y sus referencias únicas recuperados de las colecciones principales

de *Web of Science* (WoS) entre 1980 y 2020. Se empleó el software *CiteSpace* para llevar a cabo un análisis de co-citación con el fin de investigar los temas comunes, las etapas de desarrollo, así como las publicaciones de referencia de este ámbito en las últimas cuatro décadas. Se identificaron doce temas principales representados por grupos, la mayoría de los cuales se centraban en el ESP (Inglés con fines específicos). Los estudios de estos grupos progresaron a través de tres etapas de desarrollo, a saber, la etapa de conceptualización, la etapa de maduración y la etapa de florecimiento. Se identificaron tres estudios transformadores clave, que contribuyeron a marcar el comienzo de la etapa de desarrollo. El uso del análisis de variación estructural utilizado en este estudio contribuyó a resolver la gran dependencia del análisis de co-citación. Nuestro análisis identificó rasgos notables en el desarrollo de la investigación sobre metadiscurso, que pueden aportar información sobre la evolución de los campos científicos.

Palabras clave: bibliometría, estudios de metadiscurso, análisis de co-citación, análisis de variación estructural, etapas de desarrollo

1. Introduction

Bibliometric analysis is a quantitative review approach used to trace the development of research fields by identifying the stages through which a topic progresses, tracking research trends, and investigating features of academic publications such as authorship and impact (Merigó et al., 2016; Tian & Wise, 2020). Specific applications of this approach in scientific disciplines have enabled insights into the origination and the formation of the scientific domain (Yeung et al., 2017), the identification of high-impact work (Merigó et al., 2016) and co-citation patterns (Merigó et al., 2016). Efforts have also been devoted to sketching out research trends in applied linguistics through a bibliometric lens, e.g., lexicography (De Schryver, 2009), non-professional interpreting (Martínez-Gómez, 2015; Zhang, et al., 2015), genre analysis (Pérez-Llantada, 2015), and more recently, applied linguistics (Lei & Liu, 2018), English for specific purposes (ESP) (Liu & Hu, 2021), English for academic purposes (EAP) (Hyland & Jiang, 2021) and data-driven learning (Dong et al., 2022).

To date, bibliometric analysis has been primarily employed to investigate the general development of an established research field (e.g., Aryadoust & Ang, 2019; Fu et al., 2021; Hyland & Jiang, 2021). In this study, we adopted this approach to explore the evolution of metadiscourse in applied linguistics

from its initial emergence to its current status as an established domain of study.

The concept of metadiscourse, which was initially mentioned by Harris (1959) and finally defined by Schiffrin (1980), received some attention in the 1980s (e.g., Crismore, 1989; Kopple, 1985; Williams, 2008) and interest in this domain has increased over the years (Hyland et al., 2022). The rising popularity of this topic over the last two decades has helped to make it a prominent domain of research in linguistics, which is studied from a variety of perspectives (D'Angelo & Consonni, 2020; Dong & Buckingham, 2018; Hyland, 2017).

In this study, we undertook a systematic quantitative review of research in this domain to map out more specific developmental trends and predict potential future research themes in metadiscourse. Specifically, a bibliometric analysis of metadiscourse research over four decades (1980-2020) was conducted to identify research themes and evolutionary stages, using the conceptualization of scientific domain formulated by Shneider (2009), and predict future developments in this domain. The following questions guide our endeavor:

1) What are the major themes in metadiscourse research between 1980 and 2020?

2) What stages can be identified in the development of metadiscourse research since its genesis?

3) What are the most transformative publications in the recent development of metadiscourse research?

2. Literature review

Metadiscourse is mainly concerned with the commentary on a text (written or spoken) by its producer, predominantly in academic register (Hyland, 2017), and it refers to audience-oriented language that is intended to support readers or listeners in processing information. This concept is based on the premise that language not only interacts with the outside world, transmitting information of various types, but also with itself, providing readers with tools for interpreting, organizing, and evaluating what is being said (Jiang & Hyland, 2015). Metadiscourse plays a vital role in facilitating effective

communication between writers and readers and influencing readers' acceptance of the assertions expressed in a text (Hyland, 2005), and it has thus attracted substantial attention in discourse studies.

Two main approaches have been prevalent in the study of metadiscourse: broad and narrow. The broad approaches are quantitative, while the narrow approaches are qualitative. The broad approach advocates retrieving all instances of a pre-defined list of subset members on a large scale based on the premise that each form searched for has the same function. It exhibits a high reliance on linguistic form, and is quantitative in nature. The automated retrieval enabled in this approach allows for a comparison of frequency and distribution trends across large data sets, and thus researchers can effectively compare genres, registers, and contexts.

In contrast, the narrow approach involves greater manual recognition of potential metadiscourse and analyzes extended metadiscursive meaning. It is characterized by a high reliance on context and is thus smaller scale and qualitative in nature. This method, while addressing fewer aspects of metadiscourse than the broad approach, provides a more in-depth knowledge of how metadiscourse functions in particular contexts. (Ädel and Mauranen (2010) provide a more detailed distinction between these two approaches.)

One strand of previous studies on metadiscourse has focused on systematic reviews of the field (i.e., D'Angelo & Consonni, 2020; Hyland, 2017; Hyland et al., 2022; Khedri et al., 2013). For instance, Khedri et al (2013) reviewed studies on metadiscourse in 1990s-2010s with consideration of how authors employ interactive metadiscourse markers in academic abstracts in the soft sciences, and they illustrated how these studies contribute to the development of metadiscourse research. Hyland (2017) provides a narrative account of the development of this field in his review of metadiscourse research over the last four decades. In a recent study, Hyland et al. (2022, p. 3) mapped out metadiscourse research from the genre perspective and listed several commonly employed contrastive variables in this domain (i.e., genres, modes, languages, first language writers, student proficiencies and time). Although these systematic reviews have supplemented our knowledge regarding the development of metadiscourse studies, there is still a need to conduct a systematic quantitative review of research in this domain to map out the developmental trends and predict potential future research themes in metadiscourse research.

3. Methodology

3.1. Dataset

This study is based on a dataset comprised of academic research articles on the topic of metadiscourse, retrieved with the search string "metadiscourse" from the Web of Science (WoS) core collection database. As metadiscourse adopted in this study is a broad domain, and many specific concepts were found in related studies (e.g., Hyland (1998) examined hedging and boosting functions between disciplines), before the search in the database, we undertook a pilot study, using specific concepts in metadiscourse ("boosters", "hedges", "self-mention", "stance", etc.) and similar terms ("metadiscourse", "metatext" and "text reflexivity"), following Mauranen (1993) and Ädel (2006), as the search terms. An in-depth analysis of the results from the pilot study shows that the search results using these specific concepts of metadiscourse are included by using the search term "metadiscourse", while the search results using alternative terms yield a number of irrelevant references, such as "theories of reflexivity" or "intertextual practice".

In another pilot search, the use of "stance", "evaluation" and other interactional elements in the framework of Hyland (2005) yielded a number of unrelated results. For example, the search string "stance" generated results related to another well-known framework that classifies it into "epistemic", "attitude" and "style". A close inspection of the related items, on the other hand, indicated that they bear a strong connection to the metadiscourse framework (Hyland, 2005). Therefore, to ensure the purity and comprehensiveness of references on metadiscourse, we used "metadiscourse" exclusively as the search term, which is also in line with Hyland (2017) and Khedri et al. (2013). The search was restricted to the period from 1980 to 2020 as, according to Hyland (2017), widespread use of this term in applied linguistics started from 1980. A preliminary literature search in the WoS database also confirmed this premise.

In terms of the document type, the search was constrained to "articles", while excluding "book reviews" and "proceeding papers", following Khedri et al. (2013). In this process, "review articles" were also excluded as this document type tends to have a high citation rate that may thus distort the citation-based clustering analysis (Ho et al., 2017). Also, to limit our review to studies in the fields of language and communication, only studies under the "Linguistics" category in WoS were included. The dataset retrieved

consisted of 480 studies and 13,991 unique references to metadiscourse. The retrieved bibliographic records, including Author, Title, Source, Abstract, and References, were downloaded as a TXT file for further analysis.

3.2. Methods of analysis

CiteSpace is an analytical tool used for visualizing and analyzing trends in a scientific domain and employs citation data retrieved from the Web of Science (Chen, 2006). It has been widely employed to detect "the knowledge structure of a discipline, emerging trends, and developing modes of a field from a macroscopic perspective" (Fu et al., 2021, p. 901) in a wide range of fields, including the language sciences (Aryadoust & Ang, 2021; Fu et al., 2021; Liu et al., 2019).

CiteSpace 5.7.R2 was employed to process the dataset and generate a dynamic co-citation network of references that were co-cited by all studies in the dataset. To identify the theme(s) in a cluster of references, CiteSpace extracted noun phrases from the titles, keyword lists, or abstracts of articles that cited the particular cluster (Chen, 2016). To construct an optimal network (Chen, 2016), a network pruning algorithm (Pathfinder) was adopted, and the log-likelihood ratio (LLR) algorithm was also used to extract cluster labels automatically in the co-citation network.

To address Research Question 1, which inquires into the common themes in metadiscourse studies, a co-citation analysis was conducted. Co-citation is concerned with "the frequency with which two documents are cited together by other documents" (Small, 1973, p. 256). Co-citation analysis can uncover themes and characteristics of a specific domain (Chen, 2016), and it is a prevalent research method employed in bibliometric studies to analyze the theme of a research field. This analysis can help to reveal main themes and thus provides an "objective way of modeling the intellectual structure of scientific specialties" (Small, 1973, p. 256). Studies with interconnected references are grouped into the same clusters, and these represent a specific research topic in the co-citation network (Chen, 2006).

In the specific analysis, modularity and silhouette scores were used to detect the quality of a co-citation network. The modularity value measures the composition of the network and falls within the range of 0 to 1. A value greater than 0.3 indicates that the network associations are significant. The silhouette score identifies the degree of homogeneity of the studies contained in clusters; a score above 0.5 represents an acceptably homogenous cluster.

For a more refined picture of the themes, we also identified the landmark publications, which refer to publications that can greatly influence the development of a domain (Shu & Liu, 2021). In the co-citation network, landmark publications are those with high values in the three metrics instead of direct cited frequency following Chen (2011). Previous studies used three metrics: sigma, betweenness centrality, and citation burstness to find a landmark publication (Liu & Hu, 2021). This metric combines the strengths of betweenness centrality and citation burst. Betweenness centrality (0–1) has been used in previous studies to detect "potentially revolutionary" publications (Chen et al., 2010, p. 1390); and citation burst refers to a surge of citations and is used to locate research areas of high popularity (Chen, 2016). Sigma identifies a node within a network of cited references that is structurally significant due to its rapid increase in citations, and it is believed to combine the strength of the other two aforementioned metrics (Chen, 2016). Therefore, the sigma metric was used in this study to identify landmark publications.

To address Research Question 2, which concerns the developmental stages of metadiscourse research, Shneider's evolutionary model was employed as a theoretical framework. The theoretical framework has been employed in a number of previous studies (e.g., Chen, 2012; Chen & Song, 2019; Li et al., 2021) that investigation the development of scientific fields (e.g., Kuhn, 1962; Shneider, 2009). Scientists' working style, research focus and limitations often "depend on the evolutionary stage of a scientific discipline" (Shneider, 2009, p. 221). According to Shneider (2009), a scientific discipline normally includes four evolutionary stages in its development. The first stage of scientific development is to introduce a new subject to the community, and the second is to develop the main research techniques of this field. In the third stage, the field redefines the subject matter with the techniques developed in Stage 2, thereby generating new insights. At Stage 4, research will turn to the practical application of previously generated knowledge. Given that the theory has been recently proposed, few studies have employed Shneider's theory in applied linguistics. One of these is Liu and Hu (2021), which explores the development of ESP between 1908-2018 in two prestigious journals and identifies a three-stage development of ESP.

In this study, we mapped the clusters retrieved by Citespace, based on their beginning time, duration, interactions with other clusters, and the matching of the labels with the functions defined by Shneider (2009). A manual analysis of the clusters and labels was conducted to match articles with one

of the four stages defined by Schneider. The first and second author independently completed the mapping of the clusters according to the model, achieving a high inter-rater agreement of 93%. Disagreements were resolved through discussion.

With respect to Research Question 3, a structural variation analysis (SVA) was conducted in Citespace to determine the studies of transformative potential in ushering in a new stage of development. SVA is a method used to measure the transformative potential of ideas expressed in a newly published paper. The real-time capability of this method solves the intrinsic difficulty of co-citation analysis in failing to uncover recent publications and has been adopted by several studies (e.g., Azam et al., 2021; Hou et al., 2020; Sebastian & Chen, 2021). The approach analyses changes in a network, such as a broadening of the scientific domain or the creation of new links between previous studies (Chen, 2012). A study detected by this method (i.e., a potentially transformative study) refers to papers containing "innovations that are not previously established in studies cited in the respective list of references and the network of references cited contained therein" (Olmeda-Gómez et al., 2019, p. 1564).

Following Chen (2011), this study adopted three parameters to calculate the studies with transformative potential, namely: modularity change rate, cluster linkage, and centrality divergence. Modularity change rate measures the structural changes of the underlying co-citation network induced by connections created by new publications. The higher the value of modularity change rate, the greater the potential impact that the new paper is expected to have on the co-citation network. Cluster linkage uncovers "the overall structural change introduced by an article in terms of new connections added between clusters" (Chen, 2012, p. 439). A higher score of cluster linkage indicates a higher potential to change the whole co-citation network, that is, the connections between clusters. Finally, centrality divergence reviews "the relative entropy of betweenness centrality across all the nodes in the baseline co-citation network" (Chen, 2013, p. 626). This metric measures the variation of the distribution of centrality betweenness introduced by the new article. According to Chen (2012), a study with a high centrality divergence score often involves more than one cluster, which signals a tendency towards interdisciplinarity. It can therefore be considered as "a valuable early sign of transformative research at interdisciplinary level" (p. 144).

As suggested by Hou et al. (2020), the influence of a study can be gauged by its role in connecting other clusters. By extracting the potentially transformative papers (i.e., the potentially revolutionary publications), we hope to uncover the likely themes of metadiscourse research in the future.

4. Results and discussion

This section first provides a detailed analysis of the co-citation network (Section 4.1). Section 4.2 analyzes the developmental stages of metadiscourse based on Shneider's theory. Finally, in Section 4.3, we identify publications with transformative potential for the development of metadiscourse studies.

4.1. The major themes in metadiscourse research

The co-citation analysis identified 131 clusters in the co-citation network, consisting of 817 nodes and 1375 links. The modularity and silhouette scores of the co-citation network in this study are 0.9166 and 0.9401 respectively, indicating that the network is sufficient for further analysis. Figure 1 illustrates the timeline view of the co-citation network output by CiteSpace, which facilitates insights into the development of metadiscourse from a diachronic perspective and shows the duration of each cluster. Of these, the earliest cluster (Cluster #5) has a mean publication year around 1998, while the youngest (Cluster #10) has a mean year around 2017.

Table 1 displays information concerning the twelve largest clusters[1] (these account for 77.7% of all metadiscourse studies retrieved), with sizes (the number of studies) ranging from 11 (Cluster #18) to 97 (Cluster #0). Other information, such as the size of analyzed clusters, Silhouette score, Averaging Year, Label with the highest score,[2] is also included.
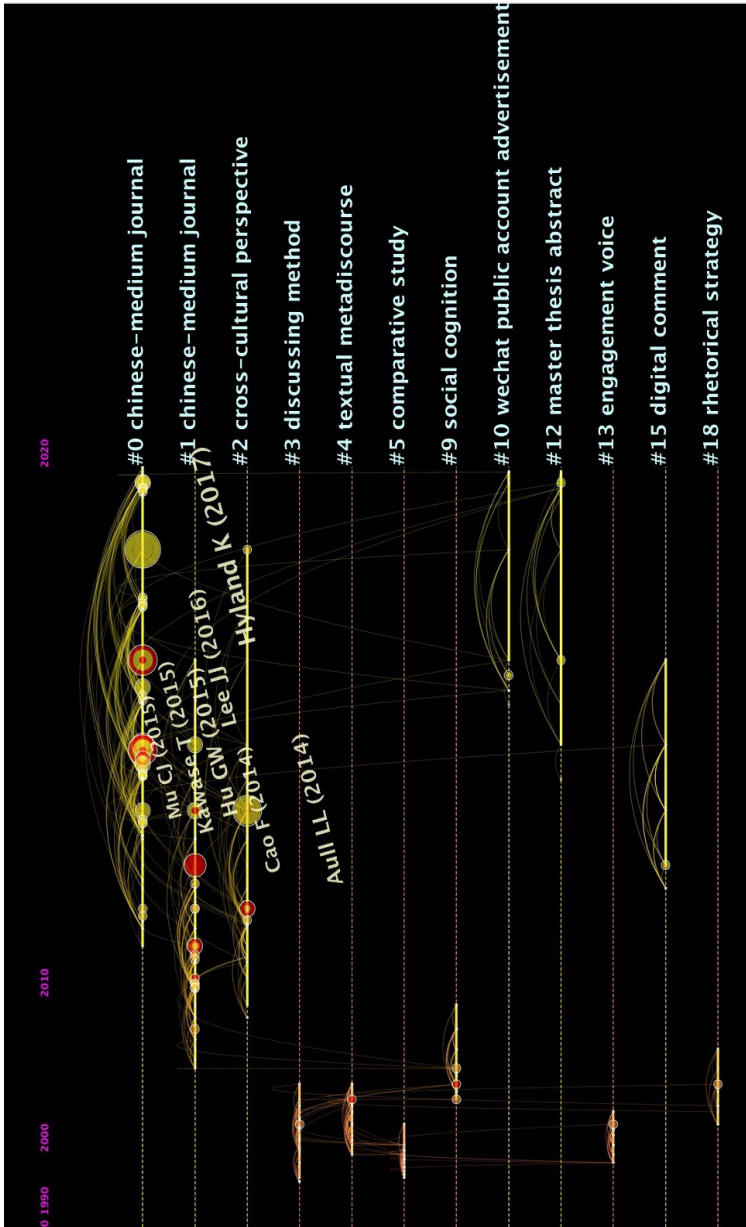
Figure 1. The timeline view of the co-citation network[3]

Cluster	Cluster Size	Silhouette Score	Mean Year	Timespan	Cluster labels (Top 10 & P value < 0.05)
#0[a]	97	0.899	2015	2012-2020	Chinese-medium journal; applied linguistics article; metadiscourse feature; interactional marker; international postgraduate business students' text; written scientific Spanish; comparative study; surprise marker; interactional metadiscourse; Spanish thesis writer
#1[a]	69	0.935	2010	2006-2014	Chinese-medium journal; applied linguistics article; metadiscourse feature; interactional marker; international postgraduate business students' text; written scientific Spanish; comparative study; Spanish thesis writer; discussion chapter; comparative analysis
#2	37	0.967	2012	2009-2017	cross-cultural perspective; authorial presence; corpus data; enhancing writing pedagogy; using corpus-based research; discussion section; online academic corpora; engagement voice; traveler forum; promotional website
#3	34	0.92	2000	1996-2005	discussing method; memory problem; computing science; corpus study; research article; surprise marker; interactional metadiscourse; academic writing; shell noun; knowledge construal
#4	27	0.962	2002	1999-2005	textual metadiscourse; national culture; academic discipline; research article
#5	26	0.931	1998	1996-2002	comparative study
#9	19	0.985	2006	2004-2008	social cognition; political communication; organizing knowledge; essay genre; discoursal resource; research article; surprise marker; interactional metadiscourse
#10	15	0.983	2017	2016-2019	WeChat public account advertisement; relational act; same degree; group discussion; exploring dominance-linked reflexive metadiscourse; research article; surprise marker; interactional metadiscourse; academic writing
#13	13	0.995	2000	1998-2002	surprise marker
#12	13	0.989	2017	2015-2019	master thesis abstract; rocky road; finish engineering student; second language; stumbling block; diachronic perspective; research article; surprise marker; interactional metadiscourse
#15	12	0.999	2014	2013-2016	digital comment
#18	11	0.923	2004	2002-2005	rhetorical strategy; modeling metadiscourse; biomedical research abstract

Table 1. Detailed information of major clusters

Cluster #0 is the largest detected cluster in the co-citation network (containing 97 studies), and it has a silhouette score of 0.899, which indicates a high level of homogeneity of themes within the cluster. The recency of this cluster (timespan: 2012-2020) and its size are evidence of the currency of the themes it encompasses. Cluster #1 is the second-largest cluster in the co-citation network, and it shares some common themes with Cluster #0 (e.g., Chinese-medium journal; applied linguistics article). The two clusters differ in that Cluster #0 contains studies focusing on specific types of metadiscourse (e.g., surprise markers and interactional metadiscourse), while Cluster #1 comprises discourse genres (e.g., Spanish thesis writing) and research methods (e.g., comparative analysis).

4.2. The developmental stages of metadiscourse studies

In this section, we present the developmental stages of research on metadiscourse over the last four decades, guided by Shneider's theory (2009). We identified three main stages by the starting year of each cluster, namely, the emerging stage (1983- about 2005), primarily focusing on introducing metadiscourse; the maturing stage (about 2005- about 2015), concerned with enlarging the territory of metadiscourse; and the flourishing stage (about 2015-now), which encompasses a broadening of the research themes and the application of the research methods established in the two preceding stages. In the following sub-sections, we provide a detailed account and discussion of the major clusters and the landmark publications identified in each stage.

4.2.1. The emerging stage (1983 – about 2005)

The emerging stage of metadiscourse studies comprises three clusters: part of Cluster #3 (*discussion of methods*), Cluster #5 (*comparative studies*), and Cluster #13 (*surprise markers*). These clusters are primarily concerned with the introduction of specific terms (e.g., *discussing method* and *knowledge construal*), the formulation of definitions (e.g., Crismore, 1984), and classification systems (e.g., Williams, 2008).

Although more typical of Stage 2, this stage also comprised themes related to research methods and text genres, such as the comparative method in Cluster #5 (e.g., Fuertes-Olivera et al., 2001) and the corpus method in Cluster #3 (e.g., Lindemann & Mauranen, 2001). A possible explanation for this discrepancy may lie in the concurrent methodological developments in corpus linguistics. That is, advancements in an allied field shaped the approach to metadiscourse research from the outset.

The three clusters are the oldest in terms of both publication timespan (1996-2005) and the mean year (2000). No landmark publication was identified at this stage. This is not unexpected, as Stage 1 research does not typically reap many citations (Shneider, 2009).

4.2.2. The maturing stage (about 2005 - about 2015)

As displayed in Table 1, Stage 2 of metadiscourse research ranges from around 2000 to 2015 and comprises seven clusters (Cluster #1, #2, #4, #9, #15, #18 and part of Cluster #3). According to Shneider (2009), this stage is "primarily concerned with broadening and deepening the

conceptualization of the focus of study, and this entails developing a richer and more sophisticated use of language to describe a broader spectrum of phenomena" (p. 221).

A central focus of metadiscourse research in these clusters concerned the refining of the classification frameworks proposed in Stage 1 or the proposal of alternatives. For instance, the classification in Vande Kopple (1985) distinguished between seven types of metadiscourse (*text connectives, code glosses, illocution markers, narrators, validity markers, attitude markers and commentaries*). This framework was further modified by Crismore, Markkanen and Steffensen (1993) in terms of introducing two higher-level categories: textual markers and interpersonal markers. Another influential framework of metadiscourse in this stage was proposed by Ädel (2006), which followed Mauranen's (1993) idea and Jakobson's (1960) functions of language and viewed metadiscourse as "text about the evolving text, or the writer's explicit commentary on her own ongoing discourse" (Ädel, 2006, p. 20). Ädel (2006) provided five clear criteria for investigating metadiscourse, i.e., intersubjectivity, non-propositionality, context-dependency, explicitness, and intra-textuality. However, among many classifications proposed in Stage 2 (and the subsequent stages), only Hyland's (2005) framework in Cluster #9 (Social cognition), which built on Thompson's (2001) distinction between interactive and interactional resources, was identified as a landmark publication (see Table 2). The framework on metadiscourse is identified as a predominant in this domain, with a high sigma value of 2.33. Interactive metadiscourse comprises self-reflexive expressions to manage the flow of information, thereby "addressing how writers guide readers by anticipating their likely reactions and needs" (Hyland, 2005, p. 44), and interactional metadiscourse encompasses linguistic resources used to make "explicit interventions to comment on and evaluate material" (Hyland, 2005, p. 44). Two subsequent publications, Hyland and Tse (2004) and Aull and Lancaster (2014), were also detected as landmark publications (as shown in Table 2).

It is necessary to point out that Ädel (2006), an influential study in metadiscourse, was not found as a landmark publication in the co-citation network for two possible reasons. The first is that Hyland's model (2005) is generally seen as the dominant model in the domain (Hyland, 2017, 2022), and the second is that in the choice of search terms, this study did not adopt "metatext" or "text reflexivity" as the search term, which is preferred by studies (e.g., Salas, 2015; Zhang, 2016), following Ädel's (2006) approach.

Another notable feature of Stage 2 concerns the endeavors to expand the terrain by proposing new themes, particularly with respect to emerging genres and languages in this domain. In the co-citation network of Stage 2, several labels regarding to the genre or part-genre under investigation (*"WeChat public account advertisement"*, *"master thesis abstract"*, *"discussion section"*, *"digital comment"*) frequently occurred (as shown in Table 1). It is also evident that a comparative study design is preferred, as the label *"comparative study"* was found to be significant in Cluster #1, LLR = 29.11. Several studies using a comparative perspective were found (Hyland, 2004; Kawase, 2015), which investigate the use of metadiscourse in different genres. Language is the other notable theme we identified in Stage 2. While metadiscourse was limited in English in Stage 1, more languages were used in Stage 2. Spanish was the second most prominent language of metadiscourse studies (e.g., Carrasco Muñoz, 2008; Meza, 2016), and this is evidenced by the high value in the log-likelihood ratio score of the cluster label "written scientific Spanish" (LLR = 29.36) in Cluster #1. Nevertheless, Stage 2 research was mainly in the domain of EAP, as evidenced by cluster labels such as "EAP lesson", "English research article", "written academic writing", etc., which confirms the finding in Hyland (2017, 2022).

As shown in Table 1, studies in Stage 2 of metadiscourse research were primarily undertaken from the following two perspectives: style-variation metadiscourse research (Cluster #1, #9, #18 and part of Cluster #2, #4 and #15); and language proficiency research (part of Cluster #2, #4 and #15). Both perspectives experienced prolonged citation bursts (between 2004 and 2016; between 1999 and 2016 respectively).

No.	Co-citation	Burst	Centrality	Sigma	Publication	Cluster
1	19	4.35	0.21	2.33	Hyland (2005). Metadiscourse: Exploring Interaction in Writing	9
2	23	4.84	0.13	1.84	Hyland and Tse (2004). Metadiscourse in academic writing: A reappraisal	4
3	19	4.78	0.17	2.09	Aull & Lancaster (2014). Written Communication	2

Table 2. Landmark publications in style-variation metadiscourse research (ranked by year).

Specifically, style-variation metadiscourse research (2002-2014) was most prominent in Cluster #1, comprising 69 articles. Two research approaches were salient: cross-disciplinary and cross-genre approaches. The cross-disciplinary approach received a high citation in the WoS core collection (e.g., Hyland & Tse,

2004; Hyland, 2004). For example, the first disciplinary-related metadiscourse publication, Hyland (1998), explored the importance of rhetorical context for the appropriate use of metadiscourse. A subsequent highly cited study, Hu and Cao (2015) identified robust evidence for disciplinary practices in the use of interactional metadiscourse in hard and soft disciplines.

Genre analysis, pioneered by Swales (2011), focuses on language in specific text types. Studies following this approach explored new genres, such as the undergraduate textbook, abstracts, slogans, and headlines. Two primary genres involved in metadiscourse research are business and academic genres (Hyland, 2017, 2022). It is evident that the academic genre is more dominant, as shown by the significant academic-pertinent cluster labels listed in Table 1 (with p values below 0.05). Regarding the cross-genre approaches, Hu and Cao (2015) pointed to the paucity of research using this approach, compared to cross-disciplinary research. Nevertheless, we found that cluster labels for the cross-genre approach possessed higher LLR scores than those for the cross-disciplinary approach, which signals the higher impact of this approach in recent years (e.g., Carrió-Pastor, 2019; Labrador et al., 2014).

Research from the perspective of "proficiency metadiscourse research" typically considers two different variables in the analysis of metadiscourse: the effect of different native language backgrounds, and different writing proficiency levels (such as undergraduate essays and journal papers) (e.g., Campbell et al., 2012; Hu & Liu, 2018). The contrastive analysis, prominent in both Clusters #2 and #4, was firstly undertaken by Vergaro (2002). This was followed by other studies which identified the differences in metadiscourse usage between English and languages such as Brazilian Portuguese (Hirano, 2009), Spanish (Mur-Dueñas, 2011), Italian (Vergaro, 2005), Chinese (Crosthwaite & Jiang, 2017; Mu et al. 2015) and Malaysian (Jomaa & Alia, 2019).

Proficiency-focused studies, represented by the cluster label "L2 ability", received a high score in Cluster #0. A landmark publication (Aull & Lancaster, 2014, with a sigma score of 2.09) identified on this theme compares the use of stance expressions (one category of metadiscourse) in academic writing produced by students and academics. Several studies were identified on this theme (Hinkel, 2005; Jiang, 2015; Mur-Dueñas, 2011) that considered the implications of metadiscourse research for language teaching, and which usually employ a contrastive approach (e.g., high-level students compared with low-level students).

With two notable exceptions, the primary focus of metadiscourse research in Stage 2 was in accordance with Shneider's (2009) theoretical framework, that is, expanding the territory of metadiscourse and establishing new research themes and approaches. The exceptions concerned the absence of innovative analytical tools for analyzing metadiscourse, and the limitation of the discourse type studied to academic writing. In our view, these limitations have hindered innovative developments in the field. Reflecting on possible future lines of enquiry, Liu and Hu (2021) encouraged future metadiscourse research to "go beyond corpus-based textual analysis to capitalize on other methodological enablers, such as experimental procedures, ethnographic tools, and eye-tracking technologies" (p.111).

4.2.3. The flourishing stage (about 2015-)

Table 3 presents detailed information relating to Stage 3 research generated by CiteSpace. Three clusters were found: part of Cluster #0 (Chinese-medium journal), part of Cluster #10 (WeChat public account advertisement), and Cluster #15 (digital comment). Of these, Cluster #0 was the most prominent cluster with 97 publications. The oldest cluster in Stage 3 began in 2013 (Cluster #15), and the latest cluster (which began in 2016) still enjoyed high impact till 2020. No landmark publications were found, which indicates the absence of a dominant research theme.

As shown in Table 3, studies in Stage 3 center on two dominant perspectives, namely "new media" and "applied linguistics". The new media perspective (part of Cluster #10) attracted growing attention (Blal & Sturman, 2014), and the cluster labels show the expansion of metadiscourse-related studies to new registers, such as WeChat advertisements, fiction, websites, and reports (e.g., González, 2005; Vásquez, 2015; Zhang, 2016).

The most influential studies in applied linguistics are in Cluster #0, which is mainly concerned with analyzing the author-reader connection, the construction of author stance and knowledge construction (see Table 5). Themes that attracted considerable attention included authorial stance, shell nouns, and writing pedagogy; new themes included intensification strategies and dominance-linked reflexive metadiscourse (i.e., reflexive metadiscourse used to achieve dominance in interaction). The most highly cited study in Cluster #0, however, was Hyland's (2017) review of the development of metadiscourse. According to Miranda and Garcia-Carpintero (2018), review papers tend to be cited (on average) three times

more frequently than regular research articles and the findings in this study confirm this.

It is interesting to note that corpus-based and corpus-driven studies, the common research approaches in Stage 2, are also dominant in Stage 3 labels (mainly in Cluster #15), which indicates the prevailing influence of Stage 2. According to Shneider (2009), Stage 3 research perspectives are strongly influenced by Stage 2, albeit with adjustments. In reflection of the dominance of corpus linguistic methods, the second most highly cited Stage 3 study was the AntConc software (Anthony, 2005).

However, contrary to Shneider (2009), new theory was found to emerge in metadiscourse research in Stage 2. Based on the work of stance nouns (Charles, 2003) in academic writing, Jiang and Hyland (2018) addressed the interactional dimension of nouns in academic writing and proposed the existence of "metadiscursive nouns". The concept refers to "a sub-set of abstract nouns and [is] distinguished from them by their unspecific semantic meaning" (Jiang & Hyland, 2015, p. 4). The study found that "metadiscursive noun + post-nominal clause" is one of the most frequently used patterns. This emerging noun pattern represents a method of "organizing discourse into a cohesive flow of information and of constructing a stance towards it" (Hyland & Jiang, 2018, p. 20). A potential influence brought by this study is that it may cause a curvilinear development of metadiscourse. As suggested by Kuhn (1962), scientific development does not proceed linearly. In this light, this study thus can contribute to giving rise to a new wave of research on metadiscourse and reinvigorating the field, according to Shneider (2009). Another notable finding concerns the introduction of a new analytical tool, namely the Authorial Voice Analyzer (in Yoon, 2017). This tool allows users to automatically detect interactional metadiscourse in the source text. Although according to Shneider (2009) the emergence of analytical tools typically occurs at Stage 2, a possible reason for its delayed appearance is the low demand for technological advancements due to the limited scale of the metadiscourse research field. However, it is necessary to point out that the application did not attract citations, as we found only two studies with the combination of keywords "metadiscourse" and "authorial voice analyzer" in the WoS (Lim, 2019; Yoon, 2018).

No.	Perspective	Cluster No./ Mean Year	Number of publications	Extracted cluster labels related to the perspective (Top 10 by LLR score, p-level < 0.05)
1	New media approach	Part of #10(2017)	15	WeChat public account advertisement (24.57); relational act (24.57); group discussion (9.45); exploring dominance-linked reflexive metadiscourse (9.45)
2	Applied linguistics approach	Part of #0 (2015); Part of #15 (2014)	110	Chinese-medium journal (27.46); applied linguistics article (27.46); metadiscourse feature (24.36); interactional marker (24.36); surprise marker (19.67); authorial presence (18.19); EAP lesson (15.11); interactive metadiscourse (12.04); shell noun (10.99); knowledge construal (10.99)

Table 3. Research perspectives identified in Stage 3.

4.3. Publications with the strongest transformative potential

Publications with transformative potential refer to those containing innovations that had not been previously introduced in the studies cited (Olmeda-Gómez et al., 2019). In this study, SVA in Citespace was used to identify the recent publications in metadiscourse of transformative potential. The specific analysis follows the criterion of Olmeda-Gómez (2019) and uses the geometric mean of the values for three separate metrics (presented in Section 3.2) to select the studies with the highest transformative potentials. The analysis identified 13 publications with transformative potential (Table 4) and the studies with the highest transformative potential are presented in Table 5. Notably, the dominance of Stage 3 publications (eleven in total) is evidence of the more innovative nature of Stage 3 metadiscourse research. This finding is not consistent with Shneider's (2009) account of Stage 3 research which constitutes only "relatively minor but still highly useful alterations of the research methods while adapting them for new tasks." (p. 220). Thus, discrepancies were identified in the development of metadiscourse research with respect to Shneider's (2009) evolutionary model, which indicates a different developmental trajectory of metadiscourse.

Also of note, papers with the highest modularity change rate in each stage are all located in the proximity of a stage transition. For instance, the transformative study of Hewings and Hewings (2002) appeared at the beginning of Stage 2, and the second transformative study (Hyland & Jiang, 2018) was located at the beginning of Stage 3. This suggests that transformative studies may function as an indicator of the transition to a new developmental stage in a domain.

Published	Modularity change rate	Cluster linkage	Centrality divergence	Stage No.	References
2002	3.3621	0.4	0.2976	1	Hewings & Hewings
2006	2.2807	0.2667	0.8316	2	Tse & Hyland
2015	-0.1274	0	0.1994	3	Pujol Dahme Ana
2015	0.00	0.00	0.00	3	Lee & Deakin
2016	0.00	0.00	0.00	3	Hu & Cao
2018	1.121	0.0303	0.5091	3	Ädel
2018	2.3356	0.0303	0.028	3	Akbas & Hardman
2018	2.5095	0.0606	0.0946	3	Ho & Li
2018	1.7137	0.0303	0.0773	3	Ho
2018	5.8082	0.0909	0.1646	3	Hyland & Jiang
2018	1.6791	0.0303	0.0832	3	Jiang & Hyland
2018	2.3356	0.0303	0.028	3	Mozayan, Allami & Fazilatfar

Table 4. Studies detected with transformative potential in the co-citation network.

Published	Modularity change rate	Cluster linkage	Centrality divergence	Stage No.	References
2002	3.3621	0.4	0.2976	1	Hewings & Hewings
2006	2.2807	0.2667	0.8316	2	Tse & Hyland
2018	5.8082	0.0909	0.1646	3	Hyland & Jiang

Table 5. The three most transformative studies detected in the co-citation network.

Table 5 displays the top three studies with the highest transformative potential in each developmental stage. The earliest publication is Hewings and Hewings (2002) from Cluster #5, appearing in the earliest stage of the evolutionary model. This study explores the "it" structure through a metadiscourse perspective in two computerized corpora, i.e., in published journal papers in the field of Business Studies and in MBA student dissertations written by non-native English speakers. The study investigated the main interpersonal roles in the hedging of the "it" structure (i.e., *marking the writer's attitude*, *emphasis*, and *attribution*) and found that students often use more hedging in articulating their claims. The retrieved information from Citespace shows that the study connects Biber et al. (1999) in Cluster #5 with Hyland (1998) in Cluster #13. This application of the register variation theory of Biber et al. (1999) in the exploration of metadiscourse constitutes an innovative step.

The second publication with transformative potential in Stage 2 is Tse and Hyland (2006). The study investigated how the pragmatic purpose in the use of metadiscourse markers varies in book reviews in three disciplines. The

analysis shows that the study strengthens the connection between Hyland (2002a) and Hyland (2002b), and also creates new links between Hyland (2002a) and Hyland (2004). An in-depth inspection revealed that Tse and Hyland (2006) first applied corpus approaches to analyze metadiscourse in philosophy book reviews, and this study functions to create a new link between Hyland (2004) and Hyland (2002b). It also used the interview approach to analyze rhetorical features of metadiscourse, which contributed to establishing the link between Hyland (2004) and Hyland (2002a). This inspired several studies in the generated co-citation network using interviews as an approach to explore the metadiscourse (e.g., Bogdanovic & Mirovic, 2018; Peng & Zheng, 2021).

The last potential transformative study is by Hyland and Jiang (2018). This study shared a strong link with Hyland (2017), Kawase (2015), and Hong and Cao (2014). The study analyzed the use of metadiscourse from a diachronic perspective and explored whether, and to what extent, metadiscourse has changed in professional writing in different disciplines over the past 50 years.

In our co-citation network, eight out of 13 transformative potential publications were found to be published in 2018. The topic of those studies signals the direction of future themes of metadiscourse research, according to the SVA results. Emerging focuses on genres like email (Ho, 2018), teacher feedback on student writing (Ädel, 2018) and qualitative studies (Mozayan et al., 2018) have extended the territory of metadiscourse. New directions such as the subdisciplinary perspective (Mozayan et al., 2018) or diachronic perspective (Hyland & Jiang, 2018) also emerged in this domain. Although studies that were detected to have the most transformative potential still focus on EAP, a broader array of perspectives has emerged (e.g., interpretation (Fu, 2017), MOOC (Zhang & Sheng, 2021)) and will probably expand the scope of metadiscourse.

5. Conclusion

Adopting a bibliometric approach, this study provided a diachronic and systematic review of the developmental features of metadiscourse studies over the last four decades and tracked the evolution of a research topic evolving into a popular research domain. The co-citation analysis revealed twelve major clusters, three landmark studies and 13 potentially transformative studies.

Using Shneider's (2009) theoretical framework of scientific development, we identified three developmental stages that metadiscourse research has undergone over the last four decades: the emerging stage (introducing the metadiscourse, 1983- about 2000), the maturing stage (expanding the research territory, about 2000- about 2015), and the flourishing stage (creating new insights into the researching object, about 2015-). The developmental trajectory of metadiscourse research was found to largely coincide with characteristics foreseen in Shneider's (2009) evolutionary model. Some unique features were identified, which may pertain to the recency of this emerging research field, and these can contribute to amending this theoretical framework of scientific development.

A second contribution of the study concerns the creative application of structural variation analysis to identify studies of transformative potential in the development of metadiscourse research. This approach helps to pinpoint the studies that play a pivotal role in ushering in the next developmental stage of this field soon after their emergence, which helps to solve the issue of heavy reliance on co-citation analysis. The ability to identify the features of a study that contribute to transformative potential, as demonstrated through our analysis, has the following two potential implications. Firstly, researchers are able to foresee such features in the design of future research, which can increase the likelihood of the study achieving transformative potential and advancing the research field. Secondly, the structural variation analysis employed in this study can also be used to help journal editors to identify manuscripts with transformative potential before the peer review process, thereby facilitating their consideration of this factor in making decisions on the acceptance of a paper in the reviewing process.

Admittedly, factors external to the research process (such as societal trends) also influence the development of a research field or scientific domain. In recognition of this, future research might explore this through the integration of particular variables into the study design, where possible. Another limitation concerns the use of a purpose-built dataset of publications and their unique references collected from WoS core collections. A broader dataset (such as the full dataset of WoS or a dataset compiled with Google Scholar) may provide more comprehensive coverage of metadiscourse studies. Finally, citations are dynamic in nature. They are subject to variation and are sensitive to the influence of various factors, including research scale, social influence, and self-citation (Bornmann &

Daniel, 2008); future studies in this line may seek to account for some of these variables when measuring co-citation and other relevant metrics.

References

Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. John Benjamins.

Ädel, A. (2018). Variation in metadiscursive "you" across genres: From research articles to teacher feedback. *Educational Sciences: Theory & Practice*, 778-798. https://doi.org/10.12738/estp.2018.4.0037

Ädel, A., & Mauranen, A. (2010). Metadiscourse: Diverse and divided perspectives. *Nordic Journal of English Studies*, *9*(2), 1-11. https://doi.org/10.35360/njes.215

Anthony, L. (2005). Antconc: Design and development of a freeware Corpus Analysis Toolkit for the technical writing classroom. *IPCC 2005. Proceedings. International Professional Communication Conference, 2005*, 729-737. https://doi.org/10.1109/ipcc.2005.1494244

Aryadoust, V., & Ang, B. H. (2019). Exploring the frontiers of Eye Tracking Research in language studies: A novel co-citation Scientometric Review. *Computer Assisted Language Learning*, *34*(7), 898-933. https://doi.org/10.1080/09588221.2019.1647251

Aull, L. L., & Lancaster, Z. (2014). Linguistic markers of stance in early and advanced academic writing. *Written Communication*, *31*(2), 151-183. https://doi.org/10.1177/07410883145270 55

Azam, A., Ahmed, A., Wang, H., Wang, Y., & Zhang, Z. (2021). Knowledge structure and research progress in wind power generation (WPG) from 2005 to 2020 using CiteSpace based scientometric analysis. *Journal of Cleaner Production*, *295*, 126496. https://doi.org/10.1016/j.jclepro.2021.126496

Biber, D., Conrad, S., & Leech, G. N. (1999). *Longman student grammar of spoken and written English (Vol. 2)*. Longman.

Blal, I., & Sturman, M. C. (2014). The differential effects of the quality and quantity of online reviews on hotel room sales. *Cornell Hospitality Quarterly*, *55*(4), 365-375. https://doi.org/10.1177/19389655 14533419

Bogdanovic, V., & Mirovic, I. (2018). Young researchers writing in ESL and the use of metadiscourse: Learning the ropes. *Educational Sciences: Theory & Practice, 18*(4), 813-830. https://doi.org/10.12738/estp.2018.4.0031

Bornmann, L., & Daniel, H. D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45-80. https://doi.org/10.1108/00220410810844 150

Campbell, T., Oh, P. S., & Neilson, D. (2012). Discursive modes and their pedagogical functions in model-based inquiry (MBI) classrooms. *International Journal of Science Education*, *34*(15), 2393-2419. https://doi.org/10.1080/09500693.2012.704552

Carrasco Muñoz, H. (2008). Discursos y metadiscursos mapuches. *Estudios Filológicos*, *43*, 39-53. https://doi.org/10.4067/s0071-17132 008000100003

Carrió-Pastor, M. L. (2019). Authorial engagement in business emails. A cross-cultural analysis of attitude and engagement markers. In C. Sancho Guinda (Ed.), *Engagement in professional genres* (pp. 47-65). John Benjamins. https://doi.org/10.1075/pbns.301.03car

Charles, M. (2003). 'This mystery…': A corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes, 2*(4), 313-326. https://doi.org/10.1016/s1475-1585(03) 00048-1

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, *57*(3), 359-377. https://doi.org/10.1002/asi.20317

Chen, C. (2012). Predictive effects of structural variation on Citation Counts. *Journal of the American Society for Information Science and Technology, 63*(3), 431-449. https://doi.org/10.1002/asi.21694

Chen, C. (2013). Hindsight, insight, and foresight: a multi-level structural variation approach to the study of a scientific field. *Technology Analysis & Strategic Management*, *25*(6), 619-640. https://doi.org/10.1080/09537325.2013.801949

Chen, C. (2016). *CiteSpace: A practical guide for mapping scientific literature*. Nova Science Pub Inc.

Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *Journal of the American Society for Information Science and Technology*, *61*, 1386-1409. https://doi.org/10.1002/asi.21309

Chen, C., & Song, M. (2019). Visualizing a field of research: A methodology of systematic scientometric reviews. *PLOS ONE*, *14*(10), e0223994. https://doi.org/10.1371/journal.pone.0223994

Crismore, A. (1984). The rhetoric of textbooks: Metadiscourse. *Journal of Curriculum Studies*, *16*(3), 279-296. https://doi.org/10.1080/0022027840160306

Crismore, A. (1989). *Talking with readers: Metadiscourse as rhetorical act*. Peter Lang.

Crismore, A., Markkanen, R., & Steffensen, M. S. (1993). Metadiscourse in persuasive writing. *Written Communication*, *10*(1), 39-71. https://doi.org/10.1177/0741088393010001002

Crosthwaite, P., & Jiang, K. (2017). Does EAP affect written L2 academic stance? A longitudinal learner corpus study. *System*, *69*, 92-107. https://doi.org/10.1016/j.system.2017.06.010

D'Angelo, L., & Consonni, S. (2020). A tale of three waves: Or, concerning the history and theory of metadiscourse. *Ibérica, Journal of the European Association of Languages for Specific Purposes*, *40*, 13-34. https://doi.org/10.17398/2340-2784.40. 13

De Schryver, G. M. (2009). Bibliometrics in lexicography. *International Journal of Lexicography*, *22*(4), 423-465. https://doi.org/10.1093/ijl/ecp027

Dong, J., & Buckingham, L. (2018). The collocation networks of stance phrases. *Journal of English for Academic Purposes*, *36*, 119-131. https://doi.org/10.1016/j.jeap.2018.10.004

Dong, J., Zhao, Y., & Buckingham, L. (2022). Charting the landscape of data-driven learning using a bibliometric analysis. *ReCALL*, 1-17. https://doi.org/10.1017/s0958344022000222

Fu, R. (2017). Metadiscourse and coherence in interpreting. *Babel, 63*, 846-860. https://doi.org/10.1075/babel.00017.ron

Fu, Y., Wang, H., Guo, H., Bermúdez-Margaretto, B., & Domínguez Martínez, A. (2021). What, where, when and how of visual word recognition: A Bibliometrics Review. *Language and Speech*, *64*(4), 900-929. https://doi.org/10.1177/00238 30920974710

Fuertes-Olivera, P. A., Velasco-Sacristán, M., Arribas-Baño, A., & Samaniego-Fernández, E. (2001). Persuasion and advertising English: Metadiscourse in slogans and headlines. *Journal of Pragmatics*, *33*(8), 1291-1307. https://doi.org/10.1016/s0378-2166(01)80026-6

González, R. A. (2005). Textual metadiscourse in commercial websites. *Ibérica, Journal of the European Association of Languages for Specific Purposes*, *9*, 33-52.

Harris, Z. (1959). The transformational model of language structure. *Linguist*, *1*(1), 27-29. https://doi.org/https://www.jstor.org/stable/30022172

Hewings, M., & Hewings, A. (2002). "It is interesting to note that…": A comparative study of anticipatory 'it' in student and published writing. *English for Specific Purposes*, *21*(4), 367-383. https://doi.org/10.1016/s0889-4906(01)00016-3

Hinkel, E. (2005). Hedging in academic text in English. *Teaching Academic L2 Writing*, *15*(1), 429-459. https://doi.org/10.4324/9780429437946-15

Hirano, E. (2009). Research article introductions in English for specific purposes: A comparison between Brazilian Portuguese and English. *English for Specific Purposes*, *28*(4), 240-250. https://doi.org/10.1016/j.esp.2009.02.001

Ho, M. H. C., Liu, J. S., & Chang, K. C. T. (2017). To include or not: The role of review papers in citation-based analysis. *Scientometrics*, *110*(1), 65-76. https://doi.org/10.1007/s11192-016-2158-0

Ho, V. (2018). Using metadiscourse in making persuasive attempts through workplace request emails. *Journal of Pragmatics*, *134*, 70-81. https://doi.org/10.1016/j.pragma.2018.06.015

Hong, H., & Cao, F. (2014). Interactional metadiscourse in young EFL learner writing. *International Journal of Corpus Linguistics*, *19*(2), 201-224. https://doi.org/10.1075/ijcl.19.2.03hon

Hou, J., Yang, X., & Chen, C. (2020). Measuring researchers' potential scholarly impact with structural variations: Four types of researchers in information science (1979-2018). *PLOS ONE*, *15*(6). https://doi.org/10.1371/journal.pone.0234347

Hu, G., & Cao, F. (2015). Disciplinary and paradigmatic influences on interactional metadiscourse in research articles. *English for*

Specific Purposes, *39*, 12-25. https://doi.org/10.1016/j.esp.2015.03.002

Hu, G., & Liu, Y. (2018). Three minute thesis presentations as an academic genre: A cross-disciplinary study of genre moves. *Journal of English for Academic Purposes*, *35*, 16-30. https://doi.org/10.1016/j.jeap.2018.06.004

Hyland, K. (1998). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics*, *30*(4), 437-455. https://doi.org/10.1016/s0378-2166(98)00009-5

Hyland, K. (2002a). Directives: Argument and Engagement in Academic Writing. *Applied Linguistics*, *23*(2), 215-239. https://doi.org/10.1093/applin/23.2.215

Hyland, K. (2002b). What do they mean? Questions in academic writing. *Text: an Interdisciplinary Journal for the Study of Discourse*, *22*(4), 529-557. https://doi.org/10.1515/text.2002.021

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum.

Hyland, K. (2017). Metadiscourse: What is it and where is it going? *Journal of Pragmatics*, *113*, 16-29. https://doi.org/10.1016/j.pragma.2017.03.007

Hyland, K., & Jiang, F. (2018). "In this paper we suggest": Changing patterns of disciplinary metadiscourse. *English for Specific Purposes*, *51*, 18-30. https://doi.org/10.1016/j.esp.2018.02.001

Hyland, K., & Jiang, F. (2021). Delivering relevance: The emergence of ESP as a discipline. *English for Specific Purposes*, *64*, 13-25. https://doi.org/10.1016/j.esp.2021.06.002

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *The essential Hyland: Studies in Applied Linguistics*, 156-177. https://doi.org/10.5040/9781350037939.0016

Hyland, K., Wang, W., & Jiang, F. (2022). Metadiscourse across languages and genres: An overview. *Lingua*, *265*, 103205. https://doi.org/10.1016/j.lingua.2021.103205

Jakobson, R. (1960). Closing statement: Linguistics and poetics. In T. A. Sebeok (Ed.), *Style in language* (pp. 350-377). The Technology Press of Massachusetts & John Wiley.

Jiang, F. (2015). Nominal stance construction in L1 and L2 students' writing. *Journal of English for Academic Purposes*, *20*, 90-102. https://doi.org/10.1016/j.jeap.2015.07.002

Jiang, F., & Hyland, K. (2015). 'The fact that': Stance nouns in disciplinary writing. *Discourse Studies*, *17*(5), 529-550. https://doi.org/10.1177/1461445615590719

Jomaa, N. J., & Alia, M. (2019). Functional analyses of metadiscourse markers in L2 students' academic writing. *Arab World English Journal*, *10*(1), 361-381. https://doi.org/10.24093/awej/vol10no1.30

Kawase, T. (2015). Metadiscourse in the introductions of PhD theses and research articles. *Journal of English for Academic Purposes*, *20*, 114-124. https://doi.org/10.1016/j.jeap.2015.08.006

Khedri, M., Chan, S. H., & Tan, B. H. (2013). Cross-disciplinary and cross-linguistic perspectives on metadiscourse in academic writing. *Southern African Linguistics and Applied Language Studies*, *31*(1), 129-138. https://doi.org/10.2989/16073614.2013.793957

Kopple, W. J. (1985). Some exploratory discourse on metadiscourse. *College Composition and Communication*, *36*(1), 82-93. https://doi.org/10.2307/357609

Kuhn, T. S. (1962). *The structure of scientific revolutions*. The University of Chicago Press.

Labrador, B., Ramón, N., Alaiz-Moretón, H., & Sanjurjo-González, H. (2014). Rhetorical structure and persuasive language in the subgenre of online advertisements. *English for Specific Purposes*, *34*, 38-47. https://doi.org/10.1016/j.esp.2013.10.002

Lei, L., & Liu, D. (2018). Research trends in Applied Linguistics from 2005 to 2016: A bibliometric analysis and its implications. *Applied Linguistics*, *40*(3), 540-561. https://doi.org/10.1093/applin/amy003

Li, J., Goerlandt, F., & Reniers, G. (2021). An overview of scientometric mapping for the safety science community: Methods, tools, and framework. *Safety Science*, *134*, 105093. https://doi.org/10.1016/j.ssci.2020.105093

Lim, J. (2019). An investigation of the text features of discrepantly-scored ESL essays: A mixed methods study. *Assessing Writing*, *39*, 1-13. https://doi.org/10.1016/j.asw.2018.10.003

Lindemann, S., & Mauranen, A. (2001). "It's just real messy": The occurrence and function of just in a corpus of academic speech. *English for Specific Purposes*, *20*, 459-475. https://doi.org/10.1016/s0889-4906(01)00026-6

Liu, J., Fan, L., & Yin, H. (2019). A bibliometric analysis on cognitive processing of emotional words. *Digital Scholarship in the Humanities*, *35*(2), 353-365. https://doi.org/10.1093/llc/fqz025

Liu, Y., & Hu, G. (2021). Mapping the field of

English for specific purposes (1980-2018): A co-citation analysis. *English for Specific Purposes*, *61*, 97-116. https://doi.org/10.1016/j.esp.2020.10.003

Martínez-Gómez, A. (2015). Bibliometrics as a tool to map uncharted territory: A study on non-professional interpreting. *Perspectives*, *23*(2), 205-222. https://doi.org/10.1080/0907676x.2015.1010550

Mauranen, A. (1993). *Cultural differences in academic rhetoric: A textlinguistic study*. Peter Lang.

Merigó, J. M., Cancino, C. A., Coronado, F., & Urbano, D. (2016). Academic research in innovation: A country analysis. *Scientometrics*, *108*(2), 559-593. https://doi.org/10.1007/s11192-016-1984-4

Meza, P. (2016). El posicionamiento estratégico del autor en artículos de investigación Una propuesta para su estudio. *Forma y Función*, *29*(2), 111-134. https://doi.org/10.15446/fyf.v29n2.60191

Miranda, R., & Garcia-Carpintero, E. (2018). Overcitation and overrepresentation of review papers in the most cited papers. *Journal of Informetrics*, *12*(4), 1015-1030. https://doi.org/10.1016/j.joi.2018.08.006

Mozayan, M., Allami, H., & Fazilatfar, A. M. (2018). Metadiscourse features in medical research articles: Subdisciplinary and paradigmatic influences in English and Persian. *Journal of Research in Applied Linguistics*, *9*(1), 83-104. https://doi.org/10.22055/rals.2018.13406

Mu, C., Zhang, L. J., Ehrich, J., & Hong, H. (2015). The use of metadiscourse for knowledge construction in Chinese and English research articles. *Journal of English for Academic Purposes*, *20*, 135-148. https://doi.org/10.1016/j.jeap.2015.09.003

Mur-Dueñas, P. (2011). An intercultural analysis of metadiscourse features in research articles written in English and in Spanish. *Journal of Pragmatics*, *43*(12), 3068-3079. https://doi.org/10.1016/j.pragma.2011.05.002

Olmeda-Gómez, C., Romá-Mateo, C., & Ovalle-Perandones, M.-A. (2019). Overview of trends in Global Epigenetic Research (2009-2017). *Scientometrics*, *119*(3), 1545-1574. https://doi.org/10.1007/s11192-019-03095-y

Peng, J. E., & Zheng, Y. (2021). Metadiscourse and voice construction in discussion sections in BA theses by Chinese University Students majoring in English. *SAGE Open*, *11*(2), 215824402110088. https://doi.org/10.1177/215

82440211008870

Pérez-Llantada, C. (2015). Genres in the forefront, languages in the background: The scope of genre analysis in language-related scenarios. *Journal of English for Academic Purposes*, *19*, 10-21. https://doi.org/10.1016/j.jeap.2015.05.005

Salas, M. D. (2015). Reflexive metadiscourse in research articles in Spanish: Variation across three disciplines (linguistics, economics and medicine). *Journal of Pragmatics*, *77*, 20-40. https://doi.org/10.1016/j.pragma.2014.12.006

Schiffrin, D. (1980). Meta-talk: Organizational and evaluative brackets in discourse. *Sociological Inquiry*, *50*(3-4), 199-236. https://doi.org/10.1111/j.1475-682x.1980.tb00021.x

Sebastian, Y., & Chen, C. (2021). The boundary-spanning mechanisms of Nobel prize winning papers. *PLOS ONE*, *16*(8). https://doi.org/10.1371/journal.pone.0254744

Shneider, A. M. (2009). Four stages of a scientific discipline; four types of scientist. *Trends in Biochemical Sciences*, *34*(5), 217-223. https://doi.org/10.1016/j.tibs.2009.02.002

Shu, S., & Liu, Y. (2021). Looking back to move forward: A bibliometric analysis of consumer privacy research. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(4), 727-747. https://doi.org/10.3390/jtaer16040042

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265-269. https://doi.org/10.1002/asi.4630240406

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Thompson, G. (2001). Interaction in academic writing: Learning to argue with the reader. *Applied Linguistics*, *22*(1), 58-78. https://doi.org/10.1093/applin/22.1.58

Tian, E., & Wise, N. (2019). An Atlantic divide? Mapping the knowledge domain of European and North American-based Sociology of Sport, 2008-2018. *International Review for the Sociology of Sport*, *55*(8), 1029-1055. https://doi.org/10.1177/1012690219878370

Tse, P., & Hyland, K. (2006). 'So what is the problem this book addresses?': Interactions in academic book reviews. *Text & Talk - An Interdisciplinary Journal of Language, Discourse Communication Studies*, *26*(6), 767-790. https://doi.org/10.1515/text.2006.031

Vergaro, C. (2002). "Dear sirs, what would you do

if you were in our position?". Discourse strategies in Italian and English money chasing letters. *Journal of Pragmatics*, *34*(9), 1211-1233. https://doi.org/10.1016/s0378-2166(02)00046-2

Vergaro, C. (2005). 'Dear Sirs, I hope you will find this information useful': discourse strategies in Italian and English 'For Your Information' (FYI) letters. *Discourse Studies, 7*(1), 109-135. https://doi.org/10.1177/1461445605048769

Vásquez, C. (2015). "Don't even get me started…": Interactive metadiscourse in online consumer reviews. *Digital Business Discourse*, 19-39. https://doi.org/10.1057/9781137405579_2

Williams, J. M. (2008). *Style: Ten lessons in clarity and grace*. Manitoba Education, Media Production Services Unit.

Yeung, A. W., Goto, T. K., & Keung Leung, W. (2017). A bibliometric review of research trends in neuroimaging. *Current Science*, *112*(04), 725. https://doi.org/10.18520/cs/v112/i04/725-734

Yoon, H. J. (2017). Textual voice elements and voice strength in EFL argumentative writing. *Assessing Writing*, *32*, 72-84. https://doi.org/10.1016/j.asw.2017.02.002

Yoon, H. J. (2018). The development of ESL writing quality and lexical proficiency: Suggestions for assessing writing achievement. *Language Assessment Quarterly*, *15*(4), 387-405. https://doi.org/10.1080/15434303.2018.1536756

Yoon, H. J., & Römer, U. (2020). Quantifying disciplinary voices: An automated approach to interactional metadiscourse in successful student writing. *Written Communication*, *37*(2), 208-244. https://doi.org/10.1177/0741088319898672

Zhang, D., & Sheng, D. (2021). EFL lecturers' metadiscourse in Chinese university moocs across course types. *Corpus Pragmatics*, *5*(2), 243-270. https://doi.org/10.1007/s41701-021-00098-0

Zhang, M. (2016). A multidimensional analysis of metadiscourse markers across written registers. *Discourse Studies*, *18*(2), 204-222. https://doi.org/10.1177/1461445615623907

Zhang, M., Pan, H., Chen, X., & Luo, T. (2015). Mapping discourse analysis in translation studies via Bibliometrics: A survey of journal publications. *Perspectives*, *23*(2), 223-239. https://doi.org/10.1080/0907676x.2015.1021260

Jihua Dong (ORCID: 0000-0001-7864-2319) is Professor, Qilu Young Scholar, and Taishan Young Scholar in the School of Foreign Languages and Literature at Shandong University, China. Her research interests include corpus linguistics, corpus-based teaching, academic writing, and discourse analysis. She has previously published in *English for Specific Purposes*, *International Journal of Corpus Linguistics*, *Journal of English for Academic Purposes*, *ReCALL*, and *System*, among others.

Shuai Dong (ORCID: 0000-0001-9358-2158) is a doctoral student in Languages and Linguistics at the School of Foreign Languages and Literature at Shandong University. He has published in *Waiyu Yanjiui*, *Waiyu Xuekan*, etc. His current research interest is EAP, corpus linguistics, discourse analysis, and bibliometric analysis.

Louisa Buckingham (ORCID: 0000-0001-9423-0664) lectures in applied linguistics at the University of Auckland. She has published on topics related to corpus informed discourse analysis, language learning and sociolinguistics. She has published in various journals including *TESOL Quarterly*, *System*, *Journal of English for Academic Purposes*, *English for Specific Purposes*, and *Journal of Multilingual and Multicultural Development*.

NOTES

[1] According to Chen's (2016) criteria, this study treats clusters containing more than ten articles as large clusters.

[2] Due to the limited studies scrutinized in this study, the keywords of Cluster #0 and Cluster #1 generated by the LLR algorithm were found to overlap in a few cluster labels, for instance Chinese medium journal.

[3] Each cluster is shown by a horizontal line with nodes/rings of varying sizes based on their co-citation statistics. In this study, recognizing that metadiscourse studies are predominantly published after 2010, we used a fish-eye lens function of Citespace to assist the interpretation of the co-citation result. The purple ring in the Figure 1 signals the greater number of publications in this domain in recent years, e.g., Hyland (2005), whose framework is predominantly used in metadiscourse research, which is also detected as a high sigma value publication.