



ARTICLE



<https://doi.org/10.1057/s41599-025-04880-2>

OPEN

The construction of authorial voice in thesis writing: a multidimensional comparative perspective

Ruoxi Liu^{1✉}, Ping Xin¹ & Ken Chen^{2✉}

Constructing appropriate authorial voices through in-text citations is the basis for quality academic writing. Writers are required to distinguish a variety of other voices presented in academic writing, including their own, and to maintain effective control over them. Voice construction has been widely studied, but little attention is given to the examination of novice writers. To fill this gap, this study aims to conduct a multidimensional exploration of masters' citation usages and the construction of authorial voices in terms of citation forms, reporting markers and citation distributions. For this purpose, a corpus of 20 Chinese master theses in Applied Linguistics and a corpus of 20 published research papers have been constructed, all written in Chinese and totaling one million Chinese characters. Findings revealed that, compared with experts' use of citations, novice writers employed significantly more citations in Research Background while less in the Main Body section. In each section, novices used significantly more "author as subject" forms to list the cited sources without conveying any evaluation while using fewer non-integral forms, which led to the authorial voice being drowned out by the cited authors' voices. Both novices and experts often used research markers and expressed neutral evaluation, and the evaluation in Chinese academic writing was usually not communicated via reporting markers directly. This study provides a dynamic view of citation practices of novice writers and also sheds new insights on the study of academic writing in Chinese.

¹School of Chinese As a Second Language, Peking University, 100871 Beijing, P. R. China. ²Chinese International Education College/Overseas Education College, Xiamen University, Amoy City, Fujian Province 361102, P. R. China. ✉email: liuruoxi@stu.pku.edu.cn; chenken@xmu.edu.cn

Introduction

The academic paper is a kind of communicative discourse based on literature resources (Hyland 1999). It is not just monologues of the writer but also the product of dialogs and exchanges between different voices within the existing literature, which means that the writer needs to articulate their own voice in addition to displaying subject knowledge (Sun et al. 2022). Therefore, constructing the authorial voice in academic writing is an essential part of academic texts. Although the voice in writing has properties that are hidden and ineffable and whose definitions are debated, the term voice, as used in our study, refers to “the amalgamative effect of the use of discursive and non-discursive features that language users choose, deliberately or otherwise, from socially available yet ever-changing repertoires” (Matsuda 2001, p. 40). In other words, it is a negotiation of meaning between writers and the world around them, as well as an engagement of writers with others’ perspectives through their own stance or evaluation.

Citation is a necessary way for other-sourced materials to enter the discourse, and it is a clear sign of “dialogism and intertextuality” in academic writing (Swales 2014, p. 119). Citation involves voices of both writers and cited authors (Bakhtin 1978), and requires the writers to not only demonstrate their knowledge of the research field through the use of source texts but also present their own evaluation and stance towards the cited content to highlight the value and uniqueness of their own research. Thus, citation is an important discursive resource for building an authorial voice. Writers are always expected to express their own voice objectively and confidently and to use it to persuade readers to accept their viewpoints. Therefore, the construction of authorial voice always poses huge challenges especially to novice writers whose previous learning experiences rarely involve critical evaluation of others’ views (Peng 2019).

It has been well acknowledged that academic writers must strategically use citations to incorporate multiple perspectives and convey authorial opinions (Marti et al. 2019; Wette 2017). Although many studies have focused on English L1 novice writers’ citations (Liardet and Black 2019; Samraj 2013), the differences between Chinese and English academic writing make it difficult to provide effective guidance for Chinese. For example, in the structure of citation sentences, the conjunction ‘that’ is required between the projecting clause and the projected information in English articles, but it is not required in Chinese (Yu and Zhang 2021). A small number of papers on Chinese academic citations mainly describe reporting verbs/markers, citation forms and other aspects of textual features (Liu et al. 2021; Liardet and Black 2019; Liu and Wang 2019; Planks and Gebriel 2012; Yu and Zhang 2021). However, none of the previous studies has paid particular attention to the construction of authorial voice in novice Chinese academic writing.

Moreover, Chinese students’ learning experiences rarely involve challenging other viewpoints. They usually receive limited training in Chinese academic writing before writing their theses (Sun et al. 2022), and citation is considered more as an academic norm issue and has not been included in the research scope of academic writing during the training process (Liu 2025). Although relevant guidebooks for thesis writing have been published (Qian and Zhang 2021; Zhou 2021), they basically do not include citation issues. There is still a gap in our knowledge of how they use citations.

By comparing the citation practices of master’s students (novice writers) and experts (experienced writers), we can further understand the difficulties that novices encounter when citing. Compared to Ph.D. students, who already have some writing experience and have already begun to establish their own authoritative voice (Wette 2017), master’s students are generally

considered to be at the beginning of a researcher’s study and are still inexperienced writers, who are more characteristic and representative of novice writers’ citation use (Li and Zhang 2021). And as a research text, master’s theses have to fulfill the function of academic persuasion and make readers believe that the conclusions of current research are true and credible (Ahn and Oh 2024), so it is necessary for writers to effectively construct their authorial voice in them. In addition, master’s theses usually take about a year to be carefully written and are relatively long which require students to systematically cite relevant literature and construct a complete academic argument. There are also strict norms regarding citation formats, literature annotation and so on. Therefore, compared with academic writing under test and coursework conditions which are shorter in length and contain fewer citations, master’s theses can thus reflect novices’ academic citation skills in a comprehensive and in-depth manner.

Thus, to further contribute to the knowledge of novices’ authorial voice construction through citations, this study intends to compare the citation practices of Chinese applied linguistics-major master’s students and experienced writers, including citation distribution, citation forms, and high-frequency reporting markers. The research results will help to provide targeted guidance for academic citation mentoring, which could guide novices to make more effective contact with academic communities, strengthen their own viewpoints and to persuade readers to accept them.

Literature review

The construction of authorial voice in academic writing. Voice is responsible for creating an impression of writers in the minds of readers. A writer’s voice is multiple and intertextual, including echoes and associations with previous research, which is an important prerequisite for achieving academic persuasion (Bakhtin 1982; Tardy 2016). Voice has individual and social characteristics (Tardy 2016), and it emphasises the expression of individual and unique voice in writing (Atkinson 2001). On the other hand, academic writing always takes place in a shared ideological and social context. A process of meaning negotiation between the writer and the world around them is behind the different discursive features. Therefore, an individual’s voice is inextricably linked to the expectations of the academic community and is only persuasive and meaningful when it contributes to and connects with the community’s ideology and value system (Bakhtin 1982; Matsuda and Tardy 2007).

Voice therefore plays a critical role in the reception of academic writing and can often be viewed as the effect of a writer’s use of textual devices (Matsuda and Tardy 2007). The construction of authorial voice with the help of different text resources is a necessary process of academic discourse learning. However, novices are often unfamiliar with how to express their views and positions in academic writing and find it difficult to design appropriate texts based on readers’ rhetorical expectations and information needs (Hyland 2012; Magalhães et al. 2019). Several empirical studies have found that novices are intimidated when sharing their views through academic texts (Caffarella and Barnett 2000; Mansourizadeh and Ahmad 2011), or unable to convey their contribution to knowledge and infuse that voice with a sense of their own personal authority through literature resources (Sun et al. 2022). Novices were also found to “mimic or ‘ventriloquise’ the dominant discourses without identifying with them” in order to disguise their lack of their own voice (Barnett and Di Napoli 2008, p. 201). However, if writers are unaware of the impression conveyed by their different textual devices, they may be labeled as inexperienced and as outsiders by the academic community.

Authorial voice in citation

Citation distribution. Citation distribution refers to the number of citations in different parts of academic texts, reflecting the degree of connection between writers' knowledge creation and other views (Bakhtin 1982). Citations may be distributed differently due to their functions in each section of a paper, for example, the introduction and discussion sections may be dense with citations to construct subject knowledge and show the significance of the study (Mansourizadeh and Ahmad 2011). Petrić (2007) studied 16 English masters' theses written by non-native speakers and found that high-level novices tend to use more citations. Some studies also explored the distribution of citations in each part of the academic writing and found that novices used the most citations in the introduction to locate their own research in the community, followed by the results and discussion sections (Swales 2014).

However, the corpora used in these studies are all course assignments, and the assignment requirements may have an impact on the distribution of citations, resulting in different research findings. For example, after analyzing students' in-class writings, Wette's (2017) study reached different conclusions from those of researchers like Swales (2014). It was found that most of the citations appeared in the body paragraphs, while the introduction and conclusion had fewer in novices' homework theses. Therefore, a more comprehensive and in-depth exploration using corpora that can better represent their writing proficiency is still needed to figure out how novices distribute their own voices and those of other researchers.

Citation forms. In addition to distributing citations reasonably, manipulating grammatical and lexical devices such as citation forms and reporting markers has also been shown to contribute to the construction of an authorial voice (Hyland 2012; Morton and Storch 2019). The influential binary categorization of citation forms into integral and non-integral citations, proposed by Swales (1990), is broadly recognized in academic literature. Integral citation means the cited author's information such as name, year, page number, etc. is part of the grammatical structure of the sentence, while non-integral is when the cited author's information is located outside the sentence structure and placed independently at the end of the sentence within parentheses, or in footnotes or endnotes. The current study refers to Swales' (1990; 2014) classification to explore citation forms in our data and further descriptions are presented in the methods section with illustrative examples.

The framework above has shaped many later studies on citation practices, which have shown that the flexible mastery of citation forms is challenging for novices (Hyland 1999; Thompson and Tribble 2001; Wette 2017). For instance, Sun et al. (2022) conducted a one-year longitudinal observation of 10 master students and found that high-group students who got high average marks for written coursework used integral and non-integral citations evenly while low-group students used primarily non-integral citations. While other studies have found that novices tend to use integral citations to attribute a finding or idea to a specific source (Hirvela and Du 2013; Shi 2004), it is easy to disrupt the flow of discursive arguments and inappropriately emphasise the authority of the cited authors (Marti et al. 2019; Peng 2019; Swales 2014; Wette 2017, 2018). Despite these findings, existing studies have focused on analyzing citation practices in research articles as a whole, rather than examining individual sub-genres. To some extent, this approach limits the exploration of the unique rhetorical purposes inherent in each section and has not yet fully addressed the complex interactions among the citation practices of different sections (Ahn and Oh 2024).

Reporting markers. Reporting markers are used to introduce other people's research and opinions and reflect writers' evaluations of cited works (Hunston 1995; Hyland 2002; Hawes and Thomas 1997), which can directly express the author's voice. There is a great deal of research on reporting markers in English academic discourse. It has been found that the use of these markers by novices has the following characteristics. For example, a small proportion of reporting verbs were used excessively repeatedly (Hyland 1999, 2002; Marti et al. 2019); expert authors and masters could use more diverse reporting verbs and are more familiar with disciplinary lexis than undergraduate students (Friginal 2013); undergraduate students strongly rely on neutral stance reporting verbs and are unable to express their own position about the cited content (Liardet and Black 2019).

Unfortunately, all of the above studies are analyses of novice English academic writing. Factors such as the differences between the English and Chinese languages and the different backgrounds of novice writers make it difficult for these conclusions to provide direct guidance for teaching citation in Chinese. Research on reporting markers in Chinese academic discourse is severely lacking. After analysing 100 Chinese journal papers, Liu et al. (2021) proposed the term 'reporting markers' because they found that in addition to the use of reporting verbs alone, Chinese also uses prepositions, verb phrases and frame structures. This research also adopts the term 'reporting markers', which also includes 'reporting verbs' in other related studies. The existing research has provided a clear classification framework for this study. However, what it presents are the usage characteristics of experts in different disciplines, but we do not know the evaluative functions of reporting markers and how reporting markers are involved in the construction of authorial voice; nor do we know the usage characteristics of novices within applied linguistics.

To summarize, though numerous studies have been done regarding the use of citations in academic writing, many remain to be explored. First, there have been fewer studies on master's citations (Hendley 2012; Keck 2006; Liu et al. 2021), so it is necessary to understand how citations are used and further demonstrate how novices construct their authorial voice in the writing. Second, previously a large number of studies have investigated novices' citation practices under English test and coursework conditions (Sun et al. 2022; Wette 2018), while there has been a lack of studies examining the challenges that novices face in theses comprehensively and thoroughly. Third, whether students can use the citation format appropriately in each sub-genre still needs to be analysed and discussed in combination with specific contexts and detailed data (Petrić 2007; Tseng 2018). Finally, the voice construction ability in the citation is a multi-dimensional structure, so a multidimensional comparative perspective is needed to fully investigate a writer's citation competence (Li and Zhang 2021).

Thus, a gap in our knowledge remains in relation to how Chinese masters with limited academic training use citations to construct authorial voice for thesis writing. While such novices are L1 students, citation use in academic writing with disciplinary norms is still challenging for them (Beaufort 2004). Thus, to better understand how Chinese masters in applied linguistics appropriate their citation use to build authorial voice in each section of their Chinese theses and the challenges they experience in the process, this study raised the following research questions:

1. Are there any differences in how authorial voice is constructed in the distribution of citations between novices' theses and experts' papers in Chinese academic writing?

2. Are there any differences in how authorial voice is constructed in the citation forms between the aforementioned sub-corpora?
3. Are there any differences in how authorial voice is constructed regarding the use of high-frequency reporting markers in citations between the aforementioned sub-corpora?

Methodology

The Corpus development. This study investigates citation practices in applied linguistics, comparing Chinese L1 master students and academic experts. Our analysis compiled a corpus that includes two sub-corpora to represent students and experts writing: Chinese Novices Theses (NOV) and Experts Journal Papers (EXP). The varied landscape of disciplinary traditions, such as soft versus hard disciplines¹, leads to distinct modes of citation practices (Becher and Trowler 2001). Investigating the citation behaviors within a single discipline is conducive to a comprehensive understanding of the citation characteristics and the ways of constructing authorial voice under the specific disciplinary norms, avoiding the influence of differences in disciplinary conventions. We draw our data from the field of applied linguistics, which is known for its sophisticated rhetorical techniques and emphasis on citation. As a representative of soft disciplines (Moed 2005), applied linguistics is often included in studies of academic writing (Hyland and Jiang 2017), so studying it makes our data comparable to previous studies.

Brown's (1988) stratified sampling strategy was used to ensure maximum equivalence between the sub-corpora for such crucial parameters as subject matter, structure and reference format. Such methodological equivalence was necessary to provide a common platform for making meaningful comparisons and drawing reliable and valid conclusions about the differences/similarities between novices and experts (Hu and Wang 2014). Theses and papers were initially selected by reading the abstracts to control for the subject matter. Empirical studies were included. These studies include the sub-genres of "introduction, methods, results, discussion and conclusion" sections, and use APA referencing, and are related to Chinese language instruction and second language acquisition. While mainly qualitative studies such as case studies and conversation analyses were not chosen, to control for the effect of structural and topical differences.

Following the above selection principles, each thesis was randomly collected from CNKI (<https://www.cnki.net/>), the largest academic resource provider in China, submitted between 2018 and 2022, to avoid the impact of differences in academic training and document accessibility caused by a large time span (Hyland and Jiang 2017). Referring to the corpus retrieval method of Li and Zhang (2021), a simple search was conducted for the theses and dissertations in linguistics, followed by the standard search for the theses in applied linguistics. The key words that were used in the simple search were "Chinese language acquisition/ learning/ teaching", "Chinese second language acquisition/ learning/ teaching", "Chinese foreign language acquisition/ learning / teaching". These topics did not include all fields in applied linguistics but were related to its sub-field. After the simple search and standard search, the duplicate documents were eliminated. Then, the selection process of the 20 master's theses was completed. Each paper written by experts was chosen from five authoritative journals published between 2018 and 2022. To ensure the authority of the journals, we consulted the influential journal citation report in mainland China, the Catalog of CSSCI (Chinese Social

Sciences Citation Index, 2021 - 2022, cssrac.nju.edu.cn), and asked specialist informants to nominate top journals in the discipline of applied linguistics. The research depth and writing norms of the papers included in the CSSCI Journals have been strictly reviewed and approved, and can be used as a reference for novice writers to learn writing.

The front matter (i.e., titles, authors, and abstracts/summaries), figures, tables, captions, footnotes, and back matter (i.e., acknowledgments, endnotes, author notes, references, and appendices) from the sampled articles were removed (see Appendix for the topics of the corpus).

The corpus description. The NOV sub-corpus includes 20 Chinese L1 master's theses, with a total of 825,091 characters (including 802,548 Chinese characters and 22,471 foreign characters), from Teaching Chinese as a Second Language and Chinese linguistics, both of which are subfields of applied linguistics. The authors of the theses who are Chinese L1 masters are identified by the following criteria: (1) the author's name on the thesis cover is written in Chinese and does not include a foreign name and nationality; (2) in the acknowledgments, the authors mention their relatives (such as partners or parents), teachers or friends in mainland China².

The EXP sub - corpus is composed of 20 journal papers and contains a total of 209,972 characters (including 202,702 Chinese characters and 5270 foreign characters) and is the reference corpus of this study. Journal papers were selected from 5 authoritative applied linguistics publications: "Language Teaching and Linguistic Studies(《语言教学与研究》)", "Chinese Teaching in The World(《世界汉语教学》)", "Chinese Language Learning (《汉语学习》)", "Applied Linguistics(《语言文字应用》)", and "TCSOL Studies(《华文教学与研究》)". To ensure diversity of writing style, no more than two papers per author are included.

Although the two sub-corpora contain most of the same sub-genres, we have noted the diversity in the structural arrangement of theses and journal papers. For example, some theses and papers include a "literature review" after the "introduction" section, while others include the "literature review" in the "introduction" section; some theses and papers combine "results" and "discussion" into a whole section, which forms "results and discussion". Thus, to better align with our focus on citation usage within these two sub-corpora, this study divided the texts into three sections: (1) the "introduction" and "literature review" were analysed together and formed into the Research Background section. The previous researches have highlighted similarities in the major moves of those two sections (Bhatia 1993; Kwan 2005), such as review background research, position the writer's own study and point out the niche to justify the study. An integrative analysis of these parts is warranted by their shared goal of providing knowledge background for research (Ahn and Oh 2024). (2) Main Body (including "methods", "results" and "discussion"), the "methods" section explains the basis and acceptability of the research method through citation, the "results" presented the findings, and the "discussion" may employ citations to further interpret their findings and establish connections with previous researchers (Li and Zhang 2021). Thus, an integrative analysis of these sections is necessary because they all share the common goal of presenting current research in detail and presenting it to the reader as objective and credible, which constitutes the main body of research that we will analyse together. (3) Conclusion (including "conclusion/suggestions"), the *conclusion/suggestions* section briefly restates the research result or puts forward

suggestions on this basis, which is the end of the research. This study takes them as a part for comprehensive analysis.

It is worth noting that although theses and journal papers are two different genres of writing, there are still some similarities between them. Papers published in authoritative journals can represent the disciplinary requirements and reflect professional academic writing skills (Nam and Beckett 2011), which can be imitated and learned by novices. And as mentioned above, this study ensured consistency in subject matter, structure, reference format and expected audiences (both include professionals within the discipline) between the two sub-corpora. Moreover, the two types of papers require writers to present the research process and findings in the articles to achieve academic persuasion, and the communicative purposes served by each sub-genre of them have the similarities. More importantly, we also take into account the differences between the two sub-corpora when analysing the results. For example, theses require a longer literature review, which may result in more citations of this section; the descriptive nature of theses could entail more sources to support their arguments, while the innovative nature of journal papers makes them focus on synthesizing previous studies and relating them to their own results. As for the other differences in citations between the two, these will be further explores in our study.

The above ensure that the two types of articles are comparable. Previous studies (e.g., Ahn and Oh 2024; Samraj 2013) have also made such comparisons, providing insights into citation practices and the relationship between the two genres. Therefore, this study explored a comparative examination of master's theses and journal papers to uncover how each navigates.

Analytical framework for the present study

Citation forms. Currently there is no framework for citation forms that is suitable for analyzing Chinese academic text. So referring to Swales's (1990, 2014) framework, this study conducted text analysis on the corpus of expert papers to clarify the citation forms of Chinese academic text and used them as the standard for corpus analysis. The coding result showed that Swales's (1990, 2014) framework could explain this study's data well, while it still needs some adjustments. For instance, example (6) belongs to "the subject as agent" of Swales (2014), but appeared only 3 times in our corpus, which was not enough to form a subcategory, so we classified it with "others" which includes the category that has been used no more than 5 times. The adapted framework of citation forms is shown in Table 1.

Reporting markers. Previous studies have found that the structures of Chinese reporting markers are diverse, including four forms: verb (e.g., 分析analyze), preposition (e.g., 根据according to), verb phrase (e.g., 注意到note), and frame structure (e.g., 由.....提出proposed by) (Liu et al. 2021). Our study also adopts the above structures. Hyland (2002) and Thompson and Ye (1991) proposed the taxonomy of reporting verbs, which has been widely referenced and adopted by subsequent studies. On this basis, combining with the pragmatic function of reporting markers, Liu et al. (2021) further classified them in combination with Chinese corpora. This study's framework of reporting markers' denotation is adapted from Liu et al.'s (2021) taxonomy and is shown in Table 2, which classifies reporting verbs into three categories according to the activities they refer to. Moreover, the

| Table 1 The framework of citation forms. | | | |
|--|-----------------------------------|---|---|
| Forms | Subcategories | Explanation | Examples |
| Non-integral | Nonreporting | Non-integral citation sentences do not include reporting markers | (1) 句法启动效应广泛存在于句子加工中 (Bernolet et al. 2007 ; McDough and Mackey 2008 ; Branigan and Pickering 2017) 。 (EXP-17) Syntactic priming widely exists in sentence processing (Bernolet et al. 2007 ; McDough and Mackey 2008 ; Branigan and Pickering 2017) ³ 。 |
| | Reporting | Non-integral citation sentences include reporting markers | (2) 诸多学者注意到线上教学和线上、线下混合教学将趋于常态化 (李泉, 2020 ; 崔永华, 2020) 。 (EXP-18) Many scholars have noticed that online teaching and mixed online and offline teaching will become normalized (Li Quan, 2020; Cui Yonghua, 2020). |
| Integral | Author as subject of the sentence | The cited authors' location of integral citation sentences was as subject | (3) 陆俭明 (2008) 曾提出应该研究如何在汉语教学过程中帮助学生尽快克服对拼音的依赖, 顺利过渡到汉字教学。 (EXP-20) Lu Jianming (2008) once proposed that we should study how to help students overcome their dependence on pinyin as soon as possible in the process of Chinese teaching and transition to Chinese character teaching smoothly. |
| | Author in noun phrase | The cited authors' location of integral citation sentences was in noun phrase | (4) 石定果 (1998) 的调查显示, 不管是否属于汉字文化圈, 大部分学生都会对汉字感兴趣。 (EXP-9) Shi Dingguo's (1998) survey showed that most students are interested in Chinese characters regardless of whether they belong to the cultural circle of Chinese characters. |
| | Author as adjunct | The cited authors' location of integral citation sentences was as adjunct | (5) 根据董秀芳 (2009), “整体—部分”关系的编码在词汇和句法系统中均有凸显性。 (EXP-10) According to Dong Xiufang (2009), the encoding of the “whole-part” relationship is salient in both lexical and syntactic systems. |
| | Other | A citation form that was used less than 5 times in the corpus, which is not enough to form a small classification | (6) 正如Tannen (1989、2007) 所指出, 重复模仿在语言的产出和理解方面优化了资源的使用。 (EXP-14) As Tannen (1989, 2007) pointed out, repeated imitation optimizes the use of resources in terms of language production and comprehension. |

Table 2 Categories of reporting markers denotations (adapted from Liu et al. 2021, p.67).

| Categories | Subcategories | Explanation | Examples |
|-------------------|---------------|--|---|
| Research markers | Procedures | Related to research activities occurring in the real world | 研究study, 分析analyze, 采用adopt, 定义define, 出版publish, 提供provide, 界定define, 阐述elaborate, 考察investigate |
| | Findings | | 发现find, 表明indicate, 证明prove, 揭示reveal, 证实confirm, 得出draw, 显示show, 发表present, 推动promote |
| Cognitive markers | | Associated with the researcher's mental and cognitive process | 在.....看来in one's opinion, 考虑consider, 承认acknowledge, 接受receive, 当作regard as, 相信believe, 关注focus on, 注意到notice, 看作see as |
| Discourse markers | | Associated with linguistic activities and focus on the verbal expression of cognitive or research activities | 认为think, 提出propose, 指出point out, 说say, 强调emphasize, 所说what sb. said, 探讨explore, 论述discuss, 介绍introduce |

Table 3 Categories of reporting markers evaluations (adapted from Hyland, 2002, p. 118).

| Categories | Explanation | Examples |
|------------|---|---|
| positive | indicating that the writers think the reported message is true and reliable. | 证实confirm, 强调emphasize, 完善improve, 倡导advocate |
| negative | indicating that writers have a negative attitude towards the quoted point of view, skeptical or negative. | 忽视neglect, 批判criticize, 怀疑doubt, 质疑question |
| neutral | the writers are characterized as staying neutral toward the proposition. | 考察investigate, 分析analyse, 讨论discuss, 采用adopt |

reporting markers also convey writers' attitudes and evaluations of the reported message, so the evaluation can be divided into three categories (Thompson and Ye 1991) as shown in Table 3.

Procedures

The 20 masters' theses and 20 experts' papers were coded as NOV-1 to 20, EXP-1 to 20, for the purpose of repeated reading, coding, and in - depth analysis. Instances of citation were identified following these guidelines (Peng 2019; Thompson and Ye 1991): (a) treat one parenthesis as a citation, in the form of "author (year)" or "(author, year)", multiple source texts within one bracket, such as example (2), were counted as one citation; (b) if a quoted sentence contains several source texts but is used in different propositions, it is counted as multiple citations; and (c) the original document and the quoted document of "reference" are regarded as one citation (Hu and Wang 2014). However, the following forms are not counted as citations: (a) internal citations referring to the same text; (b) self-citations by researchers or research participants; and (c) common research tools (such as SPSS) and statistical methods (such as Pearson's *r*).

However, categories for reporting markers are not easy to decide and may be inaccurate because they may change with context and overlap (Peng 2019). In order to improve the validity of the analysis, we invited two native Chinese teachers (they are all teachers at the same school as the authors, but were not among any of the authors) who taught academic writing to check the categories of markers generated in the analysis. They were provided with Hyland's (2002) definitions of categories and the list of 146 coded reporting markers of our corpus and independently reviewed the categories. Finally, they agreed on the classification of all reporting markers.

Corpus annotation was done by two doctoral students (one of them is the first author of this paper) majoring in Applied Linguistics. First, according to the analysis framework above, each student read the theses and papers carefully and completed the qualitative annotation of citation distribution, citation forms, reporting markers, and related information independently. Then they discussed the differences together and came to agreement, the Pearson correlation coefficient of the two annotators was 0.967, $p < 0.05$, and the consistency is good. Second, the researchers counted the original frequency of citation and citation

forms, and then used the *Chi-square test* to compare whether the frequencies were significantly different. It should be pointed out that the length of masters' theses and experts' papers is different, so the percentage was also presented to minimize any effect caused by differing sub-corpora sizes and to show differences in the data more clearly between the two sub-corpora. The percentage was calculated in these ways, the citation's occurrence in each section was added up and divided by the total instances of citations in all text to show the distribution of citations. In each section, to present the percentage of each form, the occurrence of each citation form was added up, and divided by the total instances of citations. Finally, this study analyzed the frequency of all reporting markers in two sub-corpora, calculated the top 10 reporting markers that were used most frequently, and analyzed their evaluations in combination with the specific context. Then we calculated the percentage by adding up the occurrence of each marker and dividing it by the total instances of all markers.

Findings and discussion

The construction of authorial voice through citation distribution. The citation distribution in different sections of two sub-corpora is shown in Table 4. A total of 1073 citations were found in the theses, with an average of 53.65 per thesis, while 620 citations were found in experts' papers, with an average of 31 per paper. In both sub-corpora, citations were mostly used in the Research Background section followed by the Main Body, but were rarely used in the Conclusion section. The *Chi-square test* shows that the frequency of citations in the Research Background of novices' theses is significantly higher than that of experts ($\chi^2 = 31.71$, $p < 0.001$), and that of Main Body is significantly lower than that of experts ($\chi^2 = 55.77$, $p < 0.001$). The Conclusion section usually does not need to cite many previous opinions. The frequency of citations in both sub - corpora is low, and there are no significant differences.

Specific analysis reveals that the number of citations in the Research Background and Main Body sections is distributed (about 50% each) evenly in experts' papers, which is consistent with Thompson's (2005) research findings that citations play an important function in the two sections of the paper. However, novices' references focus excessively on the Research Background section. This may be related to the need for more citations in the

long literature review of a master's thesis, and it also indicates that novices are striving to establish a rich background for their research, which is a way to show their familiarity with the research field.

In contrast, the proportion of citations used by novices in the Main Body is significantly lower than that of experts. Qualitative analysis reveals that they mainly encounter the following problems. Firstly, novices often lack cross-reference citations, in other words, the important documents and related studies that appear in the Research Background section are not mentioned in this section, even some novices (15%) do not cite any source texts in the Main Body section. Secondly, some novices (35%) do not use citations when presenting the research method, which indicates that they fail to highlight the rigor and validity of the methodology and procedures in their research through citations. Thirdly, some of the novices (40%) do not cite any views when discussing the research results, which confirms Mansourizadeh and Ahmad's (2011) idea. They lack the ability to place the current research in a broad research field for comparison and explanation. Finally, there are some novices (40%) who cited a large number of predecessors' conclusions to express their own views. This may indicate that they lack the consciousness and academic confidence to independently present their own research conclusions and highlight their own voices.

The construction of authorial voice through citation forms use

Quantitative analysis of citation forms. Table 5 presents statistics of integral and non-integral citations in each section of two

sub-corpora. The results show that novices use integral citations significantly more often than experts in Research Background, Main Body, and Conclusion section ($\chi^2 = 151.78$, $p < 0.001$; $\chi^2 = 39.96$, $p < 0.001$; $\chi^2 = 17.99$, $p < 0.001$), while the frequency of non-integral citations is significantly lower than that of experts in Research Background, Main Body, and Conclusion section ($\chi^2 = 55.44$, $p < 0.001$; $\chi^2 = 17.49$, $p < 0.001$; $\chi^2 = 8.99$, $p < 0.05$).

This study further analyzed the frequency of different sub-categories of integral and non-integral citations (see Tables 6, 7). The results show that the "author as subject" of novices' theses is significantly higher than that of experts in Research Background, Main Body, and Conclusion section ($\chi^2 = 62.53$, $p < 0.001$; $\chi^2 = 53.66$, $p < 0.05$; $\chi^2 = 6.74$, $p < 0.001$). While the frequency of "nonreporting" citation of theses is significantly lower than that of papers in three sections ($\chi^2 = 82.24$, $p < 0.001$; $\chi^2 = 5.91$, $p < 0.05$; $\chi^2 = 10.07$, $p < 0.001$), the "reporting" citation of theses is also significantly lower than that of papers in Research Background and Main Body section ($\chi^2 = 78.12$, $p < 0.001$; $\chi^2 = 5.91$, $p < 0.001$).

Our results show that novices prefer to use "author as subject" in three sections, while experts can use integral and non-integral citations in a more balanced way. In fact, "author as subject" puts cited authors in an important position, highlighting the voice of cited authors and allowing for "alternative positions and voices" (Martin and White 2005, p. 102). However, it also tends to weaken the expressiveness of writers' own positions (Weissberg and Buker 1990). While non-integral citations are usually the form more frequently used by expert authors in applied linguistics, they can emphasize the objectivity and credibility of scientific research and ensure the integrity and coherence of textual narratives (Hewings et al. 2010; Hyland 1999), thereby developing their authoritative voice. Previous studies (Peng 2019; Sun et al. 2022) found that novices prefer integral citations, while experienced writers use non-integral citations more to establish academic identity, which is consistent with this study's findings.

The quantitative examination depicted the distribution patterns of citation forms in each section of the sub-corpora. This result has triggered discussions about the potential motives for the selection of these forms, their contextual utilization, and rhetorical implications, leading us to shift our focus toward a comprehensive qualitative analysis.

Table 4 Citation distribution in each section (frequency and percentage).

| Sections | Novices | Experts | Chi-Square | p |
|---------------------|------------|------------|------------|-------|
| Research Background | 776 72.32% | 308 49.68% | 31.71 | 0.00* |
| Main Body | 266 24.79% | 286 46.13% | 55.77 | 0.00* |
| Conclusion | 31 2.89% | 26 4.19% | 2.02 | 0.16 |

* $P < 0.5$.

Table 5 Citation forms in each section of two sub-corpora (frequency and percentage).

| Forms | Sections | Novices | Experts | Chi-Square | P |
|--------------|---------------------|------------|------------|------------|-------|
| Non-integral | Research Background | 99 12.76% | 168 54.55% | 151.78 | 0.00* |
| | Main Body | 79 29.70% | 161 56.29% | 39.96 | 0.00* |
| | Conclusion | 7 22.58% | 21 80.77% | 17.99 | 0.00* |
| Integral | Research Background | 677 87.24% | 140 45.45% | 55.44 | 0.00* |
| | Main Body | 187 70.30% | 125 43.71% | 17.49 | 0.00* |
| | Conclusion | 24 77.42% | 5 19.23% | 8.99 | 0.01* |

* $P < 0.5$.

Table 6 The Subcategories of non-integral in each section (frequency percentage).

| Subcategories | Sections | Novices | Experts | Chi-Square | P |
|---------------|---------------------|-----------|------------|------------|-------|
| Nonreporting | Research Background | 80 10.31% | 112 36.36% | 82.24 | 0.00* |
| | Main Body | 68 25.56% | 107 37.41% | 5.91 | 0.02* |
| | Conclusion | 6 19.35% | 20 76.92% | 10.07 | 0.00* |
| Reporting | Research Background | 19 2.45% | 56 18.18% | 78.12 | 0.00* |
| | Main Body | 11 4.14% | 54 18.88% | 5.91 | 0.02* |
| | Conclusion | 1 3.23% | 1 3.85% | 10.07 | 0.73 |

* $P < 0.5$.

Table 7 The Subcategories of integral in each section (frequency percentage).

| Subcategories | Sections | Novices | Experts | Chi-Square | P |
|---------------|---------------------|------------|------------|------------|-------|
| Subject | Research Background | 628 80.93% | 109 35.39% | 62.53 | 0.00* |
| | Main Body | 135 50.75% | 45 15.73% | 53.66 | 0.00* |
| | Conclusion | 18 58.06% | 4 15.38% | 6.74 | 0.00* |
| NP | Research Background | 31 3.99% | 19 6.17% | 2.24 | 0.17 |
| | Main Body | 17 6.39% | 24 8.39% | 0.67 | 0.46 |
| | Conclusion | 0 0% | 1 3.85% | 1.17 | 0.93 |
| Adjunct | Research Background | 16 2.06% | 10 3.25% | 1.43 | 0.35 |
| | Main Body | 33 12.41% | 50 17.48% | 2.23 | 0.12 |
| | Conclusion | 6 19.35% | 0 0% | 5.04 | 0.06 |
| Others | Research Background | 2 0.26% | 2 0.65% | 0.92 | 0.69 |
| | Main Body | 2 0.75% | 6 2.10% | 1.67 | 0.33 |
| | Conclusion | 0 0% | 0 0% | 0.00 | 1.00 |

*P < 0.5.

Qualitative analysis of citation forms

Citation forms in the research background section. The Research Background section includes presenting the research background, building a network of previous studies to position the writer's own new research and pointing out the research gap to justify the study (Kwan et al. 2012), which is notably prominent in applied linguistics and are closely intertwined with citation practices (Swales 2004). The qualitative analysis reflects the citation forms used by experts.

1) Backgrounding and reviewing previous research This part involves providing the macro research background and constructing the research network. Experts often summarize multiple related documents by mainly using non-integral citations (46.67%) to establish the research net, such as example (1) and (2), which were mentioned above. Among them, "reporting" usually refers to researchers as the subject, such as "many scholars" and so on. Most of the documents cited at this time have commonalities and are widely accepted views, such as example (1).

Compared with experts who often use non-integral citations to synthesize previous studies (Samraj 2008), novices rarely synthesize source texts to provide the research background. The purpose of novices' citations seems only to show sources of cited literature. They often list a single cited document by using a larger number of "author as subject" citations (even half of them use all "author as subject" in the whole Research Background), resulting in a simple presentation of information and failing to effectively weave previous literature into a research background network. Example (8) is a complete section from novices containing four paragraphs in total, each of the paragraphs lists a single source by using "author as subject". The result indicates that they may lack a broad understanding of the related literature and the ability to generalize. This result is consistent with the conclusions of previous studies (Sun et al. 2022; Wette 2017, 2018).

(7) 目前, 考察注音文本阅读眼动模式的研究只有Yan et al. (2008) 这一项。该研究考察了.....但有关拼音作用的研究结果没有得到一致的解释。(EXP-20)

At present, Yan et al. (2008) is the only study on the eye movement pattern of phonetic text reading. The study looked at..... But there is no consistent explanation for the role of pinyin.

(8) 成海萍提出.....石芬提出.....贺海分析.....蒋朝莉认为..... (NOV-8)

Cheng Haiping proposed..... Shi Fen put forward..... He Hai analysis..... Jiang Chaoli believes that.....

2) Demonstrating the research significance and indicating a gap Experts in this part lay a well-structured foundation for

research significance and carve out a niche, thereby increasing the acceptance of the study within the discourse community (Swales 1990). Experts typically use the 'author as subject' citations (35.39%) to cite a specific document or use it in the first paragraph to lay the foundation for the research significance of the full text. In this case, the "author" is usually an authority or a well-known scholar. Example (3) indicates that the discourse information comes from an academic authority. This helps the writer establish contact with other academic communities and establish the writer's research status in the academic field. It may also discuss research highly relevant to the problem to be explored in detail. Then, the writer will immediately point out the shortcomings of existing research and establish a research gap among these points of view. Example (7) expresses a negative evaluation of the existing Chinese research by using "but", and the authorial voice is also constructed immediately.

Accompanying the extensive use of integral citations, novices usually lack evaluative expressions for the cited content; Also, they rarely make connections between source texts. The reason is that they are unable to indicate the research gap directly through explicit lexical and grammatical markers (Neville 2007). Eventually, this leads to "shopping list" style descriptive citation and weakens the writers' voice. They ultimately lead to "shopping list" style descriptive citation and weaken the writers' voice. They may not have realized that citations should be used as a means of showing thought development, presenting a foothold of their own research and giving one's own voice. In addition, the novices in this study were all Chinese students and may exhibit the cultural tendency to avoid conflict and judgment towards sources that are seen as authoritative. Therefore, listing previous research without any evaluation is a relatively safe practice for them.

Citation forms in the Main Body section. Citation in the Main Body section needs to explain the research method and results, and compare findings with previous scholars' views (Kwan and Chan 2014). These parts need to use citations to connect with previous literature and expand the scope of discussion (Samraj 2013).

1) Explaining the research method The dominant use of citations in this section is to show support for the methodology and procedures. Some novices do not cite any sources of research methods as mentioned above. The rest mainly use "non-reporting" to prove the acceptability and credibility of the methods used or "author as adjunct" citation to explain the basis of the research method adopted, as shown in example (9). These citations are consistent with those of experts. This trend highlights the intention of novices and experts to present

propositions as facts and form an objective writing tone in order to confirm the validity of their method choices (Hyland 1999, 2002).

(9) 根据Ure (1971) 的词汇密度计算公式, 两个语段的词汇密度均为0.79。(NOV-5)

According to Ure's (1971) lexical density calculation formula, the lexical density of both segments is 0.79.

2) Explaining the results In this part, the authors introduce the principal findings of their research and propose potential explanations for these results, elucidating the reasons underlying the observed phenomena and the implications of these findings (Ahn and Oh 2024). Experts often utilize “nonreporting” citations to synthesize relevant ideas from multiple source texts. In example (10), “Nonreporting” implies that writers hold a supportive attitude towards the cited content. They present it as an established fact and form a community with other researchers, allowing the authority of previous research to speak on behalf of the writers and reinforce the writers’ research justification (Coffin 2009).

However, most of the novices still rely excessively on “author as subject”. In fact, it is not necessary for novices to highlight too much research in such a conspicuous way. Because when experts are explaining results, “author as subject” is usually used to cite content that is highly relevant to current research or from authoritative cited authors. Example (11) emphasizes the concepts that are not highly relevant and inappropriately highlights the voice of the cited author. This reflects that novices may have uncertainties in clarifying their academic stances due to limited knowledge or difficulties in using various citation forms. They seem to cite for the sake of citation and without realizing the purpose of citing sources when interpreting results.

(10) 中级水平新手的表现和界面假说预测相反, 回答这个问题, 我们需要回到直接回指语境中的主宾语不对称现象(Jens, 2018; 马千, 2018)。(EXP-5)

The performance of novices at the intermediate level is contrary to what the interface hypothesis predicts. To answer this question, we need to return to the subject-object asymmetry in the direct anaphoric context (Jens, 2018; Ma Qian, 2018).

(11) 汉字学得, 李蕊 (2014) 定义为“通过学习而获得汉字能力的过程”。我们分析后将他们的偏误可分为别字、错字和笔顺三种。(NOV-8)

Chinese character acquisition is defined by Li Rui (2014) as “the process of acquiring Chinese character ability through learning”. After analyzing their errors, we can divide their bias into three kinds: misspelling, misspelling and brushwork.

3) Comparing with results in literature This part emphasizes the comparison between the research and the cited literature. Therefore, it is particularly important to establish links among sources and make comparisons with the results in the literature. Experts usually use non-integral and integral citations flexibly according to the importance of cited documents, with the ultimate goal of locating and building research conclusions into the development system of the entire discipline. As in example (12), experts summarize multiple sources by non-integral citations for comparison, clarify whether they are “friends” or “foes” (Samraj 2013) and further prove the contribution of current research. Or as in example (13), a small part of integral citations was used to compare the research results with the important sources which have high correlation with the current study and were mentioned in the Research Background.

In comparison, novices use a large number of “author in NP” and “author as subject” citations to list documents, such as in example (14), three cited authors constitute a description of the “who did what” list. This approach can, to a certain extent, form a contrast with previous research. However, it is not necessary to elaborate on

previous research that is the same as the current results. When comparing results, it is more important to highlight the innovation and credibility of the current results. Successful argumentative texts should always place the writers in a dominant position (Groom 2000; Sun et al. 2022). Therefore, the absence of a non-integral structure may divert the readers’ attention from the main topic and impede the logical connections between ideas (Kaplan 1996).

It is worth noting that the above results may also be related to the “genre-specific features” (Kawase 2015) between the two sub-corpora. Theses aim to demonstrate students’ understanding of their subject area and their eligibility for a degree, which may require a thorough analysis and the use of integral citations of sources to support their arguments. In contrast, journal papers focus on disseminating research findings and demonstrating the novelty in the field. These behaviors enable us to gain an in-depth understanding of the more descriptive nature of master’s theses and the more proficient way in which research articles can synthesize previous works and relate them to their own works (Ahn and Oh 2024).

(12) 我们发现工具型动机并非祖语保持的影响因素, 与对美国华裔学生和日本华裔成人的研究结论 (张莉, 2015; 温晓虹, 2012; 邵明明, 2018) 不同。(EXP-17)

We found that instrumental motivation is not an influencing factor for ancestral language retention, which is different from the research conclusions of Chinese American students and Japanese Chinese adults (Zhang Li, 2015; Wen Xiaohong, 2012; Shao Mingming, 2018).

(13) 因此, 对于整体加工减弱的现象, 我们的解释与Hsiao等 (2012) 和Tso 等 (2014) 相同。(EXP-8)

Therefore, for the phenomenon of overall processing weakening, our explanation is the same as that of Hsiao et al. (2012) and Tso et al. (2014).

(14) 我们认为左右结构的汉字较难习得, 与其他学者的结论一致, 尤浩杰 (2003) 发现……黄伟 (2012) 发现……王骏 (2015) 也发现……(NOV-15)

We believe that Chinese characters with a left-right structure are more difficult to acquire, which is consistent with the conclusions of other scholars. You Haojie (2003) found...Huang Wei (2012) found...Wang Jun (2015) also found that...

Citation forms in the Conclusion section. The Conclusion section is intended to further summarize the research findings or provide suggestions and inspirations. Experts use a lot of “nonreporting” (76.92%) citations which have the attribute of acknowledging previous opinions (Swales 2014), as in example (15), the writers use “nonreporting” citations cleverly to integrate external sources with their own voice, so the cited author’s voice is minimized in the argument dominated by the writers’ voice. “Nonreporting” citations could also reduce the heteroglossia and increase the monophony of academic opinions to support writers’ own conclusions or suggestions here.

(15) 关键事件是引导教师反思的重要方法之一 (程乐乐, 2017), 可通过关键事件引导教师对教学进行反思。(EXP-16)

Key events are one of the most important ways (Cheng Lele, 2017), which could guide teachers to reflect on their teaching.

However, novices also cite sources that they supported, but they still mainly use “author as subject” citation (58.06%) which leads to weakening their own voices as in example (16). They also used “author as adjunct” (19.35%) to cite the foundation for the suggestions they proposed. This usage does not appear in the experts’ papers (Example (17)). However, in the Conclusion section, cited authors usually do not need to be highlighted and should serve as ancillary components to strengthen the writers’ voice (Ahn and Oh 2024), making it inappropriate to use too many “author as subject/adjunct”.

(16) 王景丹 (2008) 提出学习成语要理解它的含义。在课堂上, 教师要看重成语意义教学, 掌握意义是学会成语的证明。(NOV-4)

Wang Jingdan (2008) suggests that learning an idiom requires understanding its meaning. In the classroom, teachers should look at teaching the meaning of idioms and mastering the meaning is the proof of learning idioms.

(17) 根据 Giles和Byrne的“族群间模式”, 提高学生文化认同的方法是在目的语国家学习。(NOV-1)

According to Giles and Byrne’s Interethnic Model, the way to improve students’ cultural identity is to study in the target language.

The construction of authorial voice through reporting markers use

Novices have 1272 tokens and 146 types of reporting markers, while experts have 499 tokens and 130 types. In this study, the method of Liu et al. (2021) was adopted⁴: $V = TTR \times (1 - CHFI)$ to calculate the diversity of reporting markers. It was found that the diversity of novices’ reporting markers is 0.04, which is lower than that of experts, which is 0.13. The result indicates that novices have a stronger tendency to use a small number of markers than experts. Worse still, some novices use the same marker almost throughout an entire discourse. For example, in one novice’s paper with 66 citations, “cited author + 认为 think...” is used 55 times. It suggests that some novices use reporting markers in a rather monotonous way and lack awareness of the rhetorical function of reporting markers, as well as having problems with the lack of rhetorical devices and stylistic awareness.

The top 10 high-frequency reporting markers for novices and experts are shown in Table 8, accounting for 61.48% and 49.10% of the total, respectively. In terms of structural form, no frame structures were found; and the rest are verb forms except for the preposition “根据 according to” and the phrase “分为 divide into”. Novices and experts use verbs as the most common reporting markers. The 6 markers such as “认为 think, 提出 suggest, 发现 find, 研究 study, 指出 point out, 探讨 discuss, 根据 according to” appear in both sub-corpora, suggesting that novices and experts have the same tendency to select high-frequency reporting markers, similar to the conclusions of Hyland (2002) and Marti et al. (2019), which also show that novices adapt to academic discourse norms when using such markers.

From the perspective of markers denotation, novices and experts frequently used research markers (27.13% vs. 28.26%), followed by discourse markers (17.77% vs.12.23%) and cognitive markers (16.59% vs. 8.62%). Among the research markers, novices prefer procedure markers (20.13% vs. 7.00% for finding

markers), experts prefer finding markers (12.23% vs. 16.03%), showing that writers are more inclined to refer to real evidence to prove their conclusions. Novices emphasize predecessors’ research processes, while experts emphasize results. The above results verify the tendency of the increasing use of research markers in linguistics academic writing (Hyland and Jiang 2017), Chinese novice writers are aware of the important role of research markers in constructing the objective authenticity of academic texts and enhancing the persuasiveness of the authorial voice.

From the perspective of reporting markers’ evaluation function, the results show that reporting markers frequently used by novices are all neutral evaluations (61.48%), while experts use “表明 indicate, 指出 point out” as positive ones (7.82%) and the rest 8 (41.28%) are neutral. No negative markers were used in either sub-corpus. This result suggests that in Chinese academic writing, writers tend to adopt a neutral or accepting attitude towards literature and try to avoid direct criticism. Such an attitude may seriously threaten the face of the cited person and is not conducive to writers’ construction of their identity as members of the academic community.

The extensive use of neutral reporting markers by both sub-corpora may be related to the following two factors. Firstly, it may be influenced by international academic conventions. The trend in empirical discourse in applied linguistics also emphasizes the objective presentation of previous research processes and findings (Hu and Wang 2014). For example, similar patterns have also been found in studies on English reporting verbs by international applied linguistics scholars (Liu et al. 2021; Peng 2019). Neutral reporting markers help writers to create an objective academic discussion environment, making it easier for readers to accept and recognise their viewpoints. Secondly, Chinese writers may exhibit similar cultural tendencies, that is, they use expressions that tend to adopt a neutral or accepting attitude towards literature in academic writing, avoiding overly strong emotional tendencies and subjective evaluations (Hu and Wang 2014), which may seriously threaten the face of the cited person. This conclusion is consistent with the analysis results of Chinese students and experts by Cumming et al. (2018) and Sun et al. (2022). They also found that Chinese writers try to avoid direct criticism and conflicts because they regard the cited content as authoritative.

However, objectivity in presenting the authorial voice does not mean that only neutral evaluations are used. For example, this study found that most negative evaluations of cited information need to be conveyed through contextual expressions, which are re-evaluated by being placed in the text, as in example (18). Appropriate expression of negative evaluations is helpful in promoting academic innovation (e.g. pointing out the limitations of existing research and creating conditions for the birth of new theories and methods) and facilitating academic debate. The novices’ high-frequency reporting markers all adopt a neutral stance, showing a cautious attitude towards the source text and rarely expressing agreement or disagreement. This reflects their lack of knowledge about conventions and norms of academic writing.

(18) Tso 等 (2012) 都采用完全复合范式, 也考察了多类型读者群, 但仍然留下了很多疑点。

Tso et al. (2012) used a complete composite paradigm and also investigated multiple types of reader groups, but still left many doubts.

In addition, this study also found that novices have some difficulties in using reporting markers, such as the oral tendency markers (82 times in total) “说 said, 做了 done” which suggest that novices are less aware of academic genres. Some novices used the form of “先生 (a respectful form of address for scholars in Chinese)/老师 Teacher” (28 times in total, 2.20%) to show respect for

| Table 8 Top ten high-frequency reporting markers (frequency, percentage). | | | |
|---|------------|-----------------|------------|
| Novices | | Experts | |
| 认为 think | 211 16.59% | 发现 find | 60 12.02% |
| 提出 propose | 132 10.38% | 认为 think | 43 8.62% |
| 分析 analyze | 96 7.55% | 提出 propose | 29 5.81% |
| 发现 find | 89 7.00% | 表明 indicate | 20 4.01% |
| 研究 study | 77 6.05% | 指出 point out | 19 3.81% |
| 指出 point out | 68 5.35% | 研究 study | 18 3.61% |
| 调查 investigate | 32 2.52% | 考察 inspect | 18 3.61% |
| 分为 divided into | 28 2.20% | 探讨 discuss | 13 2.61% |
| 探讨 discuss | 26 2.04% | 采用 adopt | 13 2.61% |
| 根据 according to | 23 1.81% | 根据 according to | 12 2.40% |
| Total | 782 61.48% | Total | 245 49.10% |

previous researchers, but such behavior is definitely not conducive to the expression of the author's confidence and authority.

Overall discussion and conclusion

The comparative analysis of citation practices in experts' journal papers and masters' theses reveals notable variations in citation practices and the authorial voice construction within the field of Chinese applied linguistics. Our study employed a comprehensive approach, incorporating both quantitative and qualitative analyses within each section of the sub-corpora, along with contextual cues of various rhetorical purposes.

The current study has improved our understanding of citation practices in four main aspects. Firstly, it has comprehensively analyzed citation practices in constructing the authorial voice from multiple aspects, including citation distribution, forms, and reporting markers. This analysis could help researchers gain an in-depth understanding of novices' citation ability and the development path and characteristics of their voice construction (Petrić and Harwood 2013). Secondly, it has compared Chinese masters' theses with experts' papers, providing a broader perspective on the citation practices of novice Chinese writers. This analysis enumerates the citation instances throughout the texts and emphasizes the preferences and differences in specific citation practices between the two sub-corpora within applied linguistics. Thirdly, this study attaches great importance to the relationship between citations and the rhetorical purposes of each sub-genre to carefully examine the purposeful use of citations. This study synthesizes the extensive quantitative corpus analysis with the qualitative studies (Ahn and Oh 2024) and bridges the gap between quantitative and qualitative methods to reveal the citation patterns related to the construction of authorial voice. This study demonstrates the statistical significance of the observed differences. It provides detailed examples of specific rhetorical goals identified in the corpus, thereby enriching the academic understanding of citation patterns in academic writing.

Specifically, the analysis of citation distribution in this study indicates that the citations of novice writers are mainly utilized to emphasize their comprehension of the field in the Research Background section. In contrast, the balanced distribution of citations among experts shows that it improves the credibility of the analysis and augments the persuasiveness and scientific nature of their research.

Regarding citation forms, this study demonstrates how different citation forms can be used to clarify rhetorical purposes. Experts use non-integral and integral citations flexibly to manage references strategically. While novices overuse "author as subject" to simply list previous studies in each section regardless of whether they need to highlight or silence some voices, which overemphasizes the authority of the cited authors and weakens the expression of writers' opinions. For example, in the Research Background, especially when indicating "Backgrounding and reviewing previous research," experts effectively synthesize and link previous studies through non-integral citations, providing the macro research background. In contrast, novices often use integral citations to enumerate previous studies to show the sources of cited literature. Similarly, in the "explaining the results" part of the Main Body section, experts tend to use non-integral citations to integrate relevant research results to support their explanations. On the other hand, novices overly rely on "author as subject" to cite studies or concepts that are not highly relevant to the current research.

In the analysis of reporting markers, we observed that both sub-corpora tend to use research markers and neutral evaluation. This finding indicates that novices are primed with citation awareness to maintain objectivity and neutrality in academic

discourse despite problems such as lack of variety and insufficient expression of evaluation.

Overall, successful writers can use citations to organise existing research, create a clear research network, balance different external voices, and construct an objective and humble authorial voice. By citing previous studies, writers can highlight the value and significance of current research, build an authoritative and credible authorial voice, and enhance academic persuasiveness.

It is worth noting that both journal papers and theses represent crucial research types in academic writing, yet they possess distinct characteristics, such as length, intended audience, and purpose. Master's theses aim to demonstrate students' understanding of their disciplinary fields and their eligibility for a degree. This may lead to citing merely for form's sake (Petrić and Harwood 2013) to meet the expectations of supervisors and reviewers. In contrast, journal papers focus on disseminating research findings and showcasing innovation, thus exhibiting a stronger authorial voice, as experts write with the purpose of communicating and engaging in dialog with readers. These characteristics enable us to gain a deeper understanding of the descriptive nature of master's theses and the more proficient ways in which experts can synthesize previous works, connect them with their own, and effectively construct an authorial voice. Our research indicates that differences in citation practices may reflect the "genre - specific features" (Kawase, 2015) between master's theses and journal papers. The unique tendencies and preferences of each corpus reveal its distinct writing styles, not only presenting the citation characteristics and challenges of novice writers but also reflecting the complex and varying degrees of intertextual connections in academic writing (Ahn and Oh 2024).

Given the research findings, we propose the following implications for teaching. Academic writing teachers should raise masters' awareness of the rhetorical functions of citations, and help them realize that citation is not only for explaining sources and enumerating knowledge but also a necessary behavior for voice negotiation and intertextual storytelling (Ivanič and Camps 2001). Teachers can use examples of citations collected from scholars' publications. They can clarify how experienced writers communicate with cited authors. This helps students understand the purpose of citation and corresponding textual features in each section of academic writing. What's more, teachers can provide the citation lexico-grammatical resources systematically, for example, citation forms and reporting markers, and then give students numerous opportunities to practice citation use in authentic text. Finally, teachers should develop novices' research abilities and critical thinking skills. Teachers should also help novices read widely in the research field and build a comprehensive knowledge system. In this way, novices can compare and evaluate previous studies to avoid the "shopping list" style of citation and develop their identity and authorial voice through appropriate text features, thus transitioning from knowledge-telling to knowledge-transforming (Ahn and Oh 2024).

This study has some limitations that suggest directions for future research. Firstly, our study was based on the novice and expert corpora of Chinese applied linguistics, which are a small-scale sample. The research results may not be fully generalized to other disciplines with different citation norms and practices. Future research could be extended to other disciplines or other research fields to enrich our understanding of citation behaviors and academic writing conventions. Secondly, behind the citation practices lie the writers' own understanding of the use of citations and the motivations for choosing the forms of cited texts. For example, in this study, the differences in many citation behaviors between novices and experts may be caused by their different perceptions of citations. Future research could

conduct interviews and think - aloud experiments with writers to explore why novices fail to master certain citation usages. This would provide in-depth motivations for the differences in the linguistic representation of citations. As a result, novice writers can make conscious choices regarding the representation of citations. Thirdly, we only explored the usage characteristics of citations from the students' texts. Future research could also conduct interviews with novices or experts to present the motivations and perceptions of citations more in-depth.

Data availability

We have already explained the data collection process in the methodology section, and also presented the summary statistics in the tables within the paper. The datasets generated during and/or analyzed during the current study will be used in our further research in the future, so it cannot be made public for the time being. However, the datasets are available from the corresponding author (liuruoxi@stu.pku.edu.cn) on reasonable request.

Received: 13 February 2024; Accepted: 8 April 2025;

Published online: 29 May 2025

Notes

- 1 Soft discipline: According to the disciplinary classification framework with high recognition in the academic writing community (Becher and Trowler 2001), various disciplines can be divided into soft disciplines (e.g., linguistics, sociology) and hard disciplines (e.g., biology, computer science). The goal of hard disciplines is to create public knowledge that is falsifiable and accumulatively growing, emphasizing the falsifiable nature of knowledge itself; while soft disciplines attach more importance to the interpretability of knowledge, believing that knowledge is constructed by obtaining the support of academic discourse communities, emphasizing the unanimous recognition of knowledge among members of academic discourse communities.
- 2 The chosen theses were cross-checked with combined information about the master's learning experience and personal bio, which were retrieved from the theses or online (e.g. websites of the master's current institutions in China). The obtained information confirmed that these writers were native Chinese L1 speakers.
- 3 The examples were translated by the first author and checked by the second and third author of this study.
- 4 The calculation method that we used is " $V = TTR \times (1 - CHFI)$ ". When measuring vocabulary variety, the type-token ratio (TTR) is often used, which is the ratio of the frequency of types and tokens of a language item within a certain range. However, this method does not adequately reflect the impact of individual high-frequency items on the overall diversity. The influence of the proportion of high-frequency items needs to be considered in the measurement, so the calculation method of Liu et al. (2021) is adopted. Among them, V represents variety, TTR represents the ratio of symbol-like symbols, and $CHFI$ represents the proportion of "High-frequency items". When the proportion of "High-frequency item" is larger, the weight of the TTR value is smaller, and vice versa. Therefore, this calculation method combines the global evaluation and the weight influence of high-frequency items, so it can more accurately reflect the variety.

References

- Ahn C-Y, Oh S-Y (2024) Citation practices in applied linguistics: A comparative study of Korean master's theses and research articles. *J Engl Acad Purp* 69:1–15. <https://doi.org/10.1016/j.jeap.2024.101369>
- Atkinson D (2001) Reflections and refractions on the JSLW special issue on voice. *J Second Lang Writ* 10(1-2):107–124. [https://doi.org/10.1016/S1060-3743\(01\)00035-2](https://doi.org/10.1016/S1060-3743(01)00035-2)
- Bakhtin MM (1978) Discourse typology in prose. In Matejka L, Pomorska K (Eds.), *Readings in Russian poetics: Formalist and structuralist views*, 176–196. Ann Arbor, MI: Michigan Slavic Publications
- Bakhtin MM (1982) *The Dialogic Imagination*. Emerson C, Holquist M (trans). Austin, TX: University of Texas Press
- Barnett R, Di Napoli R (2008) Identity and voice in higher education: Making connections. *Changing identities in higher education: Voicing perspectives*, 197–204
- Beaufort A (2004) Developmental gains of a history major: A case for building a theory of disciplinary writing expertise. *Res Teach Engl* 39:136–185. <https://www.jstor.org/stable/40171658>
- Becher T, Trowler P (2001) *Academic tribes and territories: Intellectual enquiry and the culture of the disciplines*. Open University Press
- Bhatia VK (1993) *Analysing genre: Language use in professional settings*. Longman
- Brown JD (1988) *Understanding research in second language learning: A teacher's guide to statistics and research design*. Cambridge University Press
- Caffarella RS, Barnett BG (2000) Teaching doctoral students to become scholarly writers: The importance of giving and receiving critiques. *Stud High Educ* 25(1):39–51. <https://doi.org/10.1080/030750700116000>
- Coffin CJ (2009) *Incorporating and evaluating voices in a film studies thesis*. *Writ Pedagog* 1(2):163–193
- Cumming A, Yang L, Qiu C, Zhang L, Ji X, Wang J, Lai C (2018) Students' practices and abilities for writing from sources in English at universities in China. *J Second Lang Writ* 39:1–15. <https://doi.org/10.1016/j.jslw.2017.11.001>
- Friginal E (2013) Developing research report writing skills using corpora. *Engl Specif Purp* 32(4):208–220. <https://doi.org/10.1016/j.esp.2013.06.001>
- Groom N (2000) Attribution and averral revisited: Three perspectives on manifest intertextuality in academic writing. *Patterns and perspectives: Insights into EAP writing practice*, 14–25
- Hawes T, Thomas S (1997) Tense choices in citations. *Res Teach Engl* 393–414
- Hendley M (2012) Citation behavior of undergraduate students: A study of history, political science, and sociology papers. *Behav Soc Sci Librarian* 31:96–111. <https://doi.org/10.1080/01639269.2012.679884>
- Hewings A, Lillis T, Vladimirou D (2010) Who's citing whose writings? A corpus based study of citations as interpersonal resource in English medium national and English medium international journals. *J Engl Acad Purp* 9(2):102–115. <https://doi.org/10.1016/j.jeap.2010.02.005>
- Hirvela A, Du Q (2013) Why am I paraphrasing?: Undergraduate ESL writers' engagement with source-based academic writing and reading. *J Engl Acad Purp* 12(2):87–98. <https://doi.org/10.1016/j.jeap.2012.11.005>
- Hu G, Wang G (2014) Disciplinary and ethnolinguistic influences on citation in research articles. *J Engl Acad Purp* 14:14–28. <https://doi.org/10.1016/j.jeap.2013.11.001>
- Hunston S (1995) A corpus study of some English verbs of attribution. *Funct Lang* 2(2):133–158
- Hyland K (1999) Academic attribution: Citation and the construction of disciplinary knowledge. *Appl Linguist* 20(3):341–367. <https://doi.org/10.1093/applin/20.3.341>
- Hyland K (2002) Activity and evaluation: reporting practices in academic writing. In: Flowerdew J (ed) *Academic discourse*. London, Longman, pp 115–130
- Hyland K (2012) Undergraduate understandings: Stance and voice in final year reports. In *Stance and voice in written academic genres*, 134–150. London: Palgrave Macmillan UK
- Hyland K, Jiang FK (2017) Points of reference: changing patterns of academic citation. *Appl Linguist* 40(1):1–23. <https://doi.org/10.1093/applin/axm012>
- Ivanić R, Camps D (2001) I am how I sound: Voice as self-representation in L2 writing [J]. *J Second Lang Writ* 10:3–33
- Kawase T (2015) Metadiscourse in the introductions of PhD theses and research articles. *J Engl Acad Purp* 20:114–124. <https://doi.org/10.1016/j.jeap.2015.08.006>
- Keck C (2006) The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *J Second Lang Writ* 15:261–278. <https://doi.org/10.1016/j.jslw.2006.09.006>
- Kwan BSC (2005) *A genre analysis of literature reviews in doctoral thesis*. Unpublished PhD thesis. University of Hong Kong
- Kwan BSC, Chan H (2014) An investigation of source use in the results and the closing sections of empirical articles in Information Systems. *J Engl Acad Purp* 14:29–47. <https://doi.org/10.1016/j.jeap.2013.11.004>
- Kwan BSC, Chan H, Lam C (2012) Evaluating prior scholarship in literature reviews of research articles: A comparative study of practices in two research paradigms. *Engl Specif Purp* 31(3):188–201. <https://doi.org/10.1016/j.jsp.2012.02.003>
- Li Q-C, Zhang X (2021) An analysis of citations in Chinese English-major master's theses and doctoral dissertations. *J Engl Acad Purp* 51:61–75. <https://doi.org/10.1016/j.jeap.2021.100982>
- Liardet CL, Black S (2019) "So and so" says, states and argues: A corpus-assisted engagement analysis of reporting verbs. *J Second Lang Writ* 44:37–50. <https://doi.org/10.1016/j.jslw.2019.02.001>
- Liu H (2025) An investigation of the citation form characteristics of the graduation thesis by Chinese undergraduates majoring in Chinese international education. *Int J Chin Lang Teach* 6:128–136. <https://doi.org/10.46451/ijclt.20250112>
- Liu R, Huang QQ, Wang S (2021) The form, function and distribution of Chinese academic discourse reporting marker [汉语学术语篇转述标记的形式、功能与分布]. *Contemp Rhetor [当代修辞学]* 6:60–72
- Liu R, Wang S (2019) How citation is signaled: A corpus-based study on reporting verbs in Chinese academic papers. *Chinese for Specific and Professional Purposes: Theory, Pedagogical Applications, and Practices*: 53–71
- Magalhães MB, Cotterall S, Mideros D (2019) Identity, voice and agency in two EAL doctoral writing contexts. *J Second Lang Writ* 43:4–14. <https://doi.org/10.1016/j.jslw.2018.05.001>

- Mansourizadeh K, Ahmad UK (2011) Citation practices among nonnative expert and novice scientific writers. *J Engl Acad Purp* 10(3):152–161. <https://doi.org/10.1016/j.jeap.2011.03.004>
- Martin JR, White PRR (2005) *The Language of Evaluation: Appraisal in English* [M]. Palgrave Macmillan, New York
- Marti L, Yilmaz S, Bayyurt Y (2019) Reporting research in applied linguistics: The role of nativeness and expertise. *J Engl Acad Purp* 40:98–114. <https://doi.org/10.1016/j.jeap.2019.05.005>
- Matsuda PK (2001) Voice in Japanese written discourse: Implications for second language writing. *J Second Lang Writ* 10(1-2):35–53. [https://doi.org/10.1016/S1060-3743\(00\)00036-9](https://doi.org/10.1016/S1060-3743(00)00036-9)
- Matsuda PK, Tardy CM (2007) Voice in academic writing: The rhetorical construction of author identity in blind manuscript review. *Engl Specif Purp* 26(2):235–249. <https://doi.org/10.1016/j.esp.2006.10.001>
- Moed HF (2005) *Citation analysis in research evaluation*. Dordrecht, the Netherlands: Springer
- Morton J, Storch N (2019) Developing an authorial voice in PhD multilingual student writing: The reader's perspective. *J Second Lang Writ* 43(1):15–23. <https://doi.org/10.1016/j.jslw.2018.02.004>
- Nam M, Beckett GH (2011) Use of resources in second language writing socialization. *TESL-EJ* 15:1–20
- Neville C (2007) *The complete guide to referencing and avoiding plagiarism*. Maidenhead: Open University Press
- Peng J-E (2019) Authorial voice constructed in citation in literature reviews of doctoral theses: Variations across training contexts. *J Engl Acad Purp* 37:11–21. <https://doi.org/10.1016/j.jeap.2018.11.001>
- Petric B (2007) Rhetorical functions of citations in high- and low-rated master's theses. *J Engl Acad Purp* 6(3):238–253. <https://doi.org/10.1016/j.jeap.2007.09.002>
- Petric B, Harwood N (2013) Task requirements, task representation, and self-reported citation functions: An exploratory study of a successful L2 student's writing. *J Engl Acad Purp* 12:110–124. <https://doi.org/10.1016/j.jeap.2013.01.002>
- Planks L, Gebril A (2012) A close investigation into source use in integrated second language writing tasks. *Assess Writ* 17:18–34. <https://doi.org/10.1016/j.asw.2011.09.002>
- Qian X-J, Zhang W-X (2021) *Academic writing in TCSOL: A course book* [汉语国际教育论文写作教程]. Peking University Press [北京大学出版社]
- Samraj B (2008) A discourse analysis of master's theses across disciplines with a focus on introductions. *J Engl Acad Purp* 7(1):55–67. <https://doi.org/10.1016/j.jeap.2008.02.005>
- Samraj B (2013) Form and function of citations in discussion sections of master's theses and research articles. *J Engl Acad Purp* 12(4):299–310. <https://doi.org/10.1016/j.jeap.2013.09.001>
- Shi L (2004) Cultural backgrounds and textual appropriation. *Lang Aware* 15(4):264–282. <https://doi.org/10.2167/la406.0>
- Sun QY, Kuzborska I, Bill S (2022) Learning to construct authorial voice through citations: A longitudinal case study of L2 postgraduate novice writers. *System* 106:1–13. <https://doi.org/10.1016/j.system.2022.102765>
- Swales JM (1990) *Genre analysis: English in academic and research settings*. Cambridge University Press
- Swales JM (2004) *Research genres: Explorations and applications*. Cambridge University Press
- Swales JM (2014) Variation in citational practice in a corpus of student biology papers: From parenthetical plonking to intertextual storytelling. *Writ Commun* 31(1):118–141. <https://doi.org/10.1177/0741088313515166>
- Tardy CM (2016) Voice and identity. In Manchón RM, Matsuda PK (Eds.), *Handbook of Second and Foreign Language Writing*, 349–363. Berlin: De Gruyter Mouton
- Thompson G, Ye Y (1991) Evaluation in the reporting verbs used in academic papers. *Appl Linguist* 12(4):365–382. <https://doi.org/10.1093/applin/12.4.365>
- Thompson P (2005) Points of focus and position: Intertextual reference in PhD theses. *J Engl Acad Purp* 4(4):307–323. <https://doi.org/10.1016/j.jeap.2005.07.006>
- Thompson P, Tribble C (2001) Looking at citations: Using corpora in English for academic purposes. *Lang Learn Technol* 5(3):91–105
- Tseng M-Y (2018) Creating a theoretical framework: On the move structure of theoretical framework sections in research articles related to language and linguistics. *J Engl Acad Purp* 33:82–99. <https://doi.org/10.1016/j.jeap.2018.01.002>
- Weissberg R, Buker S (1990) *Writing up research: Experimental research report writing for students of English*. Englewood Cliffs, NJ: Prentice Hall Regents
- Wette R (2017) Source text use by undergraduate post-novice L2 writers in disciplinary assignments: Progress and ongoing challenges. *J Second Lang Writ* 37:46–58. <https://doi.org/10.1016/j.jslw.2017.05.015>
- Wette R (2018) Source-based writing in a health sciences essay: Year 1 students' perceptions, abilities and strategies. *J Engl Acad Purp* 36:61–75. <https://doi.org/10.1016/j.jeap.2018.09.006>
- Yu H, Zhang S-J (2021) A study on heteroglossia in Chinese academic discourse [汉语学术语篇的多声系统探究]. *Contemp Rhetor* [当代修辞学] 6:49–68
- Zhou X-B (2021) *Research design and thesis writing in teaching Chinese to speakers of other languages* [汉语国际教育研究设计与论文写作]. Foreign Language Teaching and Research Press [外语教学与研究出版社]

Acknowledgements

This study was funded by the The National Social Science Fund of China, The Study on the Interpersonal Function of Chinese Academic Texts Written by L2 Speakers (Grant No. 24CYY086).

Author contributions

RL designed the study, performed data analysis, and wrote the manuscript. KC contributed to interpretation of the data, interpretation of results, wrote the manuscript and made the critical revision. PX designed the study, collected the data and made the critical revision. All authors have read and approved the final version of the manuscript. All contributed authors have been listed in this article.

Competing interests

The author declares no competing interests.

Ethical approval

I hereby confirm that all research was performed in accordance with relevant guidelines/regulations. The masters' theses and journal papers in our corpus used are from open-source online. Ethical approval was not required as the study did not involve human participants.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Correspondence and requests for materials should be addressed to Ruoxi Liu or Ken Chen.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025