Journal of Academic Writing
Vol. 14 No. 2 Winter 2024, pages 1-17
https://doi.org/10.18552/joaw.v14i2.1041

Authorial Voice in Academic Articles: A Corpus-Based Analysis of Citation Types and Citation Presentation across Disciplines

Elizabeth Hanks
Northern Arizona University, United States

Grant Eckstein, Jacob Rawlins, Haley Briggs, and Leanne Chun
Brigham Young University, United States

Abstract

Authorial voice plays a key role in helping writers establish themselves as experts in their field as well as demonstrate their individual style (e.g., Tardy, 2012). Citation usage has an important impact on authorial voice in academic writing and can be implemented in various ways; namely, through citation types (e.g., integral, non-integral) and citation presentation (e.g., direct quotes, summaries, generalizations). While many researchers have examined citation type among novice and experienced writers, researchers have largely overlooked citation presentation across disciplines – that is, how experienced authors balance the use of quotations, summaries, and generalization to index authorial voice. Beginning academic writers may be encouraged to use quotations to prevent plagiarism, but it is unclear if this advice reflects patterns in published writing across disciplines. In this study, we examine the background sections (i.e., introductions and/or literature reviews) of 270 academic research papers to evaluate the extent to which various citation types and presentations are used in background sections across six disciplines. Findings which can inform disciplinary writing guides and educational materials indicate disciplinary variation in citation type, with applied linguistics using the most citations overall and physics and biology using the fewest integral citations. Disciplines also differed in their citation presentation, with some favoring summaries and others favoring generalizations while quotation was rare overall. These results have important implications for teachers and material developers who can use these patterns of source usage to compare and contrast disciplinary norms and provide direct instruction on features of academic voice. Cross-disciplinary awareness of voice features can also highlight disciplinary patterns for students, allowing them to write more like experts in their fields.

Introduction

University students must produce complex written texts such as research reports, theses, dissertations, or publishable work during their studies. Yet many students struggle with the writing process because they are unsure how to write academically (Prescott, 2018); moreover, their advisors often struggle to define the features that make academic writing sound appropriate (Kamler & Thomson, 2006). Understanding the elements of authorial voice as well as how appropriate authorial voice can be incorporated in student writing can help to alleviate some of the challenges faced by both students and advisors. In the last decade, applied linguists have built on earlier notions of voice in writing (see Biber, 2006; Flowerdew, 2002; Matsuda, 2001) to further explore the lexical and syntactic features that make writing sound academic and have found useful patterns in terms of authorial voice (e.g., Bondi, 2014; Lorés-Sanz, 2011; Peng, 2019; Silver, 2012).

It is worth noting that voice is notoriously difficult to define since its effect results in a reader's overall impression of an author or piece of writing. Matsuda (2001) explained that voice is "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" (p. 40). Silver (2012) classifies voice in a text as a consequence of authorial engagement through reference and citation toward other experts within variously specified fields. Authorial voice is absorbed by readers at levels dependent on the exemplification of the writer (Silver, 2012) and sets authors apart from one another by individualistic tendencies of literature that are shown to expand across genre (Tardy, 2012). Just as our spoken voice can take on different emotions and eccentricities, our written voice can similarly take on various identities while remaining individualistic (Tardy, 2012). Nevertheless, and perhaps unsurprisingly, several early researchers felt that academic writing in particular was fairly devoid of voice in that it was intended to be objective, neutral, and detached (e.g., Biber & Finegan, 1989; Mauranen & Bondi, 2003), but recent research has found variability and intentionality in the use of verbs that mark both stance and voice (e.g., Biber, 2006; Di Carlo, 2015; Eckstein et al., 2022; Flowerdew, 2002; Hyland & Guinda, 2012; Hyland & Jiang, 2017). Such research has analyzed reporting verbs for the ratio of integral/non-integral citations and the use of summaries or quotations; additional studies have revealed differences in how authorial voice is enacted across disciplines (Charles, 2006; Hyland, 1999, 2004; Swales, 1990, 2014; Thompson, 2001; Thompson & Tribble, 2001).

Establishing discipline-specific authorial voice in academic writing is considered imperative for new writers because it allows them to engage in "meaningful and recognizable forms of literate activity" within specific spheres (Prior, 2001, p. 79). By learning about strategies for establishing voice, students can effectively develop their own authorial voice in order to participate in academic "spheres of activity" (Prior, 2001, p. 76). Several features that contribute to authorial voice include references to previous research and other parts of the text (Kawase, 2015), hedges (Kim & Lim, 2013), boosters (Kim & Lim, 2013), attitude markers (Bruce, 2014; Kim & Lim 2013; Matsuda & Tardy, 2007), and cues from sentence structure, writing conventions, and theoretical framework (Tardy & Matsuda, 2009). However, it is difficult for novice writers to find the line between quoting, over-citation, and plagiarism (Davis, 2013; Thompson & Tribble, 2001). They may struggle to integrate citations by connecting their thoughts with those of the cited author in both the main theme of their paper and in the manufacturing of their own sentences (Petrić, 2012), which is why students may struggle to reach a proficient level of source integration in their specific fields (Davis, 2013).

In this paper, we analyze elements of authorial voice as a step towards clarifying different disciplinary expectations which can help both students and instructors. Elements of authorial voice that we focus on in this paper include those of citation type and citation presentation because these elements both relate to building authorial voice through source attribution but have seldom been studied in conjunction with one another. By understanding citation types and citation presentation within a discipline, students can develop a discipline-specific authorial voice, effectively participate in ongoing research conversations, and establish themselves as respected members of their academic field. However, it is important to note that while writing guides for various disciplines may touch on citation types and presentation, they often lack specific details about disciplinary standards for utilizing these elements effectively, such as when to use integral or non-integral citations, or when to employ summary or generalization (Thompson & Tribble, 2001). For instance, a biology textbook may mention that citations should be included for figures, non-original concepts, and the names of individuals who discovered organisms but overlook information on citation practices beyond these general guidelines (Matthews & Matthews, 2007).

While students often have few resources to help them navigate the nuances of citation within their discipline, scholars eventually seem to acquire proficiency in these unspoken rules through experience. However, explicit and correct instruction can significantly accelerate students' learning process in developing their authorial voice. Writing instructors who possess knowledge of cross-disciplinary differences in citation standards and source presentation in academic research articles can play a vital role in guiding students to effectively apply the conventions most relevant to their respective fields. By teaching students these discipline-specific norms,

writing instructors can help them to cultivate their authorial voice and enhance their overall writing skills.

For these reasons, there is a pressing need for research-based materials and comprehensive writing guides that cater to both students and educators. These resources should provide detailed information about the expectations and norms surrounding citation types and citation presentation within specific disciplines. Such materials would facilitate the learning process for students and empower educators to effectively guide students in meeting the unique requirements of their fields. This will result in an improvement of students' awareness of different citation styles and disciplinary references, enabling them to engage more effectively with the scholarly community and excel in their academic endeavors.

The present study examines the use of citations and quotations within academic research papers to determine how these elements of authorial voice are constructed across six disciplines by examining source use in published research articles. Such empirically grounded research can aid teachers and students in understanding and using discipline-specific aspects of authorial voice.

Literature Review

Citations create an ongoing conversation between writers and readers that ensures a continuous contribution to research questions (Hyland, 1999), and various forms of citations allow authors to "situate their research in a larger narrative" on the discourse level (Hyland, 1999). The arrangement of new and old knowledge demonstrates constant evolution and creates a bond between the reader and writer (Hyland, 1999). In this sense, citations may align the writer with a particular group or provide a path for critiquing a school of thought. Used well, citations help authors to establish credibility because both the writer and the referenced authors contribute to an existing academic conversation (Peng, 2019). Moreover, citations are strategically arranged by the writer to establish association between the cited author's information and the writer's purpose and context (Hyland & Jiang, 2017; Nanba et al., 2000).

The citation practices analyzed in the present study involve both citation types (i.e., whether the citation is integrated grammatically or as a parenthetical) as well as citation presentation (i.e., whether the cited material is integrated as a direct quotation, summary, or generalization). These citation practices vary across disciplines, and successful students typically aim to follow the citation patterns characteristic of their discipline, rather than choosing it themselves. This is particularly relevant when considering that various style guides allow or restrict the use of particular citation practices. Namely, some citation styles such as IEEE primarily use non-integral citations due to their numeric referencing system, whereas styles such as APA or MLA allow both types, helping writers to highlight either the author's identity or the information being presented. The following sections outline previous research on the patterns of citation types and presentation across disciplines.

Citation types

Swales (1990) distinguishes between integral citations, in which the name of the cited researcher is used grammatically within the sentence, and non-integral citations, in which researcher names are listed in parentheses or footnotes that follow the sentence. Typically, integral citations prioritize the cited author (e.g., "person-focused" citation) while non-integral citations prioritize their message, findings, or ideas (e.g., "concept-focused" citation) (Hyland, 1999; Peng, 2019; Swales, 2014), yet both help to establish a narrative formed by the writer. Because integral citations are seemingly led by the referenced author while non-integral citations are spearheaded by the writer, Peng (2019) posits that authorial voice is more prominent in non-integral citations.

Previous studies have compared variations in citation type by language (Hu & Wang, 2014), country (Peng, 2019), discipline (Davis, 2013; Hyland, 1999), writing proficiency (Mansourizadeh & Ahmad, 2011; Swales, 2014), and genre (Jalilifar, 2012). Differences can be found in the amount, form, and presentation of citations across these variables, and research

has suggested that emerging writers must acquire discipline-specific patterns of citation type usage. For example, in hard sciences such as engineering, non-integral forms are only slightly favored (Hyland, 1999; Jalilifar, 2012; Thompson, 2001), yet non-integral citations are used widely among biology students to cite background information (Swales, 2014). In addition, published research articles in applied linguistics include a relatively even mix of integral and non-integral citations, but Jalilifar (2012) identified more integral than non-integral citations in applied linguistics master's theses. Findings like this suggest a developmental process in citation use among graduate students. Because of the variation found within citation practices, students pursuing careers in specific disciplines must "embark on the arduous process of learning to cite in such a manner that their academic papers are increasingly persuasive and convincing" using appropriate strategies for their particular field (Swales, 2014, p. 119). Although discipline-specific use of citation types is learned over time, not all students reach a proficient level (Davis, 2013). By enhancing our understanding of citation types across disciplines, instructors may be better able to assist students in acquiring discipline-specific writing skills.

Citation presentation

Another element of citation use in authorial voice is what we call citation presentation, which includes three presentation options: direct quotes, summaries of an individual source, and generalizations of two or more sources. When including citations, expert writers may incorporate more direct quotations than novice or inexperienced writers (Petrić, 2012). Of eight high-rated and eight low-rated master's theses in gender studies, the high-rated theses displayed almost three times as many direct quotations per 1,000 words. These quotes were more often fragments, shorter than a T-unit of text, which allowed advanced writers to frame information in their own textual context (Petrić, 2012). Similarly, Davis (2013) found that as students progressed through their postgraduate schooling, they began to use a variety of citation presentations and shorter quotations. Both studies conclude that more variety of citation presentations makes the writing sounds more sophisticated (Davis, 2013; Petrić, 2012). Further research has indicated that disciplinary variation exists in terms of quotation usage. While the social sciences and humanities tend to include quotations, research papers in the hard sciences rarely incorporate direct quotes, whether as fragments or clauses (Hyland, 1999), suggesting variety in disciplinary approaches to citation presentation.

While some citations can be integrated through direct quotations, others are integrated through paraphrasing. Two forms of paraphrases – *summaries* that summarize one author's contributions in one parenthetical and *generalizations* that summarize many authors' contributions (Hyland, 1999) – are both favorable to writers of shorter texts because they require less space than direct quotes (Petrić, 2012). Because of the frequency with which it is used, paraphrasing is an integral skill for students to acquire. The use of direct quotations, on the other hand, depends heavily on discipline and requires both careful attention to disciplinary conventions and "academic literacy" to make the quotations integral to the writing (Petrić, 2012).

The present study

The present study investigates the background sections (e.g., introductions and/or background sections) of published academic research articles in six disciplines (applied linguistics, biology, history, philosophy, physics, and political science). Background sections were examined in particular because they offer the most concentrated collection of citations and are comparable across disciplines.

The present study asks the following research questions:

1. To what extent are citation types (i.e., integral and non-integral) used in the background sections of academic articles across disciplines?
2. To what extent are different citation presentations (i.e., direct quotations, summaries, and generalizations) used in the background sections of academic articles across disciplines?

The findings presented below illuminate current trends and present information that may benefit students and teachers of academic writing.

Method

Data acquisition

We collected 270 background sections (combined introduction and literature review sections) from Gray's (2011) Academic Journal Registers Corpus (AJRC) for analysis. The AJRC comprises 270 articles containing 1,952,568 tokens and 48,403 types. The articles are distributed across six disciplines: applied linguistics (60 articles), biology (30 articles), history (30 articles), philosophy (30 articles), physics (60 articles), and political science (60 articles). Gray (2011) collected the articles for the AJRC from reputable peer-reviewed journals that cover broad topics within each discipline, representing prototypical examples of the field. The selection process involved consulting experts in each discipline who provided guidance on the research category suitable for inclusion in the corpus. Randomly chosen articles from three journal issues (2006, 2007, and 2008) were added to the corpus if they aligned with the expected research type for their respective discipline.

To ensure diversity within each discipline, different research categories were included based on disciplinary norms. Applied linguistics and political science each consist of 30 qualitative and 30 quantitative articles, biology includes 30 quantitative articles, history encompasses 30 qualitative articles, philosophy contains 30 theoretical articles, and physics comprises 30 quantitative and 30 theoretical articles. This systematic approach ensured a holistic representation of each field. Additionally, the range of disciplines spans from hard to soft sciences, offering a broad scope for investigation.

One reviewer of this paper noted the age of the corpus as a limitation of our research, though it was still contemporary when we began our research project. We chose to continue with the selected corpus because of the rigorous selection process and expert consultation that ensured the inclusion of prototypical examples from each field (Egbert, 2016; Gray, 2015). Furthermore, the fundamental norms and conventions of academic writing in the disciplines covered by the AJRC are relatively stable over time, and variationist researchers regularly compare newer corpora (often convenience sampled from free repositories, e.g. Swales, 2014) with original corpora that is decades old (e.g., Hyland, 1999; Swales, 1990; Thompson & Tribble, 2001) to substantiate their claims. Thus, we argue that the AJRC remains a valuable resource for understanding disciplinary writing practice, and practical constraints such as time and resource limitations to create an equally rigorous and sophisticated corpus further justify the use of this well-established resource.

We extracted data from only the background sections of each article, totaling 292,547 tokens and 19,569 types. The background section (including the introduction and literature review) offered the most concentrated collection of citations across all of the disciplines. Other sections (including methods, results, discussion, and conclusion) were excluded from our data because most disciplines engage with source attribution in the background section. This approach allowed for a direct comparison of the sections where authors in all disciplines actively reference previous research.

Our team (including the authors of this paper and three undergraduate researchers who are studying linguistics, statistics, and editing and publishing) manually tagged the background sections in Dedoose, a platform for collaborative data analysis. All taggers were trained following the guidelines presented below, and they practiced on at least 12 articles and acquired at least 5 hours of coaching before beginning to tag the corpus. At the end of training, all taggers reached an interrater reliability score of 0.95 and then tagged the introduction and literature review sections for all 270 background sections (3,958 tagged items).

Tagging guidelines

Citations were tagged as *non-integral* if referenced authors (or numbers that corresponded to referenced authors) were listed in parentheses (e.g., at the end of a sentence). All citations enclosed within one parenthetical were tagged together as a single non-integral citation. Integral citations were tagged because of their grammatical integration with the sentence. That is, a cited author could be the subject of an active sentence, the object of a passive sentence, the adjunct of a prepositional phrase in a dependent clause (e.g., "According to Lee (2015)..."),

or part of a noun phrase ("Lee's (2015) investigation..."). When cited authors were incorporated both integrally and non-integrally in the same sentence, it was tagged as *integral citation*. Table 1 provides an overview of these criteria with examples.

Table 1. Classification of citation types

Citation Type	Example
Integral	Nevo (1989) investigated the strategies activated while processing a test in their first language and in the target language.
	The comparison of young and mature voter responsiveness to GOTV appeals by Nickerson (2002) and Bennion (2005) represent the sale efforts among recent field experiments to tackle the fourth concern.
Non-integral	Different populations often appear adapted to a multitude of elevational, climatic, and edaphic habitats (Vickery 1978; Wu et al. 2008).

We tagged citation presentation in terms of summaries, generalizations, and direct quotations. Each sentence referencing another author was tagged as either a *summary*, in which one cited author was paraphrased, or *generalization*, in which two or more authors were paraphrased. In contrast, direct quotations were tagged when the writer directly quoted another author's words. We differentiated between three types of direct quotations. *Fragments* were quotations shorter than a T-unit, or an independent clause (Hunt, 1965), *brief quotations* were at least one full T-unit that was shorter than 40 words, and *block quotations* were quotations indented as blocks that were longer than 40 words. In some cases, a writer's direct quotations were either block quotations shorter than 40 words or T-units longer than 40 words. When this occurred, we referred to whether the quotation was indented as a block. If it was, we tagged it as *block quotation*; if not, we tagged it as a *brief quotation*. Table 2 provides examples of each of the tagging categories and subcategories within citation presentation.

Table 2. Classification of citation presentation

Integration Type	Subcategory	Example
Direct Quotation	Fragment	Foreign Affairs and International Trade (DFAIT) labels Mexico a "strategic partner" because of NAFTA...
	Brief Quotation	He commented, "Due to the flaws in the test or due to certain test taking strategies, students may not be displaying a representative performance of their language competence" (p. 71).
	Block Quotation	As Lee (2015) concludes, *block quotation*
Summary		In addition, research has indicated that cultural capitol, including prior knowledge and language skills, acts as a filter in disadvantaging L2 students who do not possess English language proficiency (Garcia & Gopal, 2003).
Generalization		Marxist historians have been even less indulgent to the NAACP (Lee, 2006; Rice, 2001).

Data analysis

After tagging all introductions and literature reviews for citation type and citation presentation, data was exported to Excel for analysis. We examined each feature separately by summing all instances and norming them to 100 papers. We chose to normalize frequencies by paper as opposed to word count to focus on meaningful units of text that can make interpretation easier and more relevant (Eckstein et al., 2022; Egbert et al., 2020). Analyses were first carried out across disciplines as well as article types (i.e., quantitative, qualitative, and theoretical articles). Because differences across article types within a given discipline were minimal, we chose to focus our analysis on differences across disciplines.

Results

Citation type

Our first research question queried the extent to which various citation types were used in background sections across disciplines. Figure 1 depicts counts of total citations across the six disciplines, normed to 100 papers. The disciplines with the fewest citations overall were history

(1,073) and philosophy (817), while political science (1,320), biology (1,544), and physics (1,659) used citations more frequently. Applied linguistics (3,154) utilized nearly twice as many citations as any other discipline.

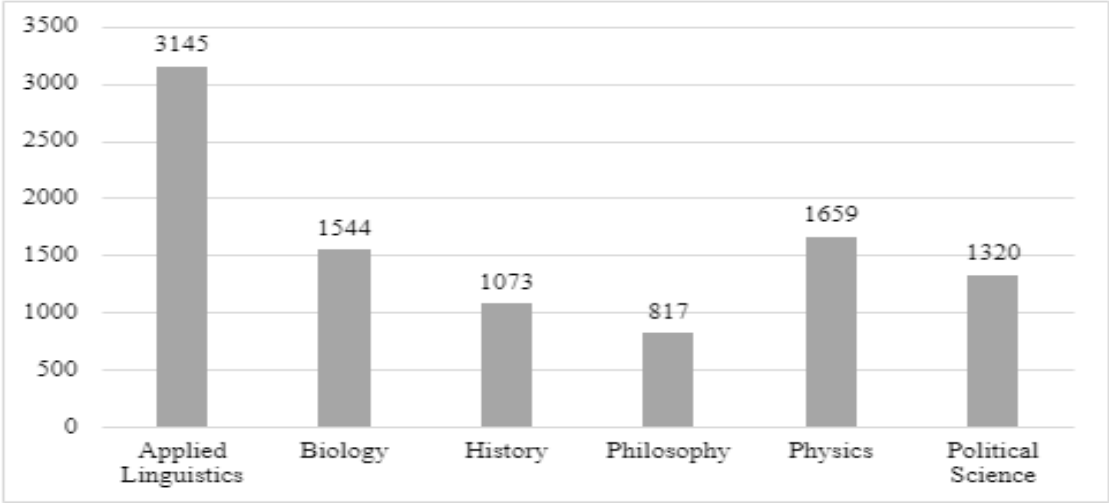

Figure 1. Total normed counts of citations per 100 background sections

The distinction between integral and non-integral citations across disciplines is shown in Fig. 2. Non-integral citations were used more often in most disciplines, and the difference between frequencies of citations type is stark in biology, physics, and history which all showed integral citation use to be less than half of non-integral usage. Biology differed from other disciplines in that all types of integral citations, calculated to be only 40 per 100 papers, were rarely incorporated. On the other hand, applied linguistics and philosophy both displayed a slightly greater frequency of integral citations at 10% and 35% more integral than non-integral citations, respectively. Raw frequency counts for types of integral and non-integral citations are available in the appendix.

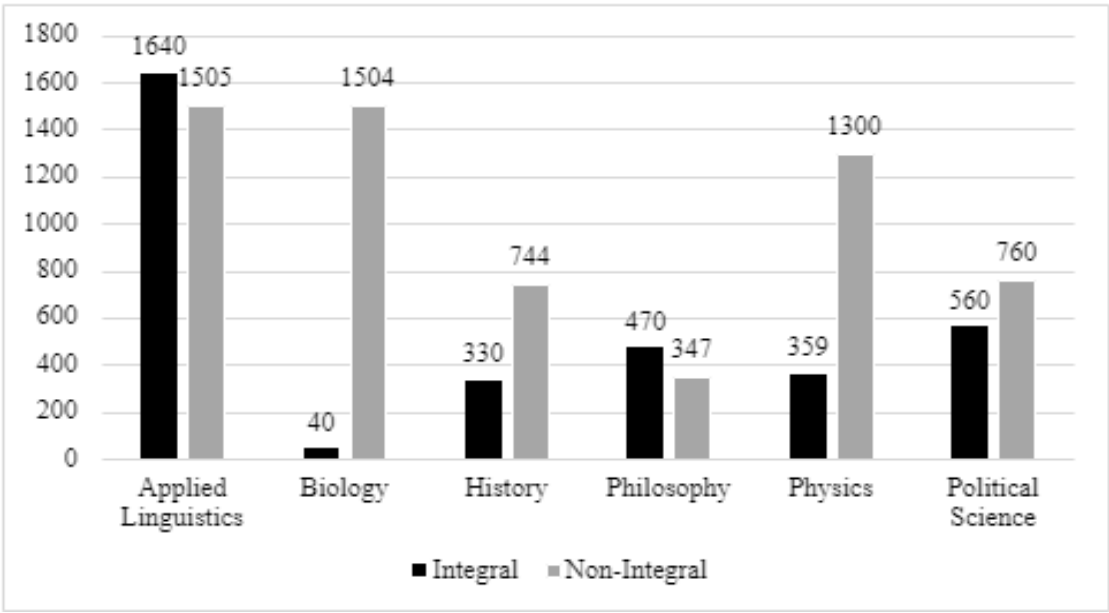

Figure 2. Normed counts of integral and non-integral citations per 100 background sections

Citation presentation

Our second research question asked to what extent different citation presentations were used in background sections across disciplines. Figure 3 includes counts of direct quotations, summaries, and generalizations normed to 100 background sections per discipline. A common

pattern emerged in applied linguistics, philosophy, and political science: summaries occurred about twice as frequently as either direct quotations or generalizations. Biology and physics both had low counts of direct quotations (7 and 24 quotes per 100 background sections, respectively), and biology and history had relatively high counts of generalizations (744 and 637 generalizations, respectively). Applied linguistics had particularly high frequency counts for all citation presentations, especially summaries (2,057). Summaries in applied linguistics were over 2.5 times more frequent than those of political science, the discipline with the next most frequent counts (779 summaries). Summary and generalization presentations had the highest means and maximum counts in all disciplines, while block quotations had the lowest means and maximum counts in all disciplines. Applied linguistics incorporated summaries and generalizations at the highest frequencies and with the greatest variation. Each discipline included fragments and brief quotations at relatively low frequencies. Biology, physics, and philosophy all had fragments, brief quotations, and block quotations that occurred particularly infrequently: at less than an average of 0.10 uses in most cases and up to an average of 0.97 uses in one case (fragments in philosophy).

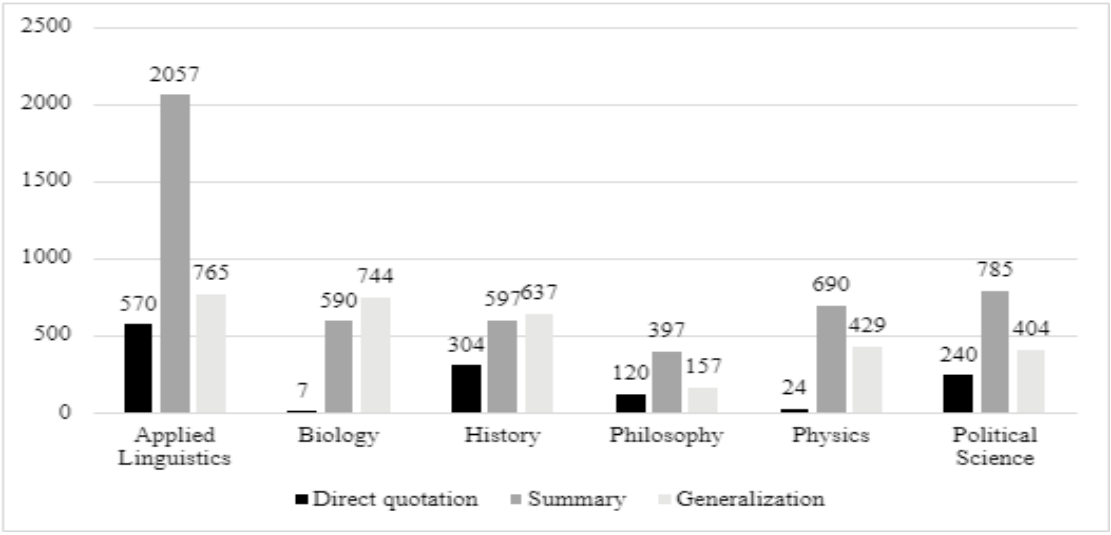

Figure 3. Total normed counts of citation presentations per 100 background sections

Discussion

Citation type

Overall citation frequency

The results suggest that disciplines may hold different standards for what type of information requires citation. Applied linguistics incorporated the most citations in their background sections (3,145 per 100 papers), which may be because linguists cite to not only attribute the ideas of others but also their findings and arguments (Murray et al., 2012), to build a foundation that can mold research to support or refute claims (Hyland, 1999; Murray et al., 2012; Peng, 2019), and to build credibility (Murray et al., 2012, p. 58). Therefore, it may be that citations are more frequent in applied linguistics background sections because they are used for various reasons beyond that of merely crediting intellectual ownership.

On the other hand, history incorporated citations among the least frequently (1,073 per 100 papers). This may at first seem surprising given that history articles draw on past events; however, the fact that citations in history do not all follow statements of fact is evidence that writers in history primarily use citations to credit others' ideas (Kaplan, 1965; Mooney, 2011). Even a guide to writing in history from Southwestern University reminds developing scholars that "dates, locations, and basic biographical details are considered to be general knowledge" (Byrnes, n.d, p. 7) and therefore do not need a citation. The following excerpt from the AJRC depicts the tendency not to cite factual information in history (emphasis added):

> While often overshadowed by Tel Aviv, its younger and larger neighbor to the north, which since *its founding in 1909* was conceived of as the *"most modern city in the Middle East,"* Jaffa was in fact the *pre-1948 economic and cultural capital of Arab Palestine*. As such, *by the late 1940s it was home to upwards of seventy thousand Palestinian Arab inhabitants, thirty thousand Jews,* and a complex set of economic, social, and cultural relations within and between both communities. (HIST-GEN-005-jwh)

Philosophy articles also had low counts of citations (817 per 100 papers), which could partially be due to the depth at which each source was discussed. For instance, it was common for one citation to refer to information discussed over several sentences or paragraphs. In some cases, philosophy articles only included one or two citations, and in fact, one background section in our corpus did not include any citations. The infrequent use of citations may also be because philosophy is a humanistic discipline that deals with the writer's ideas more than the ideas of other scholars (Williams, 2006) and therefore does not require many references.

The only other discipline that had some articles with no citations was political science, in which six articles did not use any citations in their background sections. Perhaps, like history, dates and statistics are not cited in political science because they are not the intellectual property of others (Kaplan, 1965; Mooney, 2011). A guide for students writing in political science suggests that "you have to be as specific as possible in your footnotes or endnotes when discussing *ideas developed previously* [emphasis added] by someone else" (Martel, 1997, p. 29). Thus, because historical events and current events are facts rather than ideas, they do not require a citation in political science articles.

Integral and Non-integral Citation Frequency

A few disciplines had a strong preference for the use of non-integral over integral citations in their background sections, the most evident of which was biology. This finding aligns with those of Hyland (1999), Jalilifar (2012), Swales (2014), and Thompson (2001) that show hard sciences, such as biology and physics, favor non-integral citations. By emphasizing facts through non-integral, concept-focused citations rather than integral, person-focused citations (Swales, 2014), biologists may reduce the appearance of bias in their writing. As one biologist records, "Scientists are keen to avoid bias of any kind because they threaten scientific ideals such as objectivity, transparency and rationality" (Andersen et al., 2019, para. 1).

The effort to avoid bias could also motivate writers in physics to use non-integral citations much more frequently than integral citations, since "the emphasis is on how careful and correct reasoning leads to correct results" (Jeng, 2006, p. 1). The use of non-integral citations in physics articles is such a common practice that physics style guides typically do not even mention the possibility of integral citations: "Each reference must be made immediately after the information is given" (Bottorff, n.d., p. 8).

History is the other discipline that had a clear preference for non-integral over integral citations in background sections with more than double the amount of non-integral than integral citations. A qualitative analysis of history articles in the corpus reveals a pattern of non-integral citations being used following descriptions of events in history, with integral citations being used to introduce individual views during that historical time. In other words, non-integral citations seem to be used to relate historical events, whereas integral citations are used namely to introduce the ideas of people who lived in a given historical time period. For example, the following excerpt includes a non-integral citation in the noun phrase in order to introduce personal thoughts of Smith during a specific time in history: "According to Margaret Bayard Smith, 'her appearance was such that it threw all the company into confusion, and no one dared look at her but by stealth'" (HIST-GEN-021-juh). However, more research is needed to confirm this preliminary finding. Also, as with physics articles, non-integral citations in history may be incorporated more often overall in order to place emphasis on the event rather than the person reporting it.

Conversely, a slight preference for integral citations rather than non-integral citations was demonstrated in applied linguistics and philosophy background sections. This may reflect an emphasis on engaging in conversation with previous, present, and future scholars within a published article (Hyland, 1999). Although writers in applied linguistics and philosophy may

engage with other researchers through the frequent use of integral citations, philosophy and history writers engage with researchers through non-integral citations in the footnotes. Scholars in both fields occasionally place value judgments on the sources they cite through superscripts. In history, for example, one author includes the footnote, "Lisa McGirr's recent *insightful* [emphasis added] book on the New American Right in Orange County emphasizes" (HIST-GEN-027-whq). Similarly, a philosopher cites, "see the *thoughtful* [emphasis added] essay by Rickless" (PHIL-TH-020-pos) in a footnote. This type of subjective, interactive commentary found within citations is unique to history and philosophy in our corpus.

Citation Presentation

Citation presentation, in terms of direct quotations, summaries, and generalizations, also varied by discipline. Applied linguistics had particularly high frequency counts for all citation presentations, particularly summaries, which follows the fact that applied linguistics also utilized the most citations. This contrasts with findings of Mansourizadeh and Ahmad (2011), who found that novice writers typically make use of summaries while expert writers use generalizations more often to "succinctly synthesize various sources" (p. 152). While summaries were more frequent than generalizations, generalizations were more common in each discipline than direct quotations.

Direct quotations were the least common in the background sections of all disciplines. This may be due to the fact that writing and style guides in many fields discourage the use of quotations. For instance, a web-based writing guide for linguistics at University of Washington first advises writers to break up long quotes and then admonishes that "there is *very rarely* any reason to put a direct quote in a paper. It is always much better to paraphrase the material that you want to cite in your own words" (Newmeyer, n.d., para. 11). This is somewhat ironic advice given that applied linguistics background sections had the greatest number of quotations among the fields examined in this study – nearly six per background on average, which is almost double that found in history. Meanwhile, in history, students are told that "the less you quote other people, and the more you explain their findings in your own words… the higher the quality of your paper will be" (Martel, 1997, p. 12). Biology guides warn that direct quotations could easily lead to unintentional plagiarism and mention, "if you must quote directly, use extreme care" (Matthews & Matthews, 2007, p. 3), which is advice that appears to reflect disciplinary norms given that biology backgrounds only contain about one quotation for every 14 background sections. Compared to other disciplines, biology and physics both have noticeably infrequent counts of direct quotations. Again, this may be due to the emphasis on facts as opposed to sources in these disciplines (Jeng, 2006) in order to promote objectivity (Andersen et al., 2019).

While direct quotations were less common overall in our corpus, the most frequent type in all disciplines, with the exception of a slight tendency towards brief quotations in applied linguistics, was fragments. This aligns with Petrić's (2012) finding that more advanced writing incorporated shorter quotations than writing of less expert counterparts.

Although students appear to be discouraged by style guide authors from using direct quotations, studies suggest that more advanced written work, such as high-rated master's theses, contain more direct quotes than less advanced work, such as low-rated master's theses (Petrić, 2012) and that a greater variety of citation presentations makes written work more sophisticated (Davis, 2013). Overall, learners who are not aware of legitimate disciplinary norms or are misled by well-meaning but sometimes inaccurate opinions of style guide authors may benefit from more authoritative guidance regarding how citations can be utilized to foster a discipline-specific authorial voice.

Conclusion

The present study explored two concepts that relate to a writer's authorial voice in academic writing: citation types (integral and non-integral) and citation presentation (summaries, generalizations, direct quotes). We found that writers in applied linguistics tended to use more citations, indicating their desire to give credit to other authors and boost their own credibility. On the other hand, philosophy, history, and political science writers tended not to have as many

citations because of the depth of their analysis for each cited work. Physics and biology authors rarely incorporate integral citations.

Applied linguistics writers and philosophy writers used integral citations, whereas philosophy and history writers also utilized non-integral citations to engage with cited authors in the footnotes. These are often more subjective commentary than what is seen in the body of an article. In addition to footnote interaction with cited authors, history writers seemed to use non-integral citations primarily to relate historical events, and they used integral citations to introduce the ideas of someone who lived in a given historical time period. Biology and physics writers may use non-integral citations because the style guide is different (i.e., a digit in parentheses for biology and brackets for physics). In general, these hard sciences seek to eliminate bias by using passive voice and avoiding names of cited authors (i.e., integral citations).

Implications and Applications

The research in this study has important implications and applications for students, educators, and style guide writers. As an empirical investigation into disciplinary citation conventions, the findings illustrate one way in which authors create an authoritative academic voice and how this varies across six disciplines. Students can use this information to supplement and bring nuance to advice they receive from style guides. Furthermore, student and academic writers who must navigate writing in multiple disciplines either for diverse class assignments or as they transition from one scholarly discipline to another, can likewise use information from this study to match the nuanced patterns of citation in a given field. Educators and style guide writers should consider such information prior to offering personal opinions or observations about disciplinary norms that may not hold up to empirical investigation. At the same time, this study models a valuable approach for students and scholars, who can make their own data-driven evaluation of citation use. One such activity might include selecting twenty or so high-quality articles in a discipline and then tallying the number and type of citations to compare them to those in this study in an effort to better understand a particular field or subfield. Style guide writers can likewise take information from this study or design their own as a way to inform or support their claims about disciplinary citation use and conventions.

Limitations

We recognize that our analysis is constrained by several limitations, the first of which is a corpus that, despite rigorous and sophisticated sampling, is nonetheless aging. This justifies conducting further future research with a larger, newer, and broader (but still rigorously selected) corpus. Also, we interpreted writing advice drawn from disciplinary style and writing guides mainly geared toward student writers, but a future analysis should investigate journal author guidelines as these may provide greater insight into disciplinary expectations of citation use. Additionally, the length of background sections in each article varied, which we did not account for because we wanted to focus on meaningful units of language. In addition, our study included only introductions and literature review sections. Although these sections tend to include many of the citations and citation presentation found in a given article, other sections of published articles also include some of these elements, particularly references to previous literature, which may vary by discipline (Lin & Evans, 2012). Thus, our study did not capture all the disciplinary variation in terms of authorial voice that the articles may contain. Future research should examine the rates and types of citation and presentation used within various sections of academic articles. In the meantime, teachers and students can use the findings in this report as guidance in expanding disciplinary use of these items and establishing an authorial voice that projects membership in a community of academic writers.

Further Research

Further research should investigate whether style guides correlate with a discipline's preference for integral or non-integral citations. These analyses might consider examining writing instructors' beliefs about citation practices in relation to students' and academics' actual citation practices. This would allow for a comparison of perceptions and practices in similar rhetorical contexts (i.e., university settings) as opposed to comparing two different rhetorical contexts (i.e., university settings through writing guides and academic research through published research articles) as was done in the present study. Another area of research to be explored is

the way that placement of integral citations (i.e., author as subject, author as object, author as adjunct, author in noun phrase) reflects the emphasis authors attempt to place on the name. This area of research could explore whether or not emphasizing sources is simply a matter of syntactic conventions.

References

Andersen, F., Anjum, R. L., & Rocca, E. (2019). Philosophy of biology: Philosophical bias is the one bias that science cannot avoid. *eLife*, *8*. https://doi.org/10.7554/eLife.44929

Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes, 5*(2), 97-116. https://doi.org/10.1016/j.jeap.2006.05.001

Biber, D., & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text, 9*, 93-124.

Bondi, M. (2014). Changing voices: Authorial voice in abstracts. In M. Bondi, & R. Lores Sanz (Eds.). *Abstracts in academic discourse: Variation and change* (pp. 243–270), Peter Lang.

Bottorff, M. (n.d.). *Guide for writing in physics and dual-degree engineering*. Southwestern University. https://www.southwestern.edu/live/files/4179-guide-for-writing-in-physicspdf.pdf

Bruce, I. (2014). Expressing criticality in the literature review in research article introductions in applied linguistics and psychology. *English for Specific Purposes*, *36*, 85-96. https://doi.org/10.1016/j.esp.2014.06.004

Byrnes, M. (n.d.). *Guide for writing in history*. Southwestern University. https://www.southwestern.edu/live/files/4173-guide-for-writing-in-historypdf.pdf

Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes*, *25*(3), 310-331. https://doi.org/10.1016/j.esp.2005.05.003

Davis, M. (2013). The development of source use by international postgraduate students. *Journal of English for Academic Purposes*, *12*(2), 125-135. https://doi.org/10.1016/j.jeap.2012.11.008

Di Carlo, G.S. (2015). Stance in TED talks: Strategic use of subjective adjectives in online popularisation. *Ibérica, Journal of the European association of languages for Specific Purposes, 29*, 201-222.

Egbert, J. (2016). Review: Linguistic variation in research articles: When discipline tells only part of the story. *English for Specific Purposes*, *44*, 86-87. https://doi.org/10.1016/j.esp.2016.05.001

Egbert, J., Larsson, T., & Biber, D. (2020). *Doing linguistics with a corpus: Methodological considerations for the everyday user*. Cambridge University Press.

Eckstein, G., Rawlins, J., Taylor, H., Briggs, H., Candland, A., Hanks, E., & Hill, S. (2022). Reporting verb variation across disciplines: An academic corpus study. *Journal of Academic Language & Learning, 16*(1), 59-75. https://journal.aall.org.au/index.php/jall/article/view/789

Flowerdew, J. (Ed.) (2002). *Academic discourse*. Routledge.

Gray, B. E. (2011). *Exploring academic writing through corpus linguistics: When discipline tells only part of the story* [Doctoral dissertation, Northern Arizona University].

Gray, B. (2015). *Linguistic variation in research articles: When discipline tells only part of the story*. John Benjamins Publishing Company.

Hu, G., & Wang, G. (2014). Disciplinary and ethnolinguistic influences on citation in research articles. *Journal of English for Academic Purposes*, *14*, 14-28. https://doi.org/10.1016/j.jeap.2013.11.001

Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied linguistics*, *20*(3), 341-367. https://doi.org/10.1093/applin/20.3.341

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for specific purposes*, *20*(3), 207-226. https://doi.org/10.1016/S0889-4906(00)00012-0

Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing.* University of Michigan Press.

Hyland, K. (2008). Disciplinary voices: Interactions in research writing. *English Text Construction*, 1(1), 5–22. https://doi.org/10.1075/etc.1.1.03hyl

Hyland, K., & Guinda, C. S. (Eds.). (2012). *Stance and voice in written academic genres*. Palgrave Macmillan.

Hyland, K., & Jiang, F. (2017). Points of reference: Changing patterns of academic citation. *Applied Linguistics*, *40*(1), 64-85. https://doi.org/10.1093/applin/amx012

Jalilifar, A. (2012). Academic attribution: Citation analysis in master's theses and research articles in applied linguistics. *International Journal of Applied Linguistics*, *22*(1), 23-41. https://doi.org/10.1111/j.1473-4192.2011.00291.x

Jeng, M. (2006). A selected history of expectation bias in physics. *American Journal of Physics*, *74*(7), 578-583. https://doi.org/10.1119/1.2186333

Kamler, B., & Thomson, P. (2006). *Helping doctoral students write: Pedagogies for supervision.* Routledge.

Kaplan, N. (1965). The norms of citation behavior: Prolegomena to the footnote. *American Documentation*, *16*(3), 179-184. https://doi.org/10.1002/asi.5090160305

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

Kim, L. C., & Lim, J. M. H. (2013). Metadiscourse in English and Chinese research article introductions. *Discourse Studies*, *15*(2), 129-146. https://doi.org/10.1177/1461445612471476

Lin, L., & Evans, S. (2012). Structural patterns in empirical research articles: A cross-disciplinary study. *English for Specific Purposes*, *31*(3), 150-160.

Lorés-Sanz, R. (2011). The construction of the author's voice in academic writing: The interplay of cultural and disciplinary factors. *Text & Talk: An Interdisciplinary Journal of Language, Discourse & Communication Studies, 31*(2), 173–193.

Mansourizadeh, K., & Ahmad, U. K. (2011). Citation practices among non-native expert and novice scientific writers. *Journal of English for Academic Purposes*, *10*(3), 152-161. https://doi.org/10.1016/j.jeap.2011.03.004

Mauranen, A., & Bondi, M. (2003). Evaluative language use in academic discourse. *Journal of English for Academic Purposes, 2*(4), 269-271.

Martel, A. (1997). *A student's guide for writing in political science*. Carleton University Press.

Matsuda, P. K. (2001). Voice in Japanese written discourse: Implications for second language writing. *Journal of Second language writing*, *10*(1-2), 35-53. https://doi.org/10.1016/S1060-3743(00)00036-9

Matsuda, P. K., & Tardy, C. M. (2007). Voice in academic writing: The rhetorical construction of author identity in blind manuscript review. *English for Specific Purposes*, *26*(2), 235-249. https://doi.org/10.1016/j.esp.2006.10.001

Matthews, J. R. & Matthews, R. W. (2007). *Successful scientific writing: A step-by-step guide for the biological and medical sciences*. Cambridge University Press.

Mooney, H. (2011). Citing data sources in the social sciences: Do authors do it? *Learned Publishing*, *24*(2), 99-108. https://doi.org/10.1087/20110204

Murray, N. (2012). *Writing essays in English language and linguistics: Principles, tips and strategies for undergraduates*. Cambridge University Press.

Nanba, H., Kando, N., & Okumura, M. (2000). Classification of research papers using citation links and citation types: Towards automatic review article generation. *Advances in Classification Research Online*, *11*(1), 117-134. https://doi.org/10.7152/acro.v11i1.12774

Newmeyer, F. (n.d.). *Some guidelines for writing linguistics papers*. University of Washington Department of Linguistics. https://linguistics.washington.edu/some-guidelines-writing-linguistics-papers

Peng, J. E. (2019). Authorial voice constructed in citation in literature reviews of doctoral theses: Variations across training contexts. *Journal of English for Academic Purposes*, *37*, 11-21. https://doi.org/10.1016/j.jeap.2018.11.001

Petrić, B. (2012). Legitimate textual borrowing: Direct quotation in L2 student writing. *Journal of Second Language Writing*, *21*(2), 102-117. https://doi.org/10.1016/j.jslw.2012.03.005

Prescott, F.J. (2018). In at the deep end: The struggles of first-year Hungarian university students adapting to the requirements of written academic discourse in an EFL context. In: Chitez, M., Doroholschi, C., Kruse, O., Salski, Ł., Tucan, D. (Eds.) *University Writing in Central and Eastern Europe: Tradition, Transition, and Innovation* (pp. 167-188). Springer. https://doi.org/10.1007/978-3-319-95198-0_12

Prior, P. A. (2001). Voices in text, mind and society: Sociohistoric accounts of discourse acquisition and use. *Journal of Second Language Writing*, 10, 55–81. https://doi.org/10.1016/S1060-3743(00)00037-0

Silver, M. (2012). Voice and stance across disciplines in academic discourse. In K. Hyland & C. Sancho Guinda (eds.), *Stance and voice in written academic genres* (pp. 202–217). Palgrave Macmillan.

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

Swales, J. M. (2014). Variation in citational practice in a corpus of student biology papers: From parenthetical plonking to intertextual storytelling. *Written Communication*, *31*(1), 118-141. https://doi.org/10.1177/0741088313515166

Tardy, C. M., & Matsuda, P. K. (2009). The construction of author voice by editorial board members. https://doi.org/10.1177/0741088308327269

Tardy, C. M. (2012). Current conceptions of voice. In *Stance and voice in written academic genres* (pp. 34-48). Palgrave Macmillan, London. https://doi.org/10.1057/9781137030825_3

Thompson, P. (2001). *A pedagogically-motivated corpus-based examination of PhD theses: Macrostructure, citation practices and uses of modal verbs*. Reading, UK: University of Reading.

Thompson, P., & Tribble, C. (2001). Looking at citations: Using corpora in English for academic purposes. *Language Learning & Technology*, *5*(3), 91-105. https://doi.org/10125/44568

Williams, B. (2006). *Philosophy as a humanistic discipline*. Princeton University Press.

Appendix: Raw frequency counts of authorial voice features across disciplines

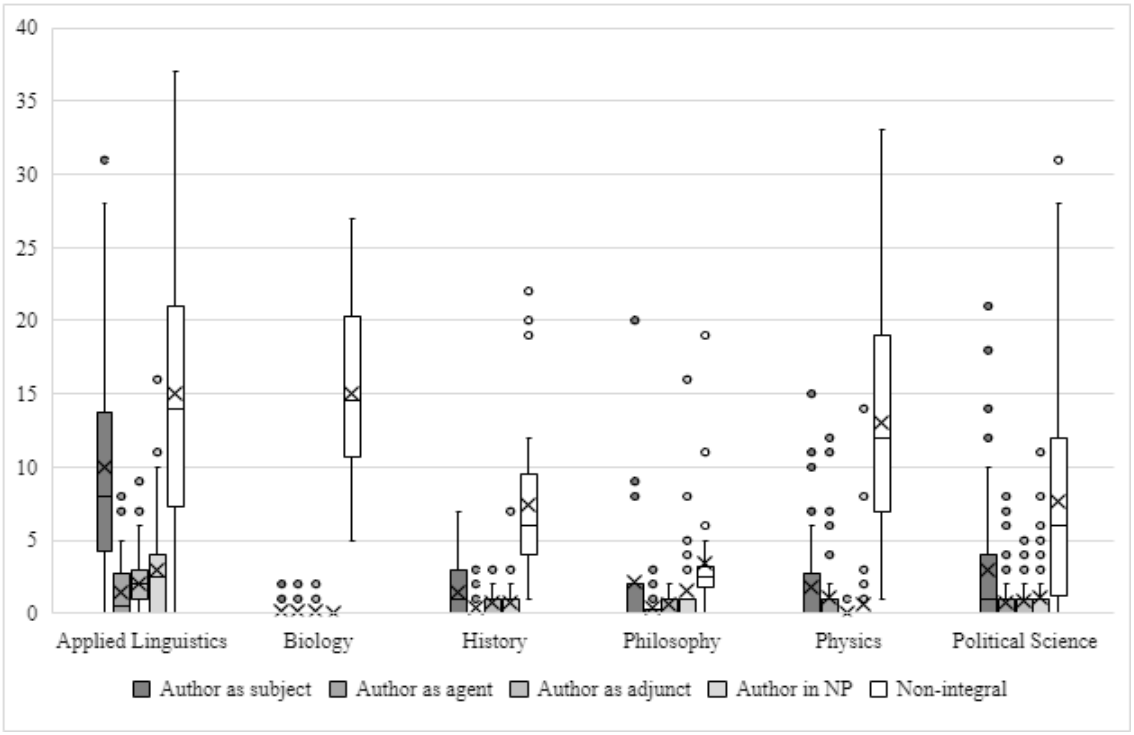

Figure 4. Citation types

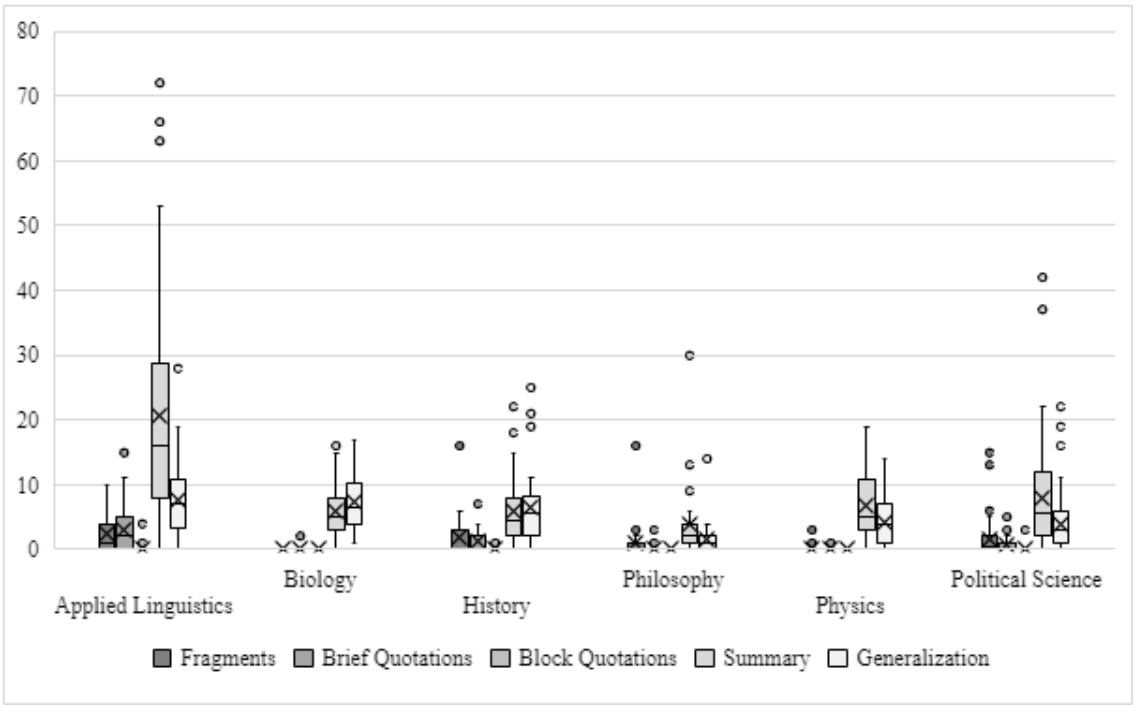

Figure 5. **Citation presentation**