

Original Paper

Comparative Analysis of Evidentiality in Spoken and Written Academic Discourse

Chenyuan Wang¹ & Jingjie Li²

¹ Shanghai, China

² Institute of Corpus Studies and Applications, Shanghai International Studies University, Shanghai, China

Received: March 27, 2024

Accepted: May 03, 2024

Online Published: May 21, 2024

doi:10.22158/eltls.v6n3p62

URL: <http://dx.doi.org/10.22158/eltls.v6n3p62>

Abstract

This paper aims to explore the differences existing in the use of evidentiality in spoken and written academic discourse. Great differences are revealed in the occurrence frequency and linguistic forms of sub-types of evidentiality (shared evidentials, reporting evidentials and personal evidentials). Comparatively, writers demonstrate a proclivity towards employing shared and reporting evidentials and opting for linguistic forms conveying objectivity and discreteness due to the temporal and spatial distance inherent in the author-reader dynamic. Speakers tend to use personal evidentials to project a confident and positive self-image and display less stringent adherence to the verifiability of evidence sources because of the instantaneous nature of (delete) speech and the interactive nature of oral communication with an audience.

Keywords

evidentiality, evidential, spoken academic discourse, Written academic discourse, Linguistic forms

1. Introduction

Since Franz Boas' groundbreaking work in 1911, evidentiality has consistently been a focal point of linguistic research (e.g., Willett, 1988; Aikhenvald, 2007; Murray, 2020; Smirnova, 2021; Rudolph, 2022; Jacquin, 2022; Bermúdez, 2023). Initially categorized solely as a grammatical phenomenon, the research emphasis was on languages that employ morphological coding of evidentiality (e.g., Barnes, 1984; Chafe, 1986; De Haan, 2001; Aikhenvald, 2004; Ahn & Yap, 2015). However, the study of languages with obligatory grammatical means of evidentiality, which only exists in about one-quarter of the world's languages (Tournadre & LaPolla, 2014), can't reveal the comprehensive use of evidentiality. A wide array of lexical expressions is also regarded as evidentials in languages like

English, where obligatory grammatical evidentiality does not exist (Marin-Arrese, 2015). And this lexical strategy extends across different linguistic registers including colloquial discourse (Fox, 2001; Clift, 2006; Estellés, 2015; Cornillie & Gras, 2015), journalistic discourse (Hsieh, 2008; Wang & Pang, 2013; Pérez Blanco, 2020; Xue & Lu, 2022; Carretero, 2023) and courtroom discourse (O'Barr, 1982; Luo & Liao, 2013; Deng, 2017; Greco, 2018). Academic discourse has been a significant area of investigation (Mushin, 2001; Janik, 2009; Yang, 2013; Albelda-Marco, 2018; Liao & Wen, 2022). These studies have demonstrated that evidentials are essential microcosmic discourse elements which reflect specific linguistic characteristics of academic discourse.

Considering the important role of evidentials in academic discourse (Mushin, 2001; Janik, 2009; Yang, 2013), many scholars are focused on the written academic discourse (Ruskan, 2012; Fetzer, 2014; Alonso-Almeida, 2015; Liao & Wen, 2022). Meanwhile, the different use of evidentiality in written and oral discourse genres interests some linguists (e.g., Chafe, 1986; González et al., 2014; Albelda-Marco, 2018). However, little attention has been paid to the spoken academic discourse, though other common oral genres have aroused linguists' interest such as debate (Cornillie & Gras, 2015; Jacquin, 2022), testimony (Luo & Liao, 2013; Greco, 2018), and especially communication (Mushin, 2012; Ahn & Yap, 2015; Estellés, 2015; Cook, 2021). In Chafe's seminal work (1986), evidentiality is regarded as one of the basic ways in which "spoken and written language differ from each other" (p. 261). His insightful comparison highlights how evidentials differ, attributing these differences to "the spontaneity and interactiveness of speaking and the deliberateness and detachment of writing" (p. 272). However, his analysis is confined to the dinnertable conversations and focuses only on the lexical means (Hunston & Francis, 2000). Considering the previous researches, the present study focuses on the different use of evidentiality in spoken and written academic discourse. And we found though the distinction is not as prominent as the comparison between daily conversation and academic writing such as in Chafe's study, the instantaneity and interactiveness of spoken academic discourse and the reliability and objectiveness requirement of written academic discourse have an impact on the choice of evidentials and the use of evidentials in different linguistic forms.

This paper extends previous studies in two ways: (1) it compares the use of evidentiality in spoken and written academic discourse, employing a substantial corpus, and (2) it describes the major linguistic forms of evidentials in spoken and written academic discourse within the evidentiality framework which facilitates a deeper understanding of the dynamic interplay among the three principal entities in academic discourse: the writer/speaker, the addressees, and the referenced third parties. And this paper intends to answer the following questions: (1) Are there divergences existing in the use of evidentiality between spoken and written academic discourse? If there are, what are the differences reflected in the frequency and linguistic forms of evidentials? (2) What possible factors have contributed to these observed differences?

2. Literature Review

Since the popularity of studies of evidentiality, there has been a considerable body of literature on it concerning its different aspects such as definition (e.g., Anderson 1986; Tournadre & LaPolla, 2014), classification (e.g., Willett, 1988; Chafe, 1986; Diewald & Smirnova, 2010; Marín-Arrese, 2015) and functions (e.g., Nuyts, 2001; Cornillie, 2010; Diewald and Smirnova, 2010; Pérez Blanco, 2020). And Mushin's representative work in 2001 reveals how to study evidentiality in a specific discourse, which also highlights the role of evidentiality in discourse research (Yang, 2013). Several factors which impact the use of evidentiality in academic discourse have been studied in recent years.

Discipline is regarded as one significant factor which can affect the use of evidentiality. As Biber (2006) points out, the linguistic distinction across academic registers is "more important than most of us would expect" (p.02). In academic discourse, academics need to be consistent with various rhetorical norms, values and assumptions of different disciplines (Van Dijk, 1997; Hyland, 2004). The study of evidentiality in academic discourse has been "a central line of investigation across disciplinary communities" (Khedri, 2018, p. 146). Some linguists compare the use of evidentials between hard and soft sciences (Yu, 2015; Liu & Xu, 2017; Khedri, 2018), results of which generally agree that evidentials are more frequently shown in soft sciences than in hard sciences. The prominent use of evidentials in soft sciences is regarded as resulting from the increase of interpretative variation (Hyland, 2005) and the detachment of immediately prior or single-line development (Becher, 1989; Hyland, 2004; Khedri, 2018). Since specific disciplines are chosen as research objects in the above studies, the representativeness of the result is somewhat arguable. For example, evidentials are frequently used in biology, which is categorized as hard science (Hyland, 2004; Liu & Xu, 2017). Some linguistics focus on two specific disciplines such as linguistics versus economics (Grossmann & Tutin, 2010), ESST (English of Specialized Science and Technology) versus ECST (English of Common Science and Technology) (Huang, 2019) and agriculture versus economics (Dong & Buckingham, 2020), which reveal the use of evidentials differs between disciplines, even within the same category.

The academic ability of writers is another factor which attracts scholars' attention. The academic discourse is characterized by rigorous language and objective argumentation, which is the standard students should learn to meet. Compared with experienced academics, the deficiency of students' academic writing has been the object of some studies, part of which concerns the differences in the use of evidentials between apprentice and advanced writers (Barton, 1993; Aull & Lancaster, 2014; Chen & Yu, 2017; Lou & Qiu, 2020; Liao & Wen, 2022), while part of which focus on the differences among students of different degrees (Hyland, 2004; Xu, 2015). Results show that the use of evidentials is affected by the academic level of writers including the frequency of evidential use (Hyland, 2004), the choice of evidentials (Chen & Yu, 2017; Liao & Wen, 2022) and the wielding of evidential strategies (Barton, 1993; Aull & Lancaster, 2014).

Except for the two factors introduced above, the effect of language used by writers (Alonso-Almeida, 2015) and the cultural background of writers (Shui & Ji, 2015; Wang, 2016; Yang, Xu, & Liu, 2019)

also arouse linguists' interest. We can see comparative researches concerning the use of evidentials in academic discourse gradually become a research hotspot. However, few linguists compare evidentiality in spoken and written discourse. Chafe (1986) reveals the differentiation of the use of evidentiality in conversation and academic writing. Great differences are regarded as stemming from "the spontaneity and interactiveness of speaking" and "the deliberateness and detachment of writing" (p. 272). Albelda-Marco (2018) studies the use of Spanish evidentials in four discursive genres - colloquial conversations, press news, academic papers, and parliamentary debates, which finds that evidential discourse markers take a higher proportion in oral genres than in written genres. Adel (2010) also notices that comparison of spoken and written academic discourse has not got enough attention but she focuses on the reflexivity in language. This paper aims to compare the use of evidentiality between written and spoken academic discourse and reveal the differences in the use of evidentials.

3. Methodology

3.1 Corpus

The paper is based on the PICAIE (Pearson International Corpus of Academic English), a large-scale corpus that is compiled at Pearson in collaboration with Lexical Computing Ltd. Currently, the large-scale corpus of academic spoken corpus in English is scarce, so the PICAIE is one of the most ideal data sources for the study intending to reveal linguistic differences between academic written and spoken discourse. Compilers of PICAIE adopt a broad definition of academic discourse where it involves not only academic productions but also the materials produced in academic life on campus. PICAIE is divided into two sub-corpus: the written corpus (32,475,526 words) consists of textbooks, article, administrative, magazines and employment and the spoken corpus (4,640,675 words) consists of lectures, seminars, broadcasts and miscellaneous. According to Ackermann et al. (2010), part of the data in spoken corpus are directly downloaded from station websites where transcriptions are available. The spoken academic discourse studied in this paper includes both the transcriptions of on-the-spot speech and the texts prepared in advance for the speech scenarios.

3.2 Identification of Evidentials

Evidentiality is grammaticalized with obligatory markers in about one quarter of languages in the world (Aikhenvald, 2007; Tournadre & LaPolla, 2014; Saratsli et al., 2020), such as hearsay evidential markers -tako, -tamyē, -tamyense, -tanun, and -tanta in Korean (Ahn & Yap, 2015). In English, due to the lack of grammaticalized evidentials, the recognition of lexical evidentials is necessary and inevitable, which, however, is complex and controversial. To determine the lexical evidentials as search object in this study, we use "evidentiality" as keyword to search for relevant papers in China National Knowledge Infrastructure, which is the largest academic paper database in China and collect all the relevant studies listed in the literature. More than 100 relevant papers are found and all the evidential markers studied in them are collected as the search object (e.g., Chafe, 1986; Marín-Arrese, 2015; Musi, 2016; Wang, 2016; Chen & Yu, 2017). Because the evidential value is contextual-functional, all

concordances of these evidential markers appeared in the PICAЕ are collected and re-identified with three of the standards proposed by Anderson (1986), in which evidentials should be (i) the justification for a factual claim which is available to the person making that claim, whether direct evidence plus observation (no inference needed), evidence plus inference; inference (evidence unspecified), and reasoned expectation from logic and other facts; (ii) not the main predication; (iii) indication of evidence as their primary meaning. Since the criteria proposed by Anderson (1986) originally hold for grammatical evidentiality, the last criterion - never occurring as a derivational morpheme nor as an element in a compound- need not be considered when it is applied to lexical evidentiality. Such criteria are adopted in this paper because as the best working definition (Willet, 1988), they can also be applied to lexical evidentiality (Cornillie & Gras, 2015). The consistent concordances of evidential markers are saved. Some evidential markers are deleted because no consistent concordance is found in PICAЕ such as “touch” and “behold” and the remaining evidential markers are classified into the evidentiality framework in Table 1.

3.3 Analytical Framework

The classification framework of evidentiality applied in this paper is a modified version of the one proposed by Chen and Yu (2017), which highlights and integrates the three groups of participants in the academic discourse - the writers/speakers, the reported subjects and addressees - who respectively have been the research object of many studies (e.g., Hyland, 2004; 2005; Lafuente-Millán, 2013; Yang, 2015). Nevertheless, this paper comprehensively concerns all of them and reveal their relationships in the discourse construction. According to Chen and Yu (2017), the knowledge individuals hold can be classified into three categories: shared knowledge, reported knowledge and personal knowledge (p. 27). Evidentiality, which indicates the source of evidence, can be classified into shared evidentials, reporting evidentials and personal evidentials. Shared evidentials refer to the source of information which is shared between writers/speakers and readers/listeners. Shared knowledge includes truth, common sense and shared knowledge in specific background, the first two kinds of which refer to information familiar to both sides of interaction and speakers/writers tend to use zero evidentials to mark (Chen & Yu, 2016), such as “Asia is in the Northern Hemisphere” or “drinking too much is harmful” in which no lexical evidential necessarily appears to indicate the shared background knowledge. However, in certain conditions, the use of shared evidentials is indispensable to point out where the supportive shared information is exhibited (e.g., “as in Experiment 1”, “(as in example of Section 6.3)”), which is the main research object in this paper.

Reporting evidentials indicate the information speakers/writers repeat from others, which are divided into quotation (the reported subject is explicit, as in “according to Casson (1987)”, and “Crocombe (2007:10) summarizes”) and hearsay (the reported subject is unknown, as in “reportedly” and “people say”). And personal evidentials indicate that speakers/writers themselves are the source of information, which is further classified into eight sub-types (sensory evidentials, inference evidentials, opinion evidentials, expectation evidentials, example evidentials, hypothesis evidentials, modal verbs and

performative verbs) in the paper of Chen and Yu (2017). In this paper, we made the following modifications:

- a. Sub-types “example evidentials”, “hypothesis evidentials” and “performative verbs” were deleted because the broad definition of evidentials they took, that is any linguistic forms indicating information source and attitude of writers, is different from what is adopted in this paper (presented in the next section). The examples of evidentials presented in their paper such as “like”, “if” and “use” can’t be identified as evidential markers in this paper.
- b. Sub-types “opinion evidentials” and “expectation evidentials” are integrated into belief evidentials because both the opinion and expectation are reasoned thinking from logic and other facts.
- c. Sub-type “modal verbs” is included in the “inference evidentials” category. Modal verbs, which concern logical reasoning and a speaker’s/writer’s commitment or detachment to propositions (Stubbs, 1996), are normally regarded as the marker of inference (Auwera & Plungian, 1998; Haßler, 2015; Yang, 2015).

The modified framework of evidentiality is presented in Table 1:

Table 1. Classification Framework of Evidentials

Evidential Types		Linguistic Forms	Examples
Shared evidentials		Non-integral	(Figure X), (see the previous ...), (in Figure X)
		Integral	The previous discussion shows that, as stated before, in example X
Reported evidentials	Quoted evidentials	Reporting verb	argue, think, believe, propose
		Prepositional phrase	according to sb., in sb.'s view /opinion
		Noun + that	sb.'s observation\ agreement\ finding\ view\ study + that...
		Parentheses reference	(sb., year), (sb., year, page), (page)
Personal evidentials	Hearsay evidentials	Adverbs	reportedly, allegedly, supposedly
		X + clause	I've heard, people say, saying goes
	Sensory evidentials	Visual verb	(We/ I) see, observe, notice
		Other sensory verbs	(We/I) hear, feel, (sth.) sound like
		Belief evidentials	Belief verbs
Prepositional phrase	in my/our opinion, in my/our view		
	Inference	Modal verb	must, can, could, may, might, will

evidentials	Modal adverb	possibly, certainly, perhaps, maybe, seemingly
	It clause	it is evident that, it's obvious that · it's apparent that

4. Results and Discussion

With the methodology delineated above, 72,884 consistent concordances are identified in spoken corpus and 409,464 consistent concordances in written corpus. Further quantitative results are presented in Table 2, which were analyzed with Python and the results of the Chi-square test are also presented. The P value of the total is 0.083 (> 0.05), which normally means no significant difference exists. However, we contend that noticeable divergences exist between the use of evidentiality in spoken and written academic discourse, which can be elucidated from the following two aspects. On the one hand, though the P value exceeds 0.05, the margin is narrow, suggesting a potential trend towards significance. On the other hand, only two subtypes, namely sensory and inference evidentials, surpass 0.05. Since inference evidentials account for a large proportion, the use of them has an evident impact on the overall data. By contrast, the P values of the other subtypes fall below 0.05, which means significant differences exist. So, it is important to scrutinize and discuss the data findings of this study, as they warrant careful consideration. Detailed differences existing between the use of evidentiality in spoken and written academic discourse will be explicitly revealed in the subsequent sections.

Table 2. Evidentials in Spoken and Written Academic Discourse

	Spoken discourse		Written discourse		P value
	RN	Percent	RN	Percent	
Evidential types	NF		NF		
Shared evidentials	116	0.16%	9,787	2.39%	0.000
	0.02		0.30		
Reporting evidentials	9,230	12.66%	98,706	24.11%	0.044
	1.99		3.04		
Quoted evidentials	8,906	96.49%	98,094	99.38%	0.045
	1.92		3.02		
Hearsay evidentials	324	3.51%	612	0.62%	0.022
	0.07		0.02		
Personal evidentials	63,538	87.18%	300,971	73.50%	0.089
	13.69		9.26		
Sensory evidentials	518	0.82%	1,306	0.44%	0.258
	0.11		0.04		
Belief evidentials	10,325	16.25%	3,082	1.02%	0.022

	2.22		0.09		
Inference evidentials	52,695	82.93%	296,583	98.54%	0.090
	11.36		9.13		
TOTAL	72,884	100%	409,464	100%	0.083
	15.70		12.60		

4.1 The Use of Shared Evidentials

The P value listed in Table 2 corroborates significant distinction exists in the use of shared evidentials between spoken and written academic discourse. Moreover, as Table 2 shows, the divergence of the use of shared evidentials is most remarkable among all the sub-types, with the normalized frequency and proportion of shared evidentials being higher in the written corpus compared to the spoken corpus. This difference can primarily be attributed to the data source. The spoken corpus comprises broadcasts, where speakers can't share information visually presented in pictures, charts, and graphs with listeners. In contrast, the written corpus consists of articles, textbooks and extra-curriculum materials where writers can easily exhibit shared information for readers. Furthermore, the necessity of using shared evidentials varies distinctly between spoken and written discourse. In the dynamic settings of lectures or seminar, the presentation of shared information typically occurs in tandem with the speaker's utterance. Conversely, writers need clearly point out specific diagram or content where the presented information supports their claims with shared evidentials.

(1) Codex Artaud (Figure 21) both summarizes the major concerns of the first twenty years of Spero's career (from the late 1940s to the late 1960s). (shared evidential, written corpus)

As in example (1), in the written discourse, writers need to use shared evidential "(Figure 21)" to point out where the shared background information is exhibited. While in the spoken discourse speakers just need to put up the relevant information, listeners will locate what the speaker is talking about by themselves.

Lastly, the transience of spoken discourse also hinders the use of shared evidentials, as in the example below:

(2) As we saw in the previous section, there is some historical evidence to support a suggestion of this kind. (shared evidential, written corpus)

When speakers proceed with their speech, what listeners can remember is restricted, so speakers must be cautious about resorting to previously presented content. However, such an obstacle doesn't exist in the written academic discourse, where readers can review all the presented content.

Except for the occurrence frequency, we further explore the existing differences in linguistic forms of shared evidentials in spoken and written corpus. The linguistic forms of shared evidentials and the proportions they respectively take are presented in Figure 2.

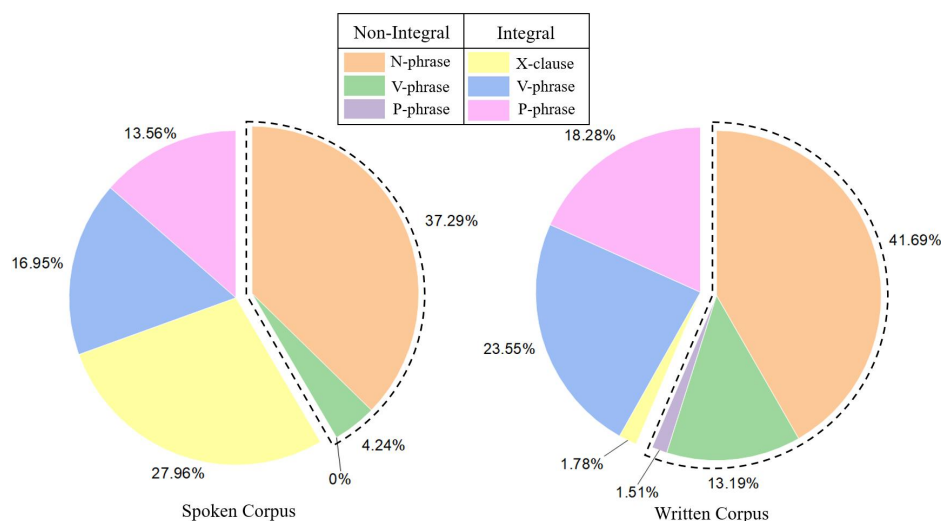


Figure 1. Proportions of Linguistic Forms of Shared Evidentials

Referring back to the classification of citation proposed by Swales (1990) and Charles (2006), the linguistic forms of shared evidentials can be categorized as “non-integral” (the evidence appears within brackets or other punctuation which explicitly separate the evidence and the main sentence structure) and “integral” (the evidence appears within a sentence structure and play a grammatical role).

The “non-integral” and the “integral” are further classified as presented in Figure 1. In the “non-integral” category, “N-phrase” is predominantly used. With the use of punctuation, which explicitly excludes the evidence source from the sentence’s main content, noun phrases can succinctly present necessary information that points out the location of shared information supporting the preceding statement, as in example (3).

(3) Asian women, in comparison with White and Mexican-born women, were the lowest consumers of tobacco (1.7 %; Figure 9.1). (shared evidential, written corpus)

(4) The reason lies in the fact that machines have a memory hierarchy and that the faster memory is at the top of that hierarchy (see Figure 2) (shared evidential, spoken corpus).

The “V-phrase” form has the second highest occurrence. As in example (4), the sensory act verb “see” is used in front of evidence source in active voice, where the speaker seems to actively invite readers to examine his/her proposed content. Not only such use can attract the addressees’ attention but also reflects the writer’s/speaker’s confidence of the presented proposition. The data results show only the sensory act verb “see” is used in this linguistic form in the spoken corpus. However, in the written corpus, some discourse and research act verbs also appear in this form, as the following examples show:

(5) In the little village of Hemnesberget (described in example 6 in the first chapter) Bokml or standard Norwegian is the variety to use when you go to the tax office to sort out your tax. (shared evidential, written corpus)

(6) These interactions may be supplemented by hydrogen bond formation (shown in Figure 29.26a)

between the OH group of Tyr-248. (shared evidential, written corpus)

As shown in the two examples above, discourse and research act verbs in this form are normally used in passive form. Different from the use of the sensory act verb, writers tend to show their research activity in the passive voice when they present the shared evidence source.

The occurrence frequency of “P-phrase” is the lowest, which may result from the redundancy of the use of prepositions. We can see in example (7), the parenthesis has separated the shared evidential from the main content and the existence of “in” doesn’t provide any extra and important information. So, academics are inclined to omit the use of prepositions.

(7) ...even if it is not correct (in experiment 2), if the animal has a high fitness, i. e., has many other adapted genes. (shared evidential, written corpus)

All three linguistic forms in the “non-integral” category are frequently used in the written discourse. It may originate from the fact that speakers do not frequently use shared evidentials in non-integral forms. With the non-integral forms, the evidence source is presented in the peripheral place and separated from the main content of a sentence, which enables addressees to quickly locate the information they want. However, it is more difficult and unnecessary in the spoken discourse because readers can not perceive the existence of punctuation.

And in the “integral” category, “V-phrase” and “P-phrase” also have a higher proportion in the written corpus, while “X-clause”, as the most frequently-used linguistic form in the spoken corpus, has a much higher proportion than that in the written corpus. Moreover, in this linguistic form, clauses only follow “as” in the spoken discourse. The subjects of perception (the speaker and the listeners) are sometimes emphasized to show the validity of the presented claim. In contrast, in the written discourse, the objective information provider normally precedes propositions.

(8) As we can see from Table 4, the Level 2 BLAS offer the potential for two floating-point operations for every reference. (shared evidential, spoken corpus)

(9) But the previous chapter showed that there may Ch 12 be many important ‘soft’ aspects to a problem situation. (shared evidential, written corpus)

In example (8), with the as-clause, the speaker shows the proposed content as the perceived result of both himself and the listeners, implying high credibility. And in example (9), the writer chooses to conceal the role of himself. We find that writers are much more inclined to avoid self-prominence in this linguistic form of shared evidentials.

4.2 The Use of Reporting Evidentials

In academic discourse, reporting evidentials have been found to have a high occurrence frequency (e.g., Hyland, 2004; Yang, 2009; Albelda-Marco, 2018; Lou & Qiu, 2020). As Xu (2012) proposes, literature citation is one of the most prominent characteristics of academic discourse, serving to anchor arguments in a broader scholarly context and to demonstrate rigor through engagement with existing research. When it comes to the comparison between spoken and written academic discourse, the P values in Table 2 indicate that the use of reporting evidentials in the spoken academic discourse is

distinct from the written academic discourse. Though reporting evidentials take the second place both in spoken and written corpus respectively, they have higher occurrence frequency in the written discourse (3.04 words per thousand) than in the spoken discourse (1.99 words per thousand), and the proportion they take in the written corpus is more than that in the spoken corpus. Reporting evidentials are further divided into quoted evidentials and hearsay evidentials, P values of which both exhibit striking differences existing in the use of them between spoken and written discourse. The proportions taken by quoted evidentials are more than 90% in both corpora, while the NF of them in the written corpus is about twice that in the spoken corpus. The quotation of relative literature is generally considered one of the most significant characteristics of academic discourse. However, speakers less frequently resort to the third-party, which may result from the temporality and instantaneity of spoken discourse which does not usually allow for extensive verification of source. And hearsay evidentials have low occurrence frequency in academic discourse, while the proportion of them in the written corpus (0.62%) is much lower than that in the spoken corpus (3.51%), which may be linked to the conversational nature of spoken discourse where speakers might reference commonly known or broadly accepted views. However, writers are more cautious about the existing risk posed by reporting information from unknown subjects.

To reveal deeper differences existing in the use, the linguistic forms of reporting evidentials are presented below, the proportions of which are calculated and listed.

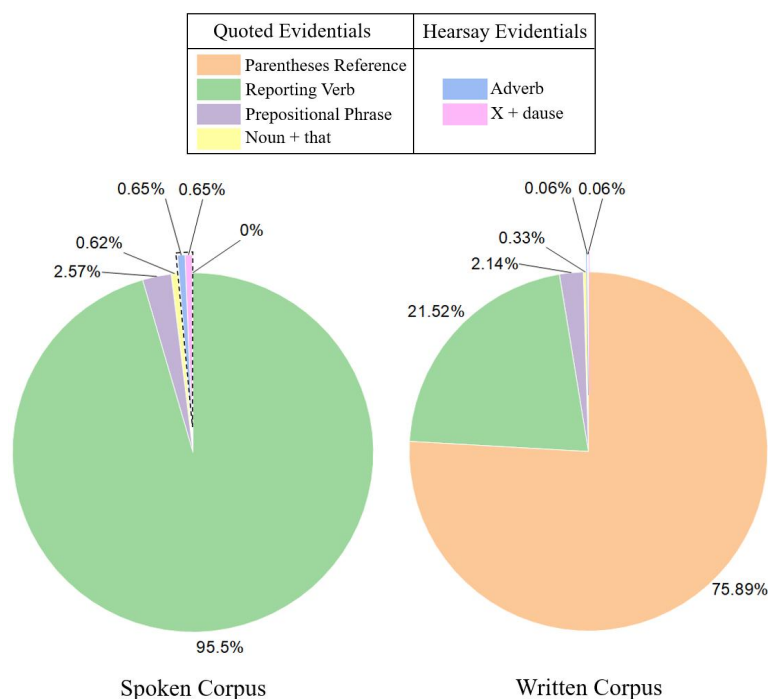


Figure 2. Proportions of Linguistic Forms of Reporting Evidentials

Figure 2 shows strong differences exist in the use of different linguistic forms. “Parenteses reference”

takes the highest proportion in the written corpus but has no concordance in the spoken corpus. Here are two examples:

(10) Apatites give particularly bright CL, with europium a common activator (Mariano & Ring, 1975). (reporting evidential, written corpus)

(11) Similarly among the Polynesian Maori of New Zealand, " to do the right thing is to follow the ancestors " (Johansen, 172). (reporting evidential, written corpus)

As shown above, writers present detailed information about the source of the cited so that readers can testify, which is necessary to exhibit the reliability of reported content. Meanwhile, reporting supportive content from professional subjects can enhance acceptability. So, with this linguistic form, we can see that writers normally cited from published academic papers. With the reporting of many relevant and authoritative works, writers can enhance the acceptability of proposed content and also show the breadth and depth of their understanding of the research field. On the other hand, the popularity of this linguistic form in the written corpus can stem from the fact that it is always put in the periphery parenthesis and appended at the end of a proposition. As Hyland (2017) thinks, writers can introduce information which supports their ideas without foregrounding the originators with this form. It leaves readers' focus on the core content that writers intend to discuss.

However, the instantaneity of spoken discourse hinders speakers from providing source information in the form "Parentheses reference", for listeners have no time to remember and check the detailed source information. Instead, reporting verbs have a dominant proportion in the spoken corpus. In this linguistic form, the originator of reported materials is foregrounded to take the major responsibility. Due to the lengthy process of spoken academic discourse, addressees can't keep their attention focused on what they hear. Considering a reliable originator can improve the acceptability of presented content, the reported originator needs to be foregrounded to attract listeners' attention. Meanwhile, despite the frequent use of parentheses reference, reporting verbs take a relatively high proportion in the written corpus, which is more than twenty percent and is ranked the second frequently-used linguistic form. Except for the occurrence frequency, to further reveal the differences, the reporting verbs are categorized into discourse act verb, research act verb and cognition act verb, according to the classification proposed by Hyland (2002). Moreover, the top three most-frequently-used discourse verbs are presented in Table 3.

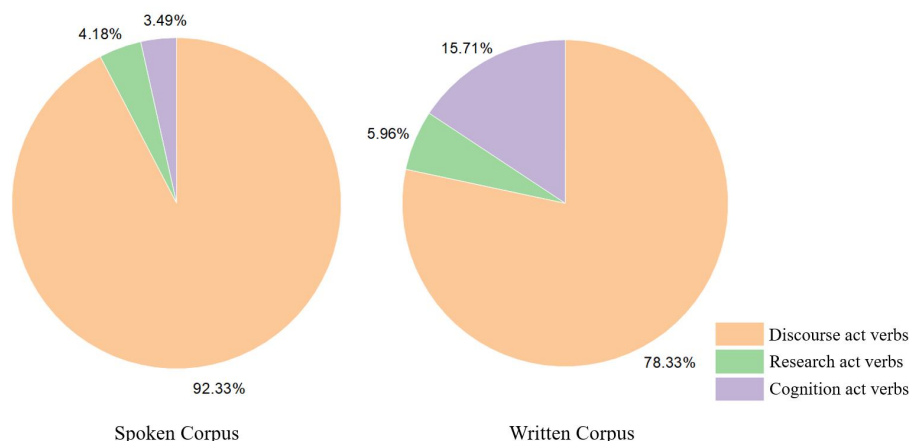


Figure 3. Proportions of Three Types of Reporting Verbs

Table 3. Top 3 Most-frequently-used Discourse Act Verbs

Spoken discourse		Written discourse	
Discourse act verbs	Percent	Discourse act verbs	Percent
say	77.96%	say	43.78%
tell	8.40%	argue	11.84%
claim	1.82%	suggest	5.45%

As Figure 3 shows, discourse act verbs have the prominently-frequent occurrence both in spoken and written discourse, but the proportion of them in the spoken corpus is more than ninety percent. As Hyland (2002) defined, discourse act verbs “focus on the verbal expression of cognitive or research activities” (06). We can see that “say” and “tell”, which take more than eighty percent in the spoken corpus, objectively pass the information without speakers’ interpretation. And the word “claim”, which is believed to attribute author’s high degree of confidence, only takes less than two percent. In the written discourse, though discourse act verbs also take about eighty percent, the frequently-used verbs show more divergence from those in the spoken discourse. The verb “say” only takes about forty percent, while the verb “argue”, which signals a supportive attitude toward reported content, takes more than ten percent. And the verb “suggest” that implies a tentative view of writers takes about five percent. The results show that in the choice of different discourse act verbs, speakers intensively use certain verbs to objectively report information from the third- party but writers are more inclined to imply different attitudes with reporting verbs.

The proportions of research act verbs are low both in spoken and written corpus and only two verbs “find” and “show” appear, which are normally used to exhibit the result or finding of research. Research act verbs describe “experimental activities or actions carried out in the real world” (Hyland, 2002: 06). What speakers and writers intend to exhibit obviously is their research activities and only the

results of others' experiments worth reporting, so such a phenomenon in this paper is understandable. The proportion of cognition act verbs in the written corpus is obviously higher than that in the spoken corpus. Cognition act verbs can reflect the attitudes of the cited subject which speakers or writers attribute to. Verbs with a positive attitude like "think", "insist" and "acknowledge" indicate that cited subjects take high responsibility for reported content:

(12) It's the American taxpayer who had nothing to do with these private contracts, and yet Democrats insist that the innocent party, the American taxpayer, pay the burden for this problem. (reporting evidential, spoken corpus)

(13) He thinks that this helps to explain the "resurgence of the arts" in evangelical churches (Webber, 2002, p. 51). (reporting evidential, written corpus)

In contrast, verbs with a tentative attitude like "believe" and "assume" reflect that the cited subjects also take a prudent stance when they express their opinions.

(14) Apo assumed that he was being framed by Viorel and Viorel had called the police. (reporting evidential, spoken corpus)

(15) Poulson (1997) believes that achieving a fair market in Colorado will not be immediate and will take some time. There are others who believe that it is working in some cases and not working in others. (reporting evidential, written corpus)

Table 4. Proportions of Cognition Act Verbs with Different Attitudes

	Spoken discourse (P)	Written discourse (P)
Positive attitude	50.83%	35.89%
Tentative attitude	49.17%	63.84%
Neutral attitude	0	0.27%
Total	100%	100%

Note. P represents the percent that the verbs with different attitudes take in all the cognition act verbs.

As Table 4 shows, speakers almost equally use cognition act verbs with a positive attitude and a tentative attitude, while writers more frequently use cognition act verbs with tentative attitude. When speakers and writers report claims from the third-party, they need to evaluate the cited subject's attitude and certainty. Writers normally present more caution in their expressing of viewpoints.

The other linguistic forms, "prepositional phrase", "noun + that", "adverb" and "X + clause", all take small proportions both in spoken and written corpus. However, the linguistic form "adverb" in the spoken corpus accounts for about two percent more than that in the written corpus. The hearsay evidentials in this linguistic form succinctly show the obscurity of information sources.

(16) They were reportedly paid more than \$4 million for pictures of their first child, Shiloh. (hearsay evidential, spoken corpus)








(17) Book of Records or exhortations to others to continue circulating a particular e-mail because some big company has supposedly promised that every forward will generate monies for the care of a languishing tot. (hearsay evidential, written corpus)

Both speakers and writers can't avoid the plight of uncertain information sources, so they need to use hearsay evidentials as in the two examples above. However, due to the unclear information source, addressees can doubt the reliability of reported information. Listeners' attention may be quickly diverted to following content, while readers have much chance to notice the lack of supporting third-party, which will impose an adverse influence on academics' discreet image.

4.3 The Use of Personal Evidentials

Table 1 shows that personal evidentials take the highest proportion both in spoken and written discourse, which is much more than shared and reporting evidentials. Although readers and the third-party are necessary subjects, the speaker/writer is the core discourse constructor. So, with the frequent use of personal evidentials, speakers/writers exhibit their arguments and create their own identity.

The P value in Table 1 shows no significant distinction exists in the use of personal evidentials in spoken and written discourse. Personal evidentials have three sub-categories, in which P values of sensory evidentials, and inference evidentials also exceed 0.05. The data results show that writers and speakers have more in common in the use of personal evidentials. Nevertheless, differences existed in the use of personal evidentials in spoken and written corpus can also be found. We can see the proportion taken by personal evidentials in the spoken corpus (87.18%) is apparently higher than that in the written corpus (73.50%). To further reveal the characteristics of the use of personal evidentials, their linguistic forms and proportions these linguistic forms take are presented in Figure 4, in which distinctions are also exhibited.

Sensory	Belief	Inference
 Visual Verb  Other Sensory Verb	 Belief Verb  Prepositional Phrase	 Modal Verb  Modal Adverb  It Clause

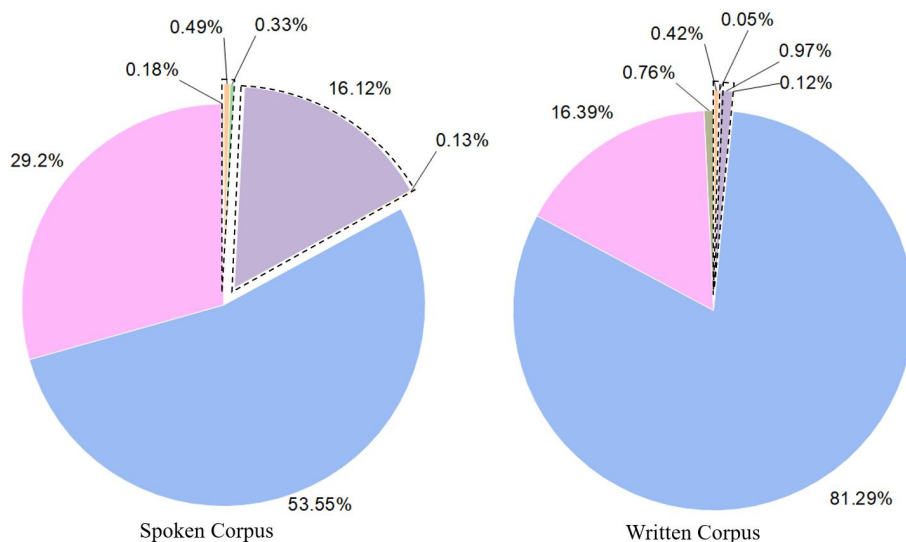


Figure 4. Proportions of Linguistic Forms of Personal Evidentials

The P value of sensory evidentials indicates that no significant difference exists between spoken and written discourse. Sensory evidentials take less than one percent both in the two corpora, as Table 1 shows. Sensory evidentials are believed to be reliable (e.g., Chafe, 1986; Papafragou et al., 2007; Marín-Arrese, 2013) because they reflect the sensory experience of the objective world, but speakers/writers will take full responsibility when they present their sensory experiences, which are often unverifiable. So, both writers and speakers are cautious about the use of them in academic discourse. Moreover, as Figure 4 shows, visual verbs are more frequently used than other sensory verbs both in spoken and written discourse. Vision is the main information source of human. And sometimes what speakers/writers visually perceive can also be presented in front of addressees or accessible to them, so the proposition can be reliable and acceptable, as the examples below show:

(18) When we look at the electron and ion densities inside the void, shown in figure 6, we see the effect of the dust on the plasma parameters. (sensory evidential, written corpus)

(19) We see polls that show 25 percent of Hillary Clinton's supporters are saying they're going to support John McCain. (sensory evidential, spoken corpus)

Comparatively, the information acquired by other senses is hard to share and prove. Here are two examples:

(20) I hear the McCain campaign is talking out of two sides of its convention. (sensory evidential, spoken corpus)

(21) It sounds like the police are taking this very seriously, because of these other high-profile incidents. (sensory evidential, spoken corpus)

When using the sensory evidentials as presented above, the credibility entirely depending on the speaker easily court doubt and challenge. So, though the use of sensory evidentials is similar in spoken

and written discourse, results show that writers are more inclined to take a prudent attitude and avoid using other sensory verbs.

The chi-square test shows the use of belief evidentials in spoken and written discourse has noteworthy differences. The proportion of belief evidentials in the spoken corpus (16.92%) is much higher than that in the written corpus (1.09%). Furthermore, the divergence is reflected in the use of belief verbs, the occurrence frequency of which is much higher in the spoken corpus. With belief verbs following the first personal pronoun, speakers/writers can directly present their views, meanwhile, exposed to the danger of being challenged. Speakers tend to explicitly express their stance than writers. In the spoken discourse, the clear expression of opinion can help listeners catch the core argument of speakers and listeners have a chance to propose their doubt or challenge on the spot and speakers can give further explanation, which will promote the interaction instead of diminishing the reliability of speakers. However, writers tend to be more discreet due to the temporal and spatial distance between them and readers. Based on Hyland (2002), belief verbs are further classified as categories with positive, neutral, tentative and negative attitudes and the proportions of each are presented below:

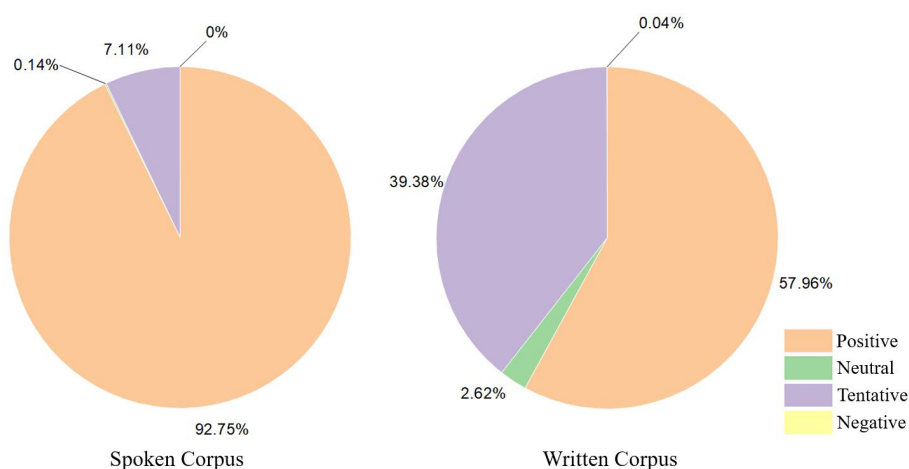


Figure 5. Proportions of Belief Verbs with Different Attitudes

As in Figure 5, belief verbs with neutral and negative attitudes are seldom used. When putting forward their own opinions, both speakers and writers avoid the neutral stance and negative attitude. Those with a positive attitude have the highest occurrence frequency, the proportion of which, however, is much higher in the spoken corpus than in the written corpus. Writers also frequently use belief verbs with a tentative attitude, which only take about 7 percent in the spoken corpus. When presenting one's own opinion, the positive attitude implied can exhibit confidence of the presenter, while the tentative attitude can leave more space for negotiation, as in the two examples below:

(22) We bring urgent pressure to bear on him, we insist that something must occur to him - and we turn out to be right. (belief evidential, spoken corpus)

(23) Kyd pioneers, on the English stage, this absurd version of homeopathic medicine-death curing

death-which I believe lies near the heart of the drama of blood-revenge. (belief evidential, written corpus)

The result shows that speakers tend to present their views with a positive attitude and writers often propose what they think with reservation. For speakers, using belief verbs with a positive attitude highlights their assurance of what they say and can increase their appeal to the audience, while for writers, the limited space of negotiation can evoke readers' suspicion.

As presented above, the use of inference evidentials has no significant difference in spoken and written academic discourse. We can see that inference evidentials take a prominently high percentage of personal evidentials both in spoken corpus and written corpus. Different from the definite and objective world, human's recognition is limited and subjective. Inference evidentials, which reflect subjects' inference of the possibility of what they present, are essential for speakers/writers to give adequate and convincing interpretation or discussion in academic discourse. It won't greatly change with the alternation of discourse. However, some detailed distinctions can be found. Modal verbs, modal adverbs and it-clause are lexical means representing modality in English. Based on previous studies (e.g., Halliday, 1994; Xue & Lu, 2021), the value of these modal expressions can be divided into high, median and low.

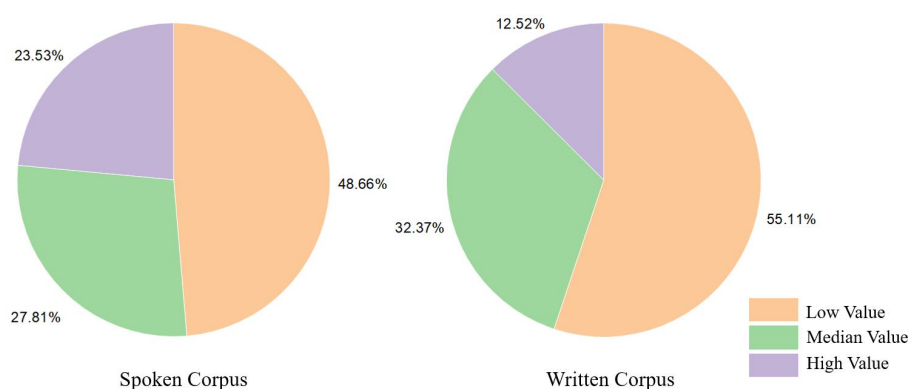


Figure 6. Modal Analysis of Inference Evidentials

Figure 6 shows that the use of inference evidentials is inversely proportional to the modal value both in spoken and written corpus. However, the proportions of low-value and median-value inference evidentials are higher in the written corpus. As Yang (2015) proposes, expressions with doubt and hesitation help to ensure the acceptability of readers (p. 23). The use of low-value inference evidentials shows writers/speakers' reservation, and meanwhile reduces their responsibility if what they present is challenged or contradicted, as in an example below:

(24) Apparently there was no ill effect, and some big marine mammals even seemed attracted to the sound. Perhaps they were curious about what the small oceanographer animals were up to. (inference evidential, written corpus)

In example (24), we can see that the attitude of the writer is converted from certainty (while he describes the objective result) to reservation (while he presents his conjecture). Because he can't prove

his inference of the marine mammals' thoughts, the expression with low certainty is reasonable and can increase readers' acceptability.

In contrast, the use of high-value inference evidentials can highlight writers/speakers' confidence and emphasis on what they present, as in:

(25) Well I mean, it's certainly a contrast, and I've had my own expeditionary comparisons as well..
(inference evidential, spoken corpus)

(26) Those are just benchmarks. Similarly, CO₂, which has already grown considerably, will indeed double and later may well triple its preindustrial concentration. (inference evidential, written corpus)

Figure 4 shows, modal verbs take the highest proportion. As an important means of reflecting modality, the use of modal verbs is common in academic discourse, which, for example, has been found to take 7.6% of the use of all verbs in the study of Francis and Kučera (1982) based on Brown Family Corpora. Comparatively, writers are more inclined to use modal verbs than speakers. Though modal adverbs take the second high percentage both in spoken and written corpus, the proportion of them in the spoken corpus is about twice of that in the written corpus. And the proportion of it-clause is low both in spoken and written corpus, which the need for language conciseness in academic discourse can account for.

5. Conclusion

As analyzed above, great divergence is revealed between the use of evidentiality in spoken and written academic discourse, which reflects both in the frequency and in the linguistic forms of evidentials. Moreover, results of this paper are consistent with the fact that discourse genre conditions the use of evidentiality. To create a persuasive and reliable discourse, both writers and speakers need to carefully keep the balance among addressees, the third-party and themselves. The differences in the use of frequency and linguistic forms of evidentials originate from the different presentation of text. The stationariness of written discourse results in writers' preferences of shared evidentials which refer to objectively supportive information exposed to addressees and reporting evidentials in linguistic form indicating the authoritative and reliable third-party. And the immediacy of spoken discourse requires speakers to make impressive self-identity, so the evidentials with positive attitude or high certainty have higher frequency. Moreover, the temporal and spatial distance between addressers and addressees contributes to the existing divergence. Different space-time location between writers and readers makes writers intentionally create the presupposed space of negotiation and rely on support from credible subjects to reduce readers' vigilance for the potential fallacy. On the contrary, the face-to-face interaction in the seminars or lectures enables speakers to explicitly express themselves.

Findings in this paper help us have a deep understanding of different uses of evidentiality in the spoken and written academic discourse and have pedagogical meaning for freshmen in academia. As the main participants in the campus life, students are exposed to a large amount of spoken and written academic discourse. It is important for them to have a deep understanding and good mastering of the use of evidentiality. However, two limitations of this study need to be acknowledged. In the seminar or lecture,

speakers sometimes use written materials so that visual material's availability may influence the use of evidentiality. Because we have no access to the original video source of the spoken corpus, we can't eliminate its impact. The effect of visual materials in the speech may be the object of later research. And with the methodology in this paper, we can't guarantee all evidential markers have been identified in the examined corpus. Given that we have collected most of the pre-studied evidential markers, the overlooked evidential markers, frequencies of which are too low to attract scholars' attention, could not make any important effect on our research results. Moreover, it is inevitable if we intend to reveal the characteristics of evidentiality use in the large-scale database.

Acknowledgements

The permission to use the PICAЕ from Veronica Benigno is gratefully acknowledged.

References

- Ackermann, K., de Jong, J., Kilgarriff, A., & David, T. (2011, July 20-22). *The Pearson international corpus of academic English*. Proceedings of the Corpus Linguistics Conference, University of Birmingham. Retrieved from <https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2011/Paper-47.pdf>
- Ädel, A. (2001). Just to give you kind of a map of where we are going: A taxonomy of metadiscourse in spoken and written. *Nordic Journal of English Studies*, 9(2), 69-97. <https://doi.org/10.35360/njes.218>
- Ahn, M., & Yap, F. H. (2015). Evidentiality in interaction: A pragmatic analysis of Korean hearsay evidential markers. *Studies in Language*, 39(1), 46-84. <https://doi.org/10.1075/sl.39.1.03ahn>
- Aikhenvald, A. (2004). *Evidentiality*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780199263882.001.0001>
- Aikhenvald, A. (2007). Information source and evidentiality: What can we conclude? *Rivista di Linguistica*, 19(1), 209-227.
- Albelda-Marco, M. (2018). Evidentials as a mark of genre: A study of four oral and written genres. *Pragmatics and Society*, 9(3), 429-453. <https://doi.org/10.1075/ps.16035.alb>
- Alonso-Almeida, F. (2015). On the mitigating function of modality and evidentiality. Evidence from English and Spanish medical research papers. *Intercultural Pragmatics*, 12(1), 33-57. <https://doi.org/10.1515/ip-2015-0002>
- Anderson, L. B. (1986). Evidentials, paths of change, and mental maps: Typologically regular asymmetries. In Chafe, W., & Nichols J. (Eds.), *Evidentiality: The Linguistic Coding of Epistemology* (pp. 273-312). New Jersey: Ablex Pub. Corp.

- Aull, L. L., & Lancaster, Z. (2014). Linguistic markers of stance in early and advanced academic writing: A corpus-based comparison. *Written Communication*, 31(2), 151-183. <https://doi.org/10.1177/0741088314527055>
- Barnes, J. (1984). Evidentials in the Tuyuca verb. *International Journal of American Linguistics*, 50(3), 255-271. <https://doi.org/10.1086/465835>
- Barton, E. L. (1993). Evidentials, argumentation, and epistemological stance. *College English*, 55(7), 745-769. <https://doi.org/10.58680/ce19939273>
- Becher, A. E. (1989). *Academic Tribes and Territories: Intellectual Inquiry and the Cultures of Disciplines*. SRHE/OUP.
- Bermúdez, F. (2023). Using prosody to express evidentiality: The case of the quotative. *Journal of Pragmatics*, 214, 127-143. <https://doi.org/10.1016/j.pragma.2023>
- Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, 5, 97-116. <https://doi.org/10.1016/j.jeap.2006.05.001>
- Boas, F. (1947). Kwakiutl grammar, with a glossary of the suffixes. *Transactions of the American Philosophical Society*, 37, 201-377. <https://doi.org/10.2307/1005538>
- Carretero, M. (2023). Knowledge and belief in the times of COVID-19: A comparative analysis of epistemicity in English newspaper discourse of two stages of the pandemic. *Applied Corpus Linguistics*, 3(2), 100054-100054. <https://doi.org/10.1016/j.acorp.2023.100054>
- Chafe, W. (1986). Evidentiality in English conversation and academic writing. In W. Chafe, & J. Nichols (Eds.), *Evidentiality: The Linguistic Coding of Epistemology* (pp. 261-272). Ablex Publishing.
- 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, 310-331. <https://doi.org/10.1016/j.esp.2005.05.003>
- Chen, Z., & Yu, D. (2016). Evidential Strategies in Argumentative Discourses. *Shandong Foreign Language Teaching*, 37(5), 24-32.
- Chen, Z., & Yu, D. (2017). A credibility-based study of evidentiality in English argumentative discourse. *Modern Foreign Languages*, 40(06), 766-777+872-873.
- Clift, R. (2006). Indexing stance: Reported speech as an interactional evidential. *Journal of Sociolinguistics*, 10(5), 569-595. <https://doi.org/10.1111/j.1467-9841.2006.00296.x>
- Cook, A. (2021). Reconsidering the shi...(de) construction in spoken Mandarin. *Chinese Language and Discourse*, 14(2), 209-231. <https://doi.org/10.1075/cld.18003.coo>
- Cornillie, B. (2010). An interactional approach to epistemic and evidential adverbs in Spanish conversation. In G. Diewald, & E. Smirnova (Eds.), *Linguistic Realization of Evidentiality in European Languages*. De Gruyter. <https://doi.org/10.1515/9783110223972.309>

- Cornillie, B., & Gras, P. (2015). On the interactional dimension of evidentials: The case of the Spanish evidential discourse markers. *Discourse Studies*, 17(2), 141-161. <https://doi.org/10.1177/1461445614564518>
- De Haan, F. (2001). The cognitive basis of visual evidentials. In A. Cienki, B. J. Luka, & M. B. Smith (Eds.), *Conceptual and Discourse Factors in Linguistic Structure* (pp. 91-106). CSLI Publications.
- Deng, Y. (2017). Communicative purpose and power relations in the choice of language, evidence, and strategy in court speech. *Journal of Hubei University (Philosophy and Social Science)*, 44(05), 151-157.
- Diewald, G., & Smirnova, E. (2010). *Evidentiality in German: Linguistic realization and regularities in grammaticalization*. Mouton de Gruyter. <https://doi.org/10.1515/9783110241037>
- Dong, J., & Buckingham, L. (2020). Stance phraseology in academic discourse: Cross-disciplinary variation in authors' presence. *Ibérica*, 39, 191-214. <https://revistaiberica.org/index.php/iberica>
- Estellés-Arguedas, M. (2015). Expressing evidentiality through prosody? Prosodic voicing in reported speech in Spanish colloquial conversations. *Journal of Pragmatics*, 85, 138-154. <https://doi.org/10.1016/j.pragma.2015.04.012>
- Fetzer, A. (2014). Foregrounding evidentiality in (English) academic discourse: Patterned co-occurrences of the sensory perception verbs seem and appear. *Intercultural Pragmatics*, 11(3), 333-355. <https://doi.org/10.1515/ip-2014-0016>
- Fox, B. A. (2001). Evidentiality: Authority, responsibility, and entitlement in English conversation. *Journal of Linguistic Anthropology*, 11, 1-29. <https://doi.org/10.1525/jlin.2001.11.2.167>
- Francis, W. N., & Kučera, H. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Houghton Mifflin.
- González, M., Roseano, P., Borràs-Comes, J., & Prieto, P. (2014). Epistemic and evidential marking in discourse: effects of register and debatability. *Lingua*, 186-187, 68-87. <https://doi.org/10.1016/j.lingua.2014.11.008>
- Greco, P. (2018). Evidentiality and epistemic modality in witness testimony in the context of Italian criminal trials. *Journal of Pragmatics*, 128, 128-136. <https://doi.org/10.1016/j.pragma.2017.10.005>
- Grossmann, F., & Tutin, A. (2010). Evidential markers in French scientific writing: the case of the French verb voir. In G. Diewald, & E. Smirnova (Eds.), *Linguistic Realization of Evidentiality in European Languages* (pp. 279-308). Mouton de Gruyter. <https://doi.org/10.1515/9783110223972.279>
- Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*. Edward Arnold.
- Hsieh, C.-L. (2008). Evidentiality in Chinese newspaper reports: Subjectivity/objectivity as a factor. *Discourse Studies*, 10, 205. <https://doi.org/10.1177/1461445607087009>
- Huang, J. (2019). Comparing evidentiality in ESST and ECST texts: a corpus-based investigation of lexicogrammar. *Journal of Chengdu University of Technology (Social Sciences)*, 27(6), 69-77.

- Hunston, S., & Francis, G. (2000). *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. John Benjamins. <https://doi.org/10.1075/scl.4>
- Hyland, K. (2002). Activity and evaluation: Reporting practices in academic writing. In J. Flowerdew (Eds.), *Academic Discourse* (pp. 115-130). Pearson Education Limited.
- Hyland, K. (2004). Disciplinary interactions: metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, 13, 133-151. <https://doi.org/10.1016/j.jslw.2004.02.001>
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192. <https://doi.org/10.1177/1461445605050365>
- Hyland, K. (2017). Points of reference: Changing patterns of academic citation. *Applied Linguistics*, 40(1), 1-23. <https://doi.org/10.1093/applin/amx012>
- Jacquín, J. (2022). A contrastive corpus study of a semantically neutral French evidential marker: tu dis/vous dites [P] [you say [P]] and its relationship with agreement and disagreement. *Journal of Pragmatics*, 199, 75-90. <https://doi.org/10.1016/j.pragma.2022.07.005>
- Janik, C. (2009). As academics we are not disposed to say “I know the world is round...”: Marking of evidentiality. In K. Hyland (Eds.), *Disciplinary Discourses: Social Interactions in Academic Writing* (pp. 1-23). Longman. <https://doi.org/10.1075/pbns.193.02jan>
- Khedri, M. (2018). Evidentials in research articles: A marker of discipline. *Pertanika Journal of Social Science and Humanities*, 26(T), 145-158.
- Lafuente-Millán, E. (2013). Reader engagement across cultures, languages, and contexts of publication in business research articles. *International Journal of Applied Linguistics*, 9, 1-23. <https://doi.org/10.1111/ijal.12019>
- Liao, Q., & Wen, X. (2022). A comparative study of the authorial identity construction in English academic writing from the perspective of evidentiality. *Journal of Xi'an International Studies University*, 30(01), 24-29.
- Liu, S., & Xu, J. (2017). A corpus-based study on evidentiality in the multidisciplinary research articles by Chinese scholars. *Foreign Languages and Their Teaching*, 06, 61-70+147.
- Lou, B., & Qiu, M. (2020). A study of the features of reporting evidentials in Chinese learners' academic writing. *Foreign Languages Research*, 37(06), 67-73.
- Luo, G., & Liao, M. (2013). Evidentiality in courtroom discourse. *Studies in Language and Linguistics*, 33(04), 92-95.
- Marín-Arrese, J. I. (2013). Stancetaking and inter/subjectivity in the Iraq Inquiry: Blair vs. Brown. In J. I. Marín-Arrese, M. Carretero, J. Arús Hita, & J. van der Auwera (Eds.), *English Modality: Core, Periphery and Evidentiality* (pp. 411-446). De Gruyter Mouton. <https://doi.org/10.1515/9783110286328.411>
- Marín-Arrese, J. I. (2015). Epistemicity and stance: A cross-linguistic study of epistemic stance strategies in journalistic discourse in English and Spanish. *Discourse Studies*, 17(2), 210-225. <https://doi.org/10.1177/1461445614564523>

- Murray, S. E. (2020). Annual review of linguistics: Evidentiality, modality, and speech acts. *Annual Reviews*, 7, 213-233. <https://doi.org/10.1146/annurev-linguistics-011718-012625>
- Mushin, I. (2001). *Evidentiality and epistemological stance: Narrative retelling*. John Benjamins. <https://doi.org/10.1075/pbns.87>
- Mushin, I. (2012). "Watching for witness" Evidential strategies and epistemic authority in Garwa conversation. *Pragmatics and Society*, 3(2), 270-293.
- Musi, E. (2016, May). *Strategies of objectification in opinion articles: The case of evidentials*. OSSA 11, Argumentation, Objectivity and Bias, Windsor, Ontario.
- Nuyts, J. (2001). Subjectivity as an evidential dimension in epistemic modal expressions. *Journal of Pragmatics*, 33, 383-400. [https://doi.org/10.1016/S0378-2166\(00\)00009-6](https://doi.org/10.1016/S0378-2166(00)00009-6)
- O'Barr, W. M. (1982). *Linguistic evidence: Language, power, and strategy in the courtroom*. Academic Press.
- Papafragou, A., Li, P., Choi, Y., & Han, C. (2007). Evidentiality in language and cognition. *Cognition*, 103, 253-299. <https://doi.org/10.1016/j.cognition.2006.04.001>
- Pérez Blanco, M. (2020). Evidential adjectives in English and Spanish journalistic opinion discourse. *Journal of Pragmatics*, 170, 112-124. <https://doi.org/10.1016/j.pragma.2020.08.014>
- Rudolph, R. E. (2022). Acquaintance and evidence in appearance language. *Linguistics and Philosophy*, 46(1), 1-29. <https://doi.org/10.1007/s10988-022-09354-1>
- Ruskan, A. (2012). Evidential adjectives in Lithuanian academic discourse. *Kalbotyra*, 64(3), 103-123. <https://doi.org/10.15388/Klbt.2012.7663>
- Saratsli, D., Bartell, S., & Papafragou, A. (2020). Cross-linguistic frequency and the learnability of semantics: Artificial language learning studies of evidentiality. *Cognition*, 197, 1-16. <https://doi.org/10.1016/j.cognition.2020.104194>
- Shui, W., & Ji, Y. (2015). Comparative study of evidentiality in Chinese and English linguistic academic discourses. *Journal of Chongqing University of Posts and Telecommunications (Social Science Edition)*, 27(01), 152-156.
- Smirnova, A. (2021). Evidentiality in abductive reasoning: Experimental support for a modal analysis of evidentials. *Journal of Semantics*, 38(4), 531-570. <https://doi.org/10.1093/jos/ffab013>
- Stubbs, M. (1996). *Text and corpus analysis: Computer-assisted studies of language and culture*. Blackwell.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Tournadre, N., & LaPolla, R. J. (2014). Towards a new approach to evidentiality: Issues and directions for research. *Linguistics of the Tibeto-Burman Area*, 37(2), 1-24. <https://doi.org/10.1075/ltba.37.2.04tou>
- Van Dijk, T. A. (1997). Discourse as interaction in society. In T. A. van Dijk (Eds.), *Discourse as Social Interaction* (Vol. 2, pp. 1-37). SAGE. <https://doi.org/10.4135/9781446221884.n1>

- Wang, G., & Pang, J. (2013). A socio-cognitive discourse analysis framework: A case study of evidentiality in news. *Foreign Languages and Their Teaching*, 19(2), 30-33.
- Wang, S. (2016). Evidentiality research on Chinese and American master's theses. *Contemporary Foreign Language Studies*, (02), 21-27.
- Willett, T. (1988). A cross-linguistic survey of the grammaticalization of evidentiality. *Studies in Language*, 12, 51-97. <https://doi.org/10.1075/sl.12.1.04wil>
- Xu, F. (2012). Citation features in English empirical research discourses. *Journal of Foreign Languages*, 35(6), 60-68.
- Xu, F. (2015). A study of stance markers in second language academic writing. *Foreign Languages and Their Teaching*, 05, 1-7.
- Xue, Y., & Lu, Y. (2021). A contrastive study of evidentiality in China and America's news reports on epidemics. *Contemporary Foreign Language Education*, 02, 51-61.
- Yang, L. (2009). *Evidentiality in English research articles* (Unpublished doctoral dissertation). Xiamen University.
- Yang, L. (2013). Review of evidential studies in China: Status and prospects. *Journal of Shanxi University (Philosophy and Social Science Edition)*, 36(06), 64-68.
- Yang, L. (2015). Academic identity in English research articles: From evidentiality perspective. *Foreign Language Education*, 36(02), 21-25.
- Yang, Y., Xu, M., & Liu, Q. (2019). A contrastive study on evidentiality of academic research articles in English and Chinese. *Journal of Xi'an International Studies University*, 27(03), 48-52.
- Yu, B. (2015). Evidentiality and interpersonal meaning in interdisciplinary doctoral dissertations. *Contemporary Foreign Language Studies*, 04, 29-36+77.

Notes

Note 1. RN represents raw number

Note 2. NF represents normalized frequencies per thousand words

Note 3. P represents percent

Note 4. According to Ackermann et al. (2010), part of the data in spoken corpus are directly downloaded from station websites where transcriptions are available. So, the spoken academic discourse studied in this paper includes both the transcriptions of on-the-spot speech and the texts prepared in advance for the speech scenarios. Given that speakers inherently cannot use parentheses, such a concordance may come from the written material of a speech. Though this kind of text is written initially, it will be presented to listeners by speech so that it is also classified into spoken discourse.