Disciplinary Differences in the Use of the Epistemic Verbs Suggest, Indicate, and Estimate in Academic Writing

Zuzana Kolaříková – Ivana Kupková

DOI: 10.18355/XL.2026.19.02.03

Abstract
This study investigates hedging verbs from a subject–verb collocational perspective, complementing research on hedging in academic discourse. Comparing pairs of academic domains in the COCA corpus, it focuses on the epistemic verbs suggest, indicate, and estimate. It examines noun subject types and their frequency patterns, distinguishing abstract rhetors from concrete sources. The findings reveal that hard sciences favor collocations with abstract rhetors and higher verb frequencies, whereas soft sciences, while also using abstract rhetors, display a greater tendency toward agentive subjects.
Key words: epistemic lexical verbs, hedging, subject-verb collocations, academic writing, disciplinary variation, COCA corpus

Introduction

In academic discourse, particularly in scientific writing, hedging plays a crucial role in shaping how knowledge is conveyed. Rather than presenting claims in absolute or categorical terms, writers use hedging to signal caution, acknowledge uncertainty, and emphasize the provisional nature of their interpretations or arguments. This rhetorical strategy marks the boundary between facts and more tentative propositions (Hyland, 1998; Hinkel, 2005). It also contributes to the tone of critical inquiry that is central to scholarly communication. Since Lakoff (1972) first introduced the concept, hedging has been widely recognized as a defining feature of academic style.

As Mercer, Di Marco, and Kroon (2004) observe, the work of scientists extends beyond the discovery of new knowledge to persuading the academic community of the validity and value of their findings. Hedging plays a crucial role in this persuasive process by helping authors soften their claims, reduce the risk of rejection, and increase their credibility. It enables writers to promote new ideas while demonstrating respect for disciplinary norms.

A substantial body of research has examined disciplinary variation in hedging in English academic writing and has also explored differences between native and non-native academic writers. Some earlier studies have focused on epistemic lexical verbs; however, little is known about the types of subject nouns that co-occur with these verbs and how they vary across different scientific disciplines. In particular, it remains unclear whether certain subject noun types overlap across fields or show discipline-specific patterns in their frequency and distribution. Understanding these patterns has significant implications for the teaching and practice of academic writing, as well as for raising awareness of disciplinary differences. The present study therefore, focuses on epistemic lexical verbs and investigates the types of subject nouns that precede them, examining both overlaps and cross-disciplinary differences in their distribution.

Hedging devices are ambiguous (Salager-Meyer, 1994), semantically complex, i.e., they may express more than one meaning simultaneously (Salager-Meyer, 1994; Hyland, 1998; Varttala, 2001), and therefore they are indeterminate from the pragmatic perspective, as it is the utterance of a statement that allows them to be interpreted as hedges (Hyland, 1998). Due to their polysemous and polypragmatic nature, various approaches to hedging devices have been developed, resulting in various taxonomies (e.g., Salager-Meyer, 1994; Crompton, 1997; Hyland, 1998), but

in general, hedges typically include epistemic modal verbs (e.g., *can, could, may, might*), lexical verbs (e.g., *appear, suggest, tend*), adverbs (e.g., *arguably, supposedly*), and other syntactic constructions (e.g., *it seems that, more or less*). Among these, epistemic modals have received particular attention due to their high frequency and rhetorical significance in research articles (Hyland, 1998; Yagiz & Demir, 2014; Panocova & Lukacin, 2019; Budiarti, Hardjanto & Faris, 2023).

Numerous cross-linguistic and cross-disciplinary studies have highlighted both similarities and differences in the use of hedging between native English-speaking (NES) scholars and non-native English-speaking (NNES) scholars. Cross-linguistic studies that indicate higher occurrence and wider range of hedging devices in NES research articles (RAs) or other types of scientific writing, regardless of the NNES language and disciplinary field, prevail (e.g., Samaie, Khosravian & Boghayeri, 2014, Persian NNES, Introduction sections of literature RAs; Yagiz & Demir, 2014, Turkish NNES, ELT RAs; Chen & Zhang, 2017, Chinese NNES, applied linguistics RAs; Laghari, Akhter & Kakepoto, 2022, Pakistani NNES, doctoral dissertations in applied linguistics). However, not all studies indicate significant differences: for instance, statistically insignificant (Chen & Zhang, 2017) or small differences (Samaie, Khosravian & Boghayeri, 2014; Demirel, 2019) between NES and NNES have been observed in the use of hedging categories.

Disciplinary variation in hedging practices has been documented in several cross-disciplinary studies investigating differences in disciplines belonging to "soft" and "hard" sciences. For example, Varttala's (2001) research explored how the frequency and types of hedging devices reflect the epistemological orientations of different academic disciplines. His findings demonstrate that in economics, a "soft" social science, hedging is often used to emphasize the tentative nature of theoretical claims. In contrast, the field of technology, which is more quantitatively driven, tends to use hedges that focus on numerical precision and accuracy. Medicine, positioned between these two extremes, exhibits a moderate level of hedging: more than in technology but less than in economics, particularly in sections that review previous research or interpret findings. Interestingly, while hedges related to quantification are more common in medicine than in economics, they are still employed less frequently than in technological disciplines. Vázquez and Giner (2008) observed differences in hedging between the soft and hard sciences, with Marketing and Mechanical Engineering representing opposite ends of the spectrum and Biology positioned in between. They suggest that the use of hedging varies not only in frequency across disciplines but also in relation to the nature of the data and research practices typical of each field. A brief overview of the literature on the disciplinary variation in hedging practices is given in Thompson (2022).

Epistemic lexical verbs

Epistemic lexical verbs (ELVs) as hedging devices and their use across disciplines are the focus of the present study. ELVs play a vital role in academic discourse by modulating degrees of certainty and conveying the writer's stance toward a given proposition (Hyland, 1998). These verbs enable authors to qualify their claims, striking a balance between assertion and caution. Drawing on the classifications proposed by Hyland (1998) and Palmer (2001), four main categories of ELVs can be identified: speculative (e.g., *assume, predict, suggest*), deductive (e.g., *conclude, estimate*), quotative (e.g., *argue, note, report*), and sensory (e.g., *appear, seem, notice*). Each category serves a distinct rhetorical function, ranging from logical inference and reasoning to attribution and subjective evaluation.

Therefore, the distribution of these verbs varies across rhetorical sections of RAs, depending on their communicative function. In medical RAs, verbs such as *indicate, speculate, suggest,* and *tend* are frequently used as hedging devices in the Introduction and Discussion sections (Salager-Meyer, 1994; Szczygłowska, 2021; Kozubikova

Sandova, 2017b – hedges in general). In contrast, in linguistics and economics RAs, the highest concentration of ELVs is typically found in the Results section, followed by the Introduction and Discussion sections (Dontcheva-Navratilova, 2018). A similar pattern of use in the Introduction and Discussion has been observed in psychology RAs (Szczygłowska, 2022).

Similarly to the use of hedges in general, both disciplinary and cultural variations have been observed in the employment of various types of ELVs in RAs. While comparing the use of ELVs in medical and psychology RAs by NES scholars, Szczygłowska (2022) revealed that authors in psychology employed these verbs significantly more often, and with greater lexical variety, than their medical counterparts, except in the case of deductive verbs. Authors in psychology frequently used speculative (e.g., *expect*, *assume*), quotative (*suggest*), and sensory verbs (*appear*, *notice*, *seem*), while medical writers relied on such verbs far less. Further evidence of variation comes from Budiarti, Hardjanto & Faris (2023), who compared ELVs in applied linguistics RAs authored by NES and Indonesian NNES writers. Although both groups used hedging verbs at comparable overall frequencies, NES authors preferred speculative and sensory verbs, while NNES authors leaned more toward deductive and quotative verbs. Notably, speculative verbs were the most frequently used category in both groups, while deductive verbs were the least common.

For the purposes of our research, most relevant would be studies on the use of ELVs by Slovak or/and (due to both linguistic and cultural proximity) Czech scholars. Yet there appears to be only one study of that kind: Dontcheva-Navratilova (2018) explored disciplinary and cultural variation by comparing the use of ELVs in economics and linguistics by NES and Czech NNES authors. Her findings revealed that NES authors used ELVs more frequently overall, but preferences varied by discipline: linguists tended to favor evidential and sensory verbs, while economists preferred judgment (speculative and deductive) verbs. These choices appear to align with the epistemological orientations and rhetorical expectations of their respective academic communities. Some other works, although not specifically addressing ELVs, may also be beneficial. For example, two Kozubikova Sandova's (2017a; 2017b) studies on hedging in English and Czech medical RAs. Her findings showed that hedging occurs more frequently in English RAs, but the most frequent types of hedges were the same in both English and Czech corpora (Kozubikova Sandova, 2017a). However, no differences were observed in the communicative function of hedges in both languages (Kozubikova Sandova, 2017b).

Beyond differences in disciplinary and rhetorical distribution, ELVs also perform two key hedging functions in academic discourse. Drawing on Hyland's (1998) framework, these functions can be categorized into two broad types: writer-oriented and reader-oriented. In writer-oriented academic writing, authors often avoid taking direct responsibility for claims while subtly encouraging readers to adopt their perspective. One common strategy is to place epistemic elements (e.g., modal verbs or verbs like *suggest* and *indicate*) early in the sentence. Another is the use of abstract rhetors (Hyland, 1998:172), which are impersonal subjects such as *the data* or *evidence.* They help distance the author from the claim and present the proposition as objective and independent of personal involvement (e.g., *These data indicate that…*, *Taxonomic evidence suggests that…*) (ibid.). Reader-oriented hedging reflects the writer's awareness that gaining the reader's acceptance is essential for recognition within a particular scientific community.

ELVs are essential for qualifying speculative claims and serve as a primary tool for writer-oriented hedging. When used with non-writer agents (ibid.), these verbs signal that the author does not personally endorse the proposition; the hesitancy expressed relates less to the truth of the statement itself and more to the degree of commitment

the writer chooses to convey. It follows that in academic writing, a key distinction between writer-oriented and reader-oriented hedging lies in the subject preceding the verb. Writer-oriented hedges often use impersonal or third-person subjects to obscure the writer's presence, while reader-protecting hedges use explicitly personal subjects to show the writer's direct engagement with the statement.

Methodology

Our research investigates the use of the verbs *indicate*, *suggest*, and *estimate* in RAs within four academic fields. The selection of these verbs was informed by our experience of teaching English for Specific and Academic Purposes (ESP/EAP) at Pavol Jozef Safarik University in Kosice, Slovakia. The textbook *Oxford Grammar for EAP* (Paterson & Wedge, 2013), regularly used in the courses, groups the three verbs together in the section on hedging. It classifies them as weaker in meaning compared to verbs such as *conclude*, *state*, *confirm*, *prove*, or *demonstrate*, and therefore recommends them for expressing hedged academic claims (ibid.:125). The authors also note that *estimate* is typically used for numbers or statistics (ibid.). Furthermore, the selection overlaps with Hyland's (1998:238) suggestion that teachers "focus on high frequency lexical devices in order to develop familiarity with the concept of hedging and confidence in its use. Early stages of ESP courses might therefore concentrate on *indicate*, *suggest*, *appear*, and *propose*, which represent 56% of all mitigating lexical verbs and constitute a spread of both evidential and judgment statements".

This pedagogical context motivated the investigation in the present study. The study aims to complement existing research on hedging in academic writing by empirically examining subject–verb collocations in the COCA corpus. It analyses the nouns that function as subjects, distinguishing between abstract entities (e.g., *study*, *data*, *findings*) and concrete sources, and examines how disciplinary context affects both verb frequency and subject type.

The verbs *suggest* and *indicate* can function as both speculative and quotative ELVs. In their speculative use, they are typically accompanied by first-person pronouns (*I/we*) or abstract rhetors, and they express the writer's tentative stance or judgment, qualifying a claim as uncertain, likely, or open to interpretation. In their quotative use, they are used to attribute a claim, opinion, or finding to an external source, such as a cited researcher or study. The verb *estimate*, by contrast, exemplifies a deductive category of ELVs, which "signal that logic has been applied to a set of known facts to reach a conclusion" (Vass, 2017:22).

This study seeks to address the following research questions:

1. Which nouns most frequently occur as grammatical subjects of the hedging verbs *indicate*, *suggest*, and *estimate* across the selected academic corpus subdomains? Are these nouns abstract rhetors or concrete sources?
2. How are these frequently occurring subject-verb collocations distributed across the disciplinary subdomains under investigation?

The data for this study were drawn from the academic subcorpus of the Corpus of Contemporary American English (COCA) (Davies:2008-), which comprises specialized scholarly texts. As with all language corpora, this freely accessible corpus is a significant tool for vocabulary research in ESP (Panocova, 2017:22). COCA comprises 1,001,610,938 words compiled between 1990 and 2019 and is divided into eight subcorpora. The academic subcorpus includes 120,988,348 words sourced from approximately 100 peer-reviewed journals and is further categorized into ten disciplinary subdomains. For the purposes of this analysis, four subdomains were selected: History (ACAD:History), Medicine (ACAD:Medicine), Humanities (ACAD:Humanities), and Science and Technology (ACAD:Sci/Tech), as they correspond to the academic disciplines represented at the researchers' home

institution. Among these, ACAD:Sci/Tech is the largest, containing 17,454,120 words, followed by ACAD:Humanities (16,203,015 words), ACAD:History (13,395,427 words), and ACAD:Medicine (10,809,528 words).

Two comparative analyses were carried out in this study. The first involved the subcorpora ACAD:Humanities and ACAD:Sci/Tech, selected for their similar word counts and their representation of broad academic fields rather than highly specialized domains. These categories were also intended to reflect the contrast between soft sciences (Humanities) and hard sciences (Science and Technology). The second analysis focused on the more narrowly defined subdomains of History and Medicine, with the former exemplifying a soft science and the latter representing a hard science.

The Collocate function was utilized to identify the most frequent nouns that function as grammatical subjects of the target hedging verbs. A maximum span of one word to the left of the verb was set, thereby focusing on nouns that immediately preceded the verb in subject position. Each instance was manually reviewed to ensure that only those in which the verbs served a hedging function were included. Cases where the identified nouns formed part of larger noun phrases, such as *the results of the study*, were excluded. Additionally, all quoted material was omitted from the frequency counts. In contrast, noun phrases containing uncountable nouns, such as *a body of research* or *a piece of evidence*, were included in the analysis.

The ten most frequent collocations in each subcorpus were identified based on their raw frequency. For countable nouns, singular and plural forms were combined to obtain an overall frequency. Normalized frequencies were also calculated for each collocation to enable valid comparisons across subcorpora of different sizes.

In the next stage, collocations shared across pairs of subdomains were identified. Their observed frequencies were subjected to statistical analysis using the log-likelihood (LL) test to assess the significance of differences. The LL test is particularly suitable for corpus-based studies, as it compares observed word frequencies between two corpora with expected frequencies under the null hypothesis of no difference, while adjusting for corpus size. For this purpose, the online LL calculator (Rayson, 2008) was used. A significance threshold of $p < 0.01$ was applied, corresponding to a critical LL value of 6.63. LL values exceeding this threshold indicate statistically significant differences unlikely to have occurred by chance, whereas values below this threshold suggest that the observed differences are not statistically meaningful.

Results

Noun + INDICATE Collocations

Table 1. Most frequent *noun + indicate* collocations in ACAD:Humanities and ACAD:Sci/Tech.

	NOUN + INDICATE	ACAD:Humanities number of occurrences	ACAD:Humanities normalized frequency per million words	NOUN + INDICATE	ACAD:Sci/Tech number of occurrences	ACAD:Sci/Tech normalized frequency per million words
1	result	41	2.53	result	373	21.37
2	research	33	2.04	data	126	7.22
3	study	27	1.67	study	106	6.07
4	evidence	17	1.05	analysis	98	5.61
5	finding	16	0.99	finding	62	3.55
6	data	12	0.74	respondent	56	2.64
7	score	10	0.62	research evidence	55 55	3.15 3.15
8	title source form	9 9 9	0.56 0.56 0.56	value	34	1.95
9	note	8	0.49	report	32	1.83
10	novel	7	0.43	test	25	1.43

The *noun* + *indicate* collocations exhibit marked distributional differences between the Humanities and Science/Technology subcorpora, with a clear dominance of empirical and data-oriented nouns in the latter. In the ACAD:Sci/Tech subcorpus, *result* appears as the most frequent collocate (21.37 pmw), followed by *data* (7.22 pmw), *study* (6.07 pmw), and *analysis* (5.61 pmw). In contrast, the Humanities subcorpus displays significantly lower frequencies for these collocates, with *result* at only 2.53 pmw, *study* at 1.67 pmw, and *data* at 0.74 pmw. Furthermore, certain collocates, such as *score*, *note*, and *title*, appear exclusively in Humanities, signifying a more interpretive and rhetorical use of *indicate* in this subdomain.

Table 2. LL values for shared *noun* + *indicate* collocations in ACAD:Humanities and ACAD:Sci/Tech.

shared collocations with **INDICATE**	ACAD:Humanities number of occurrences	ACAD:Humanities normalized frequency per million words	ACAD:Sci/Tech number of occurrences	ACAD:Sci/Tech normalized frequency per million words	LL value
result	41	2.53	373	21.37	282.40
data	12	0.74	126	7.22	101.48
study	27	1.67	106	6.07	44.48
finding	16	0.99	62	3.55	25.66
evidence	17	1.05	55	3.15	18.38
research	33	2.04	55	3.15	4.04

The statistical comparison of shared collocates with *indicate* revealed marked disciplinary variation. All shared nouns, except *research*, were significantly more frequent in the ACAD:Sci/Tech subcorpus. The noun *result* showed the greatest disparity (LL = 282.40), followed by *data* (LL = 101.48), *study* (LL = 44.48), *finding* (LL = 25.66), and *evidence* (LL = 18.38), reflecting the empirical and result-oriented discourse characteristic of scientific writing. The difference for *research* was not statistically significant (LL = 4.04), indicating that this collocate may be used more evenly across both domains.

Table 3. Most frequent *noun* + *indicate* collocations in ACAD:History and ACAD:Medicine.

	NOUN + INIDICATE	ACAD:History number of occurrences	ACAD:History normalized frequency per million words	NOUN + INDICATE	ACAD:Medicine number of occurrences	ACAD:Medicine normalized frequency per million words
1	evidence	44	3.28	result	316	29.23
2	report	34	2.54	data	129	11.93
3	respondent	30	2.24	finding study	122 122	11.29 11.29
4	data	29	2.16	respondent	84	7.77
5	research	24	1.79	analysis	75	6.94
6	study	23	1.71	score	57	5.27
7	poll	18	1.34	evidence value	53 53	4.90 4.90
8	record figure	14 14	1.05 1.05	research	48	4.44
9	survey result	11 11	0.82 0.82	report	30	2.78
10	estimate analysis teacher	10 10 10	0.75 0.75 0.75	participant	26	2.41

The collocational patterns of *indicate* reveal notable disciplinary differences between History and Medicine. The Medicine subcorpus exhibits markedly higher frequencies for nouns such as *result* (29.23 pmw), *data* (11.93 pmw), and *study* (11.29 pmw), underscoring its reliance on quantitative evidence and statistical reporting.

Conversely, History features unique collocates, including *poll* (1.34 pmw), *figure* (1.05 pmw), and *estimate* (0.75 pmw), highlighting its engagement with narrative and qualitative sources.

Table 4. LL values for shared *noun + indicate* collocations in ACAD:History and ACAD:Medicine.

shared collocations with **INDICATE**	ACAD:History number of occurrences	ACAD:History normalized frequency per million words	ACAD:Medicine number of occurrences	ACAD:Medicine normalized frequency per million words	LL value
result	11	0.82	316	29.23	426.24
study	23	1.71	122	11.29	97.07
data	29	2.16	129	11.93	91.65
analysis	10	0.75	75	6.94	71.18
respondent	30	2.24	84	7.77	39.52
research	24	1.79	48	4.44	14.13
evidence	44	3.28	53	4.90	3.88
report	34	2.54	30	2.78	0.13

The comparison of shared collocations with *indicate* between the History and Medicine subcorpora demonstrates marked disciplinary distinctions. Medicine exhibits significantly higher normalized frequencies for nouns such as *result* (LL = 426.24), *study* (LL = 97.07), *data* (LL = 91.65), and *analysis* (LL = 71.18), supported by very high log-likelihood values, indicating that these differences are statistically robust. This pattern reflects Medicine's orientation toward empirical research, quantitative data, and statistical reporting. In contrast, History presents lower frequencies for these collocates, emphasizing its more qualitative approach. Collocates like *evidence* and *report* occur with similar frequencies in both disciplines, suggesting some overlap in academic discourse.

The analysis of *noun + indicate* collocations across four academic subdomains reveals a consistent reliance on abstract rhetors such as *result*, *study*, *data*, *evidence*, and *research*. These nouns frequently serve as grammatical subjects for *indicate*, and reflect a key writer-oriented hedging function aimed at distancing the author from the proposition.

Across all domains, *result + indicate* emerges as the most frequent collocation, with particularly high normalized frequencies in Medicine (29.23 pmw) and Sci/Tech (21.37 pmw), compared to much lower frequencies in History (0.82 pmw) and Humanities (2.53 pmw). This disparity underscores the greater emphasis on data-driven reporting and quantifiable outcomes in the scientific disciplines. Sci/Tech and Medicine further demonstrate a broader range of technical or statistical agents (e.g., *respondent, score, participant, test, value*), reinforcing the impersonal, evidence-based rhetorical stance that characterizes scientific writing.

In contrast, Humanities and History, while still dominated by abstract rhetors, include a modest presence of more interpretive or narrative nouns such as *note, record, poll*, and *teacher*. This suggests a somewhat more interpretive style of writing in these fields, with a greater variety of ways of presenting knowledge, although a clear preference for distancing the writer and grounding claims in evidence remains apparent.

The findings confirm disciplinary distinctions in epistemological practices and highlight the importance of tailoring hedging instruction in EAP to reflect these domain-specific norms.

Noun + SUGGEST Collocations

Table 5. Most frequent *noun + suggest* collocations in ACAD:Humanities and ACAD:Sci/Tech.

	NOUN + SUGGEST	ACAD:Humanities number of occurrences	ACAD:Humanities normalized frequency per million words	NOUN + SUGGEST	ACAD:Sci/Tech number of occurrences	ACAD:Sci/Tech normalized frequency per million words
1	evidence	68	4.20	result	322	18.45
2	research	57	3.52	evidence	161	9.22
3	result	52	3.21	study	129	7.39
4	novel	39	2.40	finding	125	7.16
5	study	30	1.85	data	121	6.93
6	finding	25	1.54	analysis	57	3.27
7	data	21	1.30	research	54	3.09
8	title	20	1.23	theory	44	2.52
9	name teacher	14 14	0.86 0.86	observation	36	2.06
10	story author	13 12	0.80 0.74	work	31	1.78

The analysis revealed notable disciplinary and distributional differences in the use of *suggest* with noun subjects. In the ACAD:Sci/Tech subcorpus, *suggest* most frequently followed nouns denoting empirical or data-driven content such as *result* (18.45 pmw), *evidence* (9.22 pmw), *study* (7.39 pmw), *finding* (7.16 pmw), and *data* (6.93 pmw). These high-frequency collocations reflect a discourse style grounded in the reporting of observable phenomena and measurable outcomes. In contrast, the Humanities subcorpus exhibited a more heterogeneous set of collocates, including academic terms such as *evidence* (4.20 pmw), *research* (3.52 pmw), and *result* (3.21 pmw), though with much lower frequencies, as well as more interpretive or culturally situated nouns like *title*, *story*, *author*, and *teacher*.

Table 6. LL values for shared *noun + suggest* collocations in ACAD:Humanities and ACAD:Sci/Tech.

shared collocations with **SUGGEST**	ACAD:Humanities number of occurrences	ACAD:Humanities normalized frequency per million words	ACAD:Sci/Tech number of occurrences	ACAD:Sci/Tech normalized frequency per million words	LL value
result	52	3.21	322	18.45	197.31
data	21	1.30	121	6.93	70.61
finding	25	1.54	125	7.16	65.55
study	30	1.85	129	7.39	59.27
evidence	68	4.20	161	9.22	32.28
research	57	3.52	54	3.09	0.46

The log-likelihood analysis of shared collocations with the verb *suggest* demonstrates statistically significant differences in five out of six items across the two subdomains. The nouns *result* (LL = 197.31), *data* (LL = 70.61), *finding* (LL = 65.55), *study* (LL = 59.27), and *evidence* (LL = 32.28) all occur significantly more frequently in the Science/Technology subcorpus than in Humanities, indicating a strong preference for data-driven hedging structures such as "*The results suggest...*" or "*The data suggest...*". In contrast, the collocation *research* does not differ significantly between the two subdomains (LL = 0.46), suggesting that it functions as a more general-purpose noun used in hedging across disciplines.

Table 7. Most frequent *noun + suggest* collocations in ACAD:History and ACAD:Medicine.

	NOUN + SUGGEST	ACAD:History number of occurrences	ACAD:History normalized frequency per million words	NOUN + SUGGEST	ACAD:Medicine number of occurrences	ACAD:Medicine normalized frequency per million words
1	evidence	111	8.29	finding	356	32.93
2	study	45	3.36	result	331	30.62
3	data	32	2.39	study	225	20.82
4	finding	23	1.72	data	217	20.07
5	research	20	1.49	evidence	211	19.52
6	analysis	18	1.34	research	69	6.38
7	example experience	15 15	1.12 1.12	analysis	60	5.55
8	history report	13 13	0.97 0.97	observation	44	4.07
9	result	12	0.90	author	43	3.98
10	figure survey literature record	11 11 11 11	0.82 0.82 0.82 0.82	report	42	3.89

The distribution of *noun + suggest* collocations reveals notable differences between the History and Medicine subcorpora. Overall normalized frequencies are generally higher in Medicine than in History. In medical writing, the verb *suggest* is most frequently preceded by nouns such as *finding* (32.93 pmw), *result* (30.62 pmw), *study* (20.82 pmw), *data* (20.07 pmw), and *evidence* (19.52 pmw), reflecting the field's strong emphasis on empirical research and statistical reporting. In contrast, the History subcorpus features *evidence* as the most common collocate (8.29 pmw), followed by considerably lower frequencies of *study* (3.36 pmw), *data* (2.39 pmw), and *finding* (1.72 pmw). This suggests that historical writing uses *suggest* more sparingly in conjunction with data-driven terminology and leans toward broader interpretive discourse. Additionally, collocates unique to History, such as *example*, *experience*, *figure*, *literature*, and *record*, further highlight the narrative and analytical character of historical argumentation.

Table 8. LL values for shared *noun + suggest* collocations in ACAD:History and ACAD:Medicine.

shared collocations with **SUGGEST**	ACAD:History number of occurrences	ACAD:History normalized frequency per million words	ACAD:Medicine number of occurrences	ACAD:Medicine normalized frequency per million words	LL value
result	12	0.90	331	30.62	443.81
finding	23	1.72	356	32.93	427.71
data	32	2.39	217	20.07	196.72
study	45	3.36	225	20.82	172.70
evidence	111	8.29	211	19.52	56.72
research	20	1.49	69	6.38	40.07
analysis	18	1.34	60	5.55	33.76
report	13	0.97	42	3.89	22.94

The comparison of shared *noun + suggest* collocations between the History and Medicine subcorpora reveals a consistent and statistically significant preference for these structures in medical texts. All eight collocates analyzed – *result, finding, data, study, evidence, research, analysis,* and *report* – occur far more frequently in Medicine than in History, with LL values well above the significance threshold (e.g., *result*, LL = 443.81; *finding*, LL = 427.71). These differences underscore the central role of empirical evidence in medical discourse, where hedging is typically grounded in measurable outcomes, quantitative findings, and structured analyses. In contrast,

although *evidence* appears with relatively high frequency in the History subcorpus (8.29 pmw), its use is still significantly outpaced by Medicine (19.52 pmw), and other collocates such as *study* and *data* occur only marginally in historical writing. This suggests that historical texts employ *suggest* less frequently in combination with technical or research-based subjects, reflecting a discourse style that prioritizes interpretive reasoning over empirical generalization.

The analysis of *noun + suggest* collocations reveals both cross-disciplinary consistency and notable variation. Across all domains, the verb *suggest* most frequently co-occurs with abstract rhetors such as *evidence*, *result*, *study*, *data*, and *finding*, reflecting a general preference for depersonalized, externally grounded knowledge claims. This pattern is more consistent in the Sci/Tech and Medicine subcorpora, where all top collocates are abstract and data-driven. In Medicine, for instance, *finding* and *result* exceed 30 occurrences per million words, underscoring the centrality of data-driven epistemologies in scientific discourse.

While the Humanities and History subcorpora also rely on abstract rhetors, they exhibit a broader range of subject types, including more interpretive or semi-concrete nouns such as *title*, *name*, *novel*, *teacher*, and *experience*. This suggests a comparatively more flexible or personalized rhetorical style in these fields, where subjective interpretation may be more readily accepted as part of scholarly argumentation.

These findings suggest the impact of disciplinary conventions on hedging practices, particularly in relation to epistemic stance and source attribution. From a pedagogical perspective, they highlight the importance of raising students' awareness not only of hedging verbs themselves, but also of how these verbs typically interact with subject nouns in different academic domains.

Noun + ESTIMATE Collocations

Table 9. Most frequent *noun + estimate* collocations in ACAD:Humanities and ACAD:Sci/Tech.

	NOUN + ESTIMATE	ACAD:Humanities number of occurrences	ACAD:Humanities normalized frequency per million words	NOUN + ESTIMATE	ACAD:Sci/Tech number of occurrences	ACAD:Sci/Tech normalized frequency per million words
1	analysis	2	0.12	model study	15	0.86
2	anthropologist citizen program researcher research statistics teacher	1 1 1 1 1 1 1	0.06 0.06 0.06 0.06 0.06 0.06 0.06	researcher	13	0.74
3				analysis expert	8	0.46
4				agency report scientist	7 7 7	0.40 0.40 0.40
5				engineer group official	5 5 5	0.29 0.29 0.29
6				astronomer colleague government organisation survey parameter	1 1 1 1 1 1	0.06 0.06 0.06 0.06 0.06 0.06

In general, the verb *estimate* appeared very sparingly, however, it was used more frequently in ACAD:Sci/Tech than in ACAD:Humanities. In the Humanities subcorpus, *estimate* occurred only rarely, with the most frequent subject noun –

analysis, appearing just twice (0.12 pmw), and all others registering a frequency of 0.06 pmw. In contrast, Sci/Tech texts showed greater lexical variety and higher frequencies, with *model* (0.86 pmw), *researcher* (0.74 pmw), and *analysis* (0.46 pmw) among the most common collocates.

Table 10. LL values for shared *noun + estimate* collocations in ACAD:Humanities and ACAD:Sci/Tech.

shared collocations with **ESTIMATE**	ACAD:Humanities number of occurrences	ACAD:Humanities normalized frequency per million words	ACAD:Sci/Tech number of occurrences	ACAD:Sci/Tech normalized frequency per million words	LL value
analysis	2	0.12	8	0.46	3.42
researcher	1	0.06	13	0.74	11.33

When examining shared collocates of the verb *estimate* in the Humanities and Sci/Tech subcorpora, two nouns appeared in both domains. The noun *analysis* occurred more frequently in Sci/Tech (0.46 pmw) than in Humanities (0.12 pmw), though the difference was not statistically significant (LL = 3.42). The noun *researcher* showed a stronger and statistically significant association with Sci/Tech (0.74 pmw) compared to Humanities (0.06 pmw), with a log-likelihood (LL) value of 11.33, which exceeds the critical value for $p < 0.01$.

Table 11. Most frequent *noun + estimate* collocations in ACAD:History and ACAD:Medicine.

	NOUN + ESTIMATE	ACAD:History number of occurrences	ACAD:History normalized frequency per million words	NOUN + ESTIMATE	ACAD:Medicine number of occurrences	ACAD:Medicine normalized frequency per million words
1	study	12	0.90	study	11	1.02
2	government	8	0.60	author report	5 5	0.46 0.46
3	source	5	0.37	researcher	2	0.19
4	analyst	4	0.30	agency bureau expert finding	1 1 1 1	0.09 0.09 0.09 0.09
5	expert	3	0.22			

The distribution of *noun + estimate* collocations shows that both History and Medicine rely most on the noun *study*, with similar normalized frequencies (0.90 and 1.02 pmw, respectively). Beyond this, History displays variation with nouns like *government* (0.60 pmw), *source* (0.37 pmw), and *analyst* (0.30 pmw). Medicine shows lower overall frequencies for other sources, such as *agency*, *bureau*, *finding*, *expert*, *researcher*, *author*, and *report*, which occur less frequently (ranging from 0.09 to 0.46 pmw).

Table 12. LL values for shared *noun + estimate* collocations in ACAD:History and ACAD:Medicine.

shared collocations with **ESTIMATE**	ACAD:History number of occurrences	ACAD:History normalized frequency per million words	ACAD:Medicine number of occurrences	ACAD:Medicine normalized frequency per million words	LL value
study	12	0.90	11	1.02	0.09
expert	3	0.22	1	0.09	0.66

Two shared collocates of *estimate* were found in both the History and Medicine subcorpora: *study* and *expert*. The normalized frequencies of *study* were very similar

in both domains, 0.90 pmw in History and 1.02 pmw in Medicine, resulting in a non-significant log-likelihood (LL) value of 0.09. The noun *expert* occurred slightly more often in History (0.22 pmw) than in Medicine (0.09 pmw), but the LL value (0.66) once more indicates that this difference is not statistically significant.

Compared to *indicate* and *suggest*, *estimate* is used much less frequently as a verb preceded by a noun subject across all disciplines. When it does occur, the subjects tend to be more concrete, often referring to specific institutions, organizations, agencies, or named entities rather than abstract concepts. The overall frequency of *estimate* with a preceding subject is quite low, with only 12 occurrences in Humanities, 113 in Sci/Tech, 104 in History, and 40 in Medicine. Within these, the more general or less specific subjects (such as *study* or *analysis*) appear relatively rarely; most are concrete, institution-related nouns which were excluded from the present analysis. This pattern highlights that *estimate* typically functions in contexts where concrete sources or authoritative bodies provide evaluations or figures. Additionally, corpus data from COCA show that *estimate* occurs far more frequently as a noun than as a verb in noun-noun collocations.

Despite the sparse distribution of *noun + estimate* collocations, it can be concluded that both abstract rhetors and concrete sources serve as subjects in Sci/Tech and Medicine. For example, in Sci/Tech, abstract nouns like *model* and *analysis* appear alongside concrete nouns such as *researcher*, *expert*, and *scientist*. Similarly, Medicine balances abstract sources like *study* with concrete agents such as *author* and *researcher*. Humanities and History predominantly feature abstract rhetors like *study* and *analysis* but also include a more diverse range of institutional and interpretive sources such as *analyst*, *citizen*, *government*, *teacher*, *researcher*, and *anthropologist*. As noted earlier, *estimate* also frequently follows institutional names, which underscores its association with formal, often authoritative, sources.

Discussion and Conclusions

The analysis of the epistemic lexical verbs *indicate*, *suggest*, and *estimate* across academic subdomains in COCA reveals notable differences in frequency, subject types, and collocational patterns. Across all four domains, the top ten nouns functioning as grammatical subjects of the studied ELVs typically serve as either abstract rhetors or concrete sources. While top collocates in the hard sciences predominantly include abstract nouns such as *result*, *data*, *finding*, and *study*, the Humanities and History exhibit a more diverse mix, incorporating both abstract nouns and concrete or agentive subjects such as *novel*, *title*, *teacher*, *government*, *source*, or *analyst*.

Several shared collocates were identified across pairs of disciplines. Both Humanities and Sci/Tech include abstract nouns such as *data*, *finding*, *evidence*, *research*, *result*, and *study* among their top ten collocates with *indicate* and *suggest*. Similarly, History and Medicine share *analysis*, *data*, *evidence*, *report*, *research*, *result*, and *study*, with all but *report* functioning as abstract rhetors. In addition, *respondent + indicate* (a concrete source) and *finding + suggest* (an abstract rhetor) further complete the set of shared collocations between History and Medicine. This overlap suggests that while disciplinary conventions vary, certain academic discourse practices transcend individual fields.

Despite overlap in collocate types, the frequency patterns of shared collocates differ considerably across disciplines. The hard sciences show much higher distributions. For instance, *result + suggest* occurs 18.45 times per million words (pmw) in Sci/Tech and 30.62 pmw in Medicine, compared to only 3.21 pmw in Humanities and 0.90 pmw in History. This finding aligns with previous research (Varttala, 2001; Vázquez & Giner, 2008; Dontcheva-Navratilova, 2018; Panocova & Lukacin, 2019; Szczygłowska, 2022), which has also observed disciplinary variation in the frequency and form of hedging.

The verb *estimate* departs from this overall pattern in both usage and distribution. It appears far less frequently than *suggest* or *indicate* and is most commonly used with institutional agents or technical nouns (e.g., *government*, *agency*, *expert*, *report*), especially in History, Medicine, and Sci/Tech. In the Humanities, its use is rare. Unlike *suggest* and *indicate*, *estimate* is rarely preceded by abstract rhetors and often occurs in noun–noun constructions (e.g., *cost estimate*) in COCA.

The findings of this study also have implications for teaching academic writing, particularly in EAP/ESP contexts. The study contributes to the use of corpora in EAP teaching and course design, as proposed by Thompson (2022). EAP/ESP textbooks often present generalizations as helpful guidelines; however, these should be supplemented with discipline-specific insights informed by research-based data. For example, *Oxford Grammar for EAP* (Paterson & Wedge, 2013) groups *indicate*, *suggest*, and *estimate* together as hedging verbs and recommends them for expressing cautious claims. This is a useful starting point, but the corpus findings reveal considerable disciplinary variation in how these verbs are actually used. Therefore, discipline-specific patterns must be acknowledged and integrated into teaching practice.

To conclude, the findings show that noun + *indicate* and noun + *suggest* collocations are used in discipline-specific ways that reflect broader rhetorical and epistemological preferences. Sci/Tech and Medicine tend to pair these hedging verbs with abstract, data-oriented subjects and use them at significantly higher frequencies. Humanities and History also use abstract subjects, but more frequently accommodate concrete, interpretive sources. *Estimate*, while relevant in certain domains, plays a more marginal and specialized role in academic prose.

These corpus-based findings contribute to linguistic analyses of hedging. They illustrate both shared and discipline-specific patterns of subject–verb collocations in academic writing. The results provide empirical support for linguistic analyses of hedging and can inform applied approaches to discourse analysis, genre studies, and field-specific academic writing research.

Limitations and further research

This study has several limitations that should be acknowledged. First, although COCA is a robust and balanced corpus, it primarily reflects U.S.-based published academic writing and may not fully represent global academic discourse practices. Second, the disciplinary subdomains used in this analysis (i.e., Sci/Tech, Medicine, Humanities, History) encompass a broad range of subfields, which may fail to capture subtle disciplinary differences that exist within each subdomain. Third, while based on prior research, the division of subject nouns into abstract rhetors and concrete sources is partly subjective and not always clear-cut. Fourth, the study did not analyze the rhetorical orientation (writer- vs. reader-oriented) of the verbs in detail, which may influence their interpretation and therefore deserves further investigation. Finally, the study focused exclusively on subject-verb collocations and did not account for other hedging constructions, such as those involving epistemic modals (e.g., *may suggest*) or personal pronouns (e.g., *we suggest*). These additional structures also shape how hedging is expressed in academic prose and can influence the frequency distribution of the verbs under analysis.

Future research could explore how hedging verbs *indicate*, *suggest*, and *estimate* co-occur with a broader range of hedging devices, such as epistemic modals (*may*, *might*), adverbs (*possibly*, *likely*), and impersonal structures (*It is suggested that*). Such investigations would enhance our understanding of how hedging strategies manifest in English academic writing across disciplines.

Declaration on AI use

The authors used ChatGPT to assist with language editing and stylistic refinement. All research content and interpretations remain the sole responsibility of the authors.

Funding

This study was supported by the grant No. VVGS-2023-2762 The analysis of the use of hedging devices in English academic writing in selected human, social, and natural sciences.

Bibliographic references

Budiarti, D., Hardjanto, T. D., & Faris, I. N. I. (2023). Lexical verbs of hedging in English research articles by native and non-native speakers. Rainbow: Journal of Literature, Linguistics and Culture, 12(1), 29-39. https://doi.org/10.15294/rainbow.v12i1.67067

Chen, Ch., & Zhang, L. J. (2017). An intercultural analysis of the use of hedging by Chinese and Anglophone academic English writers. Applied Linguistics Review, 8(1), 1-34. https://doi.org/10.1515/applirev-2016-2009

Crompton, P. (1997). Hedging in Academic Writing: Some Theoretical Problems. English for Specific Purposes, 16(4), 271-287. ISSN 0889-4906

Davies, M. (2008-). The Corpus of Contemporary American English (COCA). Available online at https://www.english-corpora.org/coca/

Demirel, T. E. (2019). The Use of Hedging Strategies in Research Articles: A Corpus Comparison of Native and Non-Native Researchers. Karabuk University Journal of Institute of Social Sciences, 9(1), 349-362. ISSN 2147-7841

Dontcheva-Navratilova, O. (2018). Intercultural and interdisciplinary variation in the use of epistemic lexical verbs in linguistics and economics research articles. Linguistica Pragensia, 28(2), 154-167. ISSN 1805-9635

Hinkel, E. (2005). Hedging, Inflating, and Persuading in L2 Academic Writing. Applied Language Learning, 15(1-2), 29-53. ISSN 2164-0912

Hyland, K. (1998). Hedging in Scientific Research Articles. Amsterdam/Philadelphia: John Benjamins. ISBN 9027250677

Kozubikova Sandova, J. (2017a). "…and our study might therefore have been slightly underpowered": A Cross-Linguistic Analysis of Hedging in English and Czech Medical Research Articles. AUC Philologica, 2017 (1), 115-130. DOI: 10.14712/24646830.2017.7

Kozubikova Sandova, J. (2017b). Hedging Expressions in Different Sections of English and Czech Medical Research Articles. A Comparative Study. Ostrava Journal of English Philology, 9 (1), 7-22. ISSN 2571-0257

Laghari, T., Akhter, T., & Kakepoto, I. (2022). Softening the Propositional Claims: Investigation of Appropriate Use of Hedging Strategies in Academic Writing. Global Language Review, VII(II), 429-441. https://doi.org/10.31703/glr.2022(VII-II).35

Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. Journal of Philosophical Logic, 2, 458-508.

Mercer, R. E., Di Marco, Ch., & Kroon, F. W. (2004). The Frequency of Hedging Cues in Citation Contexts in Scientific Writing. In Tawfik, A. Y., & Goodwin, S. D. (Eds.). Advances in Artificial Intelligence. Berlin/Heidelberg: Springer, pp. 75-88.

Palmer, F. R. (2001). Mood and Modality. Cambridge: CUP. ISBN 0521804795

Panocova, R. (2017). The Vocabulary of Medical English: A Corpus-based Study.Newcastle upon Tyne: Cambridge Scholars Publishing. ISBN 1443895784, ISBN 9781443895781

Panocova, R., & Lukacin, L. (2019). Epistemic modal markers in two domains of academic research papers in English. Brno Studies in English, 45(2), 121-138. https://doi.org/10.5817/bse2019-2-6

Paterson, K., & Wedge, R. (2019). Oxford Grammar for EAP. Oxford: Oxford University Press. ISBN 9780194329996
Rayson, P. (2008). Log-likelihood and effect size calculator. https://ucrel.lancs.ac.uk/llwizard.html
Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. English for Specific Purposes, 13(2), 149-170. https://doi.org/10.1016/0889-4906(94)90013-2
Samaie, M., Khosravian, F., & Boghayeri, M. (2014). The Frequency and Types of Hedges in Research Article Introductions by Persian and English Native Authors. Procedia – Social and Behavioral Sciences, 98, 1678-1685. https://doi.org/10.1016/j.sbspro.2014.03.593
Szczygłowska, T. (2022). Lexical Verbs of Epistemic Modality in Academic Written English: Disciplinary Variation. Linguistica Silesiana, (43), 91-11. https://doi.org/10.24425/linsi.2022.141219
Thompson, P. (2022). Corpus analysis of disciplinary variation and the teaching of EAP/EAP. Jablonkai, R. R., & Csomay, E. (eds.). The Routledge Handbook of Corpora and English Language Teaching and Learning. London/New York: Routledge, 177-192. ISBN 9781003002901
Varttala, T. A. (2001). Hedging in scientifically oriented discourse: Exploring variation according to discipline and intended audience. Electronic dissertation. University of Tampere. ISBN 9514451953
Vass, H. (2017). Lexical verb hedging in legal discourse: The case of law journal articles and Supreme Court majority and dissenting opinions. English for Specific Purposes, 48, 17-31. https://doi.org/10.1016/j.esp.2017.07.001
Yagiz, O., & Demir, C. (2014). Hedging strategies in academic discourse: A comparative analysis of Turkish writers and native writers of English. Procedia, Social and Behavioral Sciences, 158, 260-268. https://doi.org/10.1016/j.sbspro.2014.12.085

Words: 6 203
Characters: 44 216 (25 standard pages)

Mgr. Zuzana Kolaříková, PhD.
Language Training Centre, Faculty of Arts
Pavol Jozef Šafárik University in Košice
Moyzesova 9, 044 01 Košice
Slovakia
zuzana.kolarikova@upjs.sk
ORCID: 0000-0001-6756-6752

Mgr. Ivana Kupková, PhD.
Language Training Centre, Faculty of Arts
Pavol Jozef Šafárik University in Košice
Moyzesova 9, 044 01 Košice
Slovakia
ivana.kupkova@upjs.sk
ORCID: 0000-0002-1735-4329