English Verbs Tenses and Modal Verbs of Social and Humanities Academic Texts from COCA

Gusti Ayu Praminatih

{gusti.praminatih@ipb-intl.ac.id}

Faculty of Tourism and Business, Institut Pariwisata dan Bisnis Internasional, Denpasar, Indonesia

Abstract. Comprehending English verb tenses (EVT) and modal verbs (MV) is crucial for L2 academic writing. This study aims to analyse EVT and MV occurrences from a significant corpus, The Corpus of Contemporary American English (COCA). The field of study's primary focus comprises social and humanities disciplines, primarily language, literature, and education. This study also explores how EVT and MV are generated from the corpora presented inside and outside Coxhead's Academic Word List (AWL). Corpus linguistics was employed as the methodology of this study. COCA data were retrieved following the determined academic texts. The academic text results were annotated using CLAWS Free Tagger. Further, the annotated data were inputted into the AntConc software. C7 tagsets were entered alternately to make the desired tenses accessible from the corpus. The findings revealed that each discipline comprised the base form of the lexical verb, past tense of the lexical verb, *-ing* participle of the lexical verb, infinitive, past participle of the lexical verb, *-s* form of the lexical verb, and modal verb. However, the frequencies of EVT and MV were different in each discipline. When compared against Coxhread's AWL, some of the EVT came from the most frequent of 1 sublist to the least frequent of 10 sublists. Meanwhile, EVT outside the existing AWL could be the new academic word list for each discipline. Furthermore, theoretical and practical implications are further elaborated in this study.

Keywords: academic writing, COCA, English corpora, English verbs tenses, modal verbs

1 Introduction

Writing and publishing academic papers in English have become compulsory in the publication realms [1], [2], [3], [4]. Although there is a demand to promote language diversity in publications to prevent other language loss in the scientific domain [5], [6], [7], English is claimed as the lingua franca among academia [8] with 75 per cent social and humanities science and 90 per cent of natural science are published in English (Hamel, 2007). Moreover, driven by the motivation to level up universities ranking globally, the demand for publications in English to gain more audiences and establish a sense of professionalism in academic writing is becoming indisputable [10], [11], [12], [13]. The existing literature mentions that the researchers perceived English as a means to promote their works, leading to internationalisation and recognition by the academic community peers [7]. Language does not solely function to communicate but also construct the knowledge written in the

paper. The existing literature demonstrates that L1-English tend to be more consistent in the narrative and argumentative than L2-English in academic writing [14].

Despite the significance of using English in the academic community, a study mentions that many novices struggle with English to meet the standard practices, which require advanced native English to check their work before publication [3]. Similarly, a study also describes that writing academic papers in English makes it difficult for novices, especially L2, to express their writing in native-like formulaic progressions [15]. Consequently, to achieve academic writing success in academic writing, one of the crucial steps is to identify the formal properties of the language, i.e., the set of grammatical rules. Moreover, the existing literature states that academic writing in English may differ from academic writing in L2 authors' language, including the words, grammar, and how to organise ideas [16]. The existing literature also suggests that grammatical accuracy is mandatory in L2 academic writing [17]. Consequently, the present study investigates English grammar, focusing on verb tenses (EVT) and modal verbs (MV).

EVT is the foundation for constructing meaningful sentences in academic papers [18]. Further, the existing literature shows the significance of using the correct tenses in academic papers for different citation styles, such as in the Modern Language Association (MLA) and the Chicago Manual of Style (CMS), the cited work is written in the present tense; meanwhile, the American Psychological Association (APA), the cited work is written in past or past perfect tense, and the present tense is written in discussion section [19]. Thus, it is a call to identify EVT by conducting the present study because recognising EVT is paramount as the underpinning of constructing academic papers and adapting the preferred EVT depending on the citation style applied in the papers. Further, the literature reveals that in RA abstracts from Elsevier across disciplines, several types of EVT are employed, i.e., present simple and past tense as the most frequent tenses, followed by present perfect, past perfect, and present progressive tenses [20]. EVT has various classes, inflexions, and tenses; writing academic papers becomes a struggle for L2, such as Japanese, Chinese, and Korean authors from various education levels and writing expertise [21], [22]. Subsequently, the study's first aim is to identify EVT forms and frequencies in social and humanities science by employing a significant corpus called Corpus of Contemporary American English (COCA).

Scholars examine the roles and functions of MV academic writing. In academic writing, MV usage is significantly influenced by academic essay topics among L1 and L2 authors [21]. A study demonstrates that MV is an interpersonal grammatical metaphor [23]. Furthermore, a study also mentions that MV strengthen hypothesis interpretation sections in writing research articles [24]. Subsequently, it is reported that MV is found more frequently in soft sciences, i.e., linguistics and economics, than in hard sciences [25]. Then, a study reveals that Elsevier RA abstracts across disciplines, among eight types of MV, *can* become the most frequent form [20]. In addition, the existing literature reveals that MV can be categorised as modals and semi-modals, the second most frequent type of grammar, which functioned as one of the stance markers in research article abstracts [26]. A recent study indicates how L2 learners develop the pattern of using and misusing MV in their English essay writing [22]. Moreover, in a significant corpus such as The Freiburg-LOB Corpus of British English (FLOB), MV indicates a strong authoritarian tone, i.e., it *can* and *could* increase over time despite indicating a weaker modality [27]. Drawing from the literature, MV

frequently appears in social and humanities. Accordingly, the study's second aim is to explore MV forms and frequencies in social and humanities science using COCA.

Therefore, conducting a study that investigates EVT using significant data that could reflect authentic language use in academic writing is crucial and is beneficial to novice and L2 authors. An established corpus, i.e., COCA, is utilised. Scholars describe that a corpus is a massive amount of data stored in a computer that reflects the true nature of academic text language patterns [28], [29], [30]. Since the data generated from the corpus can represent the EVT and MV in academic writing, verifying the findings to Coxhead's [31] Academic Word List (AWL). AWL is a reputable guideline to identify English word lists for academic purposes and their frequency. It also becomes the basis for constructing new academic word lists [18], [32]. Thus, the third aim of this study is to examine EVTs present in the AWL and those that appeared outside AWL, which can potentially be constructed as the new social and humanities academic word list.

2 Methodology

Corpus has been extensively utilised to examine formal language properties in academic text across disciplines, including social and humanities. For this study, a significant corpus from English Corpora was utilised. The Corpus of Contemporary American English (COCA) was selected following two considerations. First, COCA was a significant corpus consisting of one billion words [33]. This extensive data ensured the availability and representativeness of the findings. Second, COCA contained written and spoken texts from various sources and genres. One of the genres was academic texts from multiple disciplines. Accordingly, with the extensiveness of the data and the wide range of academic text genres, the selection of COCA was considered appropriate to achieve the objectives of this study. Further, the present study included the analysis of the social and humanities field of study with three disciplines, i.e., Language, Literature, and Education. The selection was based on the existing literature, which indicated that social and humanities were more significant regarding word counts than other fields of studies [34], [35].

After determining the field of study and disciplines, seven steps were taken to obtain and analyse the data from COCA. First, the Keyword in Context (KWIC) menu was selected, and then the keywords related to each discipline were entered alternately. Second, after inputting the keywords, verb.ALL in the POS section was chosen. This step was crucial to ensure that the verb would be generated from COCA. Third, the automated window span from COCA was applied. The window spans were zero left and three right. Fourth, after these settings were all set, the next step was clicking to retrieve the findings. Fifth, after the results were generated, meticulous selection was followed to verify that only data from desired academic texts were obtained. In this step, the top 50 of the results generated by COCA were selected from each discipline. Thus, 150 were gathered for this study.

Sixth, these selected data were annotated using The Constituent Likelihood Automatic Word-tagging System (CLAWS) Tagger [36]. It was crucial to annotate the data to maintain the evidentiality of data when retrieved from the corpus software. The selection of the CLAWS Tagger was due to the accuracy, which reached 96.97%, with only 1.5% error rates and 3.3% ambiguities

[37]. In this step, the UCREL CLAWS7 Tagset (C7) was utilised to make annotations because it was the latest standard of the data tagset [36].

Table 1. The Selected UCREL CLAWS7 Tagset (C7)

C7 Tagset	Forms
VV0	base form of lexical verb
VVD	past tense of lexical verb
VVG	*-ing* participle of lexical verb
VVI	infinitive
VVN	past participle of lexical verb
VVZ	*-s* form of lexical verb
VM	modal verb

Seventh, the data retrieved were categorised based on the EVT and were further compared against Coxhead's AWL. The data that existed in Coxhead's AWL were completed with the frequency of the AWL. Meanwhile, the data that were not presented in Coxhead's AWL were listed and proposed as a new social and humanities academic word list.

3 Findings and Discussion

3.1 Findings

The EVT data obtained from COCA was presented to achieve the first objective of the study. It is crucial to note that due to the nature of the source, the generated results contained some American-spelled words. Subsequently, to maintain the authenticity of the form of data retrieved, these words were kept in their original spelling. The EVT retrieved in this study are illustrated in Table 2.

Table 2. EVT in Academic Texts from COCA

Field of Studies	Base Form of Lexical Verb	Past Tense of Lexical Verb	*-ing* Participle of Lexical Verb	Infinitive	Past Participle of Lexical Verb	*-s* Form of Lexical Verb
Language	consider	indicated	using	investigate	restricted	needs
	absorb	granted	teaching	to complete	attached	takes
	appear		referring	to instil	related	explains
	suggest		trying	perform	asked	reveals

take		reducing	to transcribe	reviewed	reflects
distinguish		adopting	code	solved	allows
resource		rating	to illuminate	used	exposes
understand		receiving	to frame	assessed	creates
focus		validating	undergo	eliminated	functions
see			believe	couched	defines
discuss			to express	based	represents
improve			to see	created	assumes
				required	
Literature find	specialized	conducting	to read	found	displays
generalize	emerged	reading	master	encountered	provides
read	appropriated	taking	inform	characterized	includes
see	filled	discussing	consider	dismissed	invokes
examine	began	experiencing	recognize	informed	examines
remain	discussed	examining	suggest	focused	calls
investigate	called		to add	detached	observes
know	worked		begin	referred	exposes
	located		to raise	sent	contains
	focused		help	adopted	exists
	grew		to gain	focused	relies
	demanded		to measure	closed	supports
			to provide	composed	indicates
				demonstrated	
				entitled	
				grown	
				suggested	
				given	
				included	
Education offer	indicated	conducting	to address	developed	seems

construct	estimated	creating	to access	made	fails
transfer	argued	developing	to complete	replaced	contributes
use	criticized	clarifying	to hold	received	affects
design	followed	expanding	work	customized	delineates
claim	completed	changing	encourage	designed	crosses
address	emphasized	getting	receive	failed	address
complete	noted	distinguishing	serve	drawn	
need	served	assuming	to raise	pursued	
	suggested	providing	benefit	developed	
		marketing	to inform	indicated	
		learning	to observe	received	
		drawing	to resolve	misplaced	
		helping	to enhance	referred	
		taking	to affect	rejected	
			assume	seen	
			adopt	linked	
			to introduce	based	
			to give	assessed	
			know		

Drawing from Table 2. it can be inferred that EVT from Language, Literature, and Education appeared in the form of the base form of the lexical verb, past tense of the lexical verb, *-ing* participle of the lexical verb, infinitive, past participle of the lexical verb, *-s* form of the lexical verb, modal verb. The MV data obtained from COCA was presented to attain the second objective of the study. The MV generated in this study are depicted in Table 3.

Table 3. MV in Academic Texts from COCA

Field of Study	Modal Verb
Language	could
	can
	would
	will
Literature	would
	can
	should
	may
	could
Education	will
	might
	may
	would
	should
	could

Drawing from Table 3. it can be suggested that MV from language, literature, and education are distinct. Furthermore, when observed from the frequencies, there were some differences in the distribution of EVT and MV, as illustrated in Figure 1.
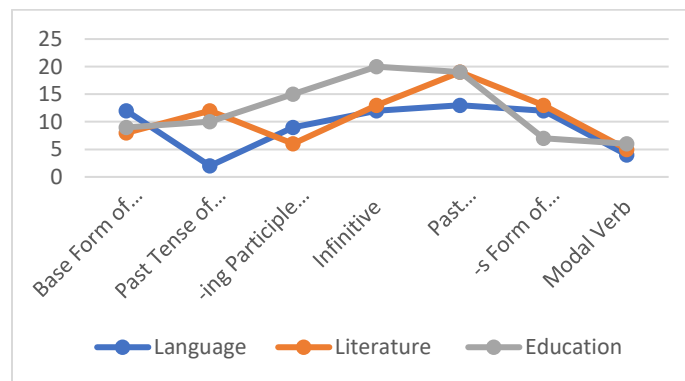

Figure 1. The Frequency Distributions of EVT and MV from COCA

From Figure 1. it can be described that EVT in Language academic texts generated from COCA, the frequency of base form lexical verb was twelve occurrences, two occurrences of the past tense of lexical verb, nine occurrences of *-ing* participle of lexical verb, twelve occurrences of infinitive, thirteen occurrences of past participle lexical verb, and twelve occurrences in *-s* for of lexical verbs. Accordingly, the highest EVT occurrence in Language academic text was the past participle of the lexical verb, and the lowest occurrence was the past tense of the lexical verb, followed by the *-ing* participle of lexical verbs. Meanwhile, the base form of the lexical verb, infinitive, and *-s* form of lexical verb occurrences were alike.

Further, it can be elaborated that EVT in Literature academic texts obtained from COCA, the frequency of base form lexical verb was eight occurrences, twelve occurrences of the past tense of lexical verb, six occurrences of *-ing* participle of lexical verb, thirteen occurrences of infinitive, nineteen occurrences of past participle lexical verb, and thirteen occurrences in *-s* form of lexical verb. Subsequently, the highest EVT occurrence in Literature academic text was the past participle of the lexical verb, and the lowest occurrence was the *-ing* participle of the lexical verb followed by the base form of the lexical verb. Concurrently, the infinitive and *-s* forms of lexical verb occurrences were alike, followed by past tense with slightly lower occurrences.

Eventually, it can be asserted that in EVT in Education academic texts retrieved from COCA, the frequency of base form lexical verb was nine occurrences, ten occurrences of the past tense of lexical verb, fifteen occurrences of *-ing* participle of lexical verb, twenty occurrences of infinitive, nineteen occurrences of past participle lexical verb, and seven occurrences in *-s* form of lexical verbs. Consequently, the highest EVT occurrences in education academic texts were the past participle of lexical verbs, and the lowest occurrence was the past tense of lexical verbs, followed by the *-ing* participle of lexical verbs. Simultaneously, the base form of lexical verb, infinitive, and *-s* form of lexical verb occurrence were alike.

Furthermore, following the data gathered from the three academic disciplines from COCA, it can be explained that in Language academic texts, the frequency of MV was four occurrences. Then, in Literature academic texts, the frequency of MV was five occurrences. Ultimately, in Education academic texts, the frequency of MV was six occurrences. Therefore, it can be demonstrated that the highest occurrence of MV was in academic texts on education, followed by literature, and language held the lowest position in terms of MV occurrence.

To accomplish the third objective of the study, the generated findings were compared against Coxhead's AWL. MV were excluded from this analysis. Meanwhile, the *-ing* participle of the lexical verb, infinitive, past participle of the lexical verb, and *-s* form of the lexical verb were analysed in their base forms, as presented in Table 4.

Table 4. EVT Inside Coxhead's AWL

Language		Literature		Education	
assess	1	appropriate	2	affect	2
assume	1	conduct	2	assess	1
attach	6	demonstrate	3	assume	1
code	4	display	6	benefit	1
create	1	emerge	4	clarify	8
define	1	encounter	10	conduct	2
eliminate	7	expose	5	construct	2
expose	5	focus	2	contribute	3
focus	2	indicate	1	create	1
function	1	investigate	4	design	2
grant	4	invoke	10	emphasize	3
indicate	1	locate	3	enhance	6
investigate	4	rely	3	estimate	1
require	1			expand	5
resource	2				
restrict	2				
reveal	6				
undergo	10				

Table 4 shows that EVT in the social and humanities academic text derived from COCA, especially language, literature, and education, appeared in Coxhead AWL from 1, the most frequent sublist, to 10, the least frequent sublist. The data implied that EVT in language appeared more, followed by education and literature than the AWL. Further, EVTs that did not appear in the AWL were compiled into the new social and humanities academic word list, as presented in Table 5.

Table 5. EVT Outside Coxhead's AWL

Language	Literature	Education
absorb	add	access
adopt	adopt	address

allow	begin	adopt
appear	call	argue
ask	characterize	base
base	close	change
believe	compose	claim
coach	consider	complete
complete	contain	criticize
consider	demand	cross
create	detach	customize
discuss	discuss	delineate
distinguish	dismiss	design
explain	encounter	develop
express	entitle	distinguish
frame	examine	draw
illuminate	exists	encourage
improve	experience	fail
instil	expose	follow
need	fill	get
perform	find	give
rate	focus	help
receive	gain	hold
reduce	generalize	indicate
refer	give	inform
reflect	grow	introduce
relate	help	know
represent	include	learn
review	inform	link
see	know	make
see	master	market
solve	measure	misplace
suggest	observe	need

take	provide	note
teach	raise	observe
transcribe	read	offer
try	recognize	provide
understand	refer	pursue
use	remain	raise
using	see	receive
validate	send	refer
	specialize	reject
	suggest	replace
	support	resolve
	take	see
	work	serve
		suggest
		take
		transfer
		use
		work

Table 5 shows that EVT outside of Coxhead's AWL was extensive and has the potential to be developed further. Contrary to EVT, which appeared in Coxhead's AWL, this data demonstrated that education had new and potential EVT, followed by literature and language.

3.2 Discussion

The increasing demand for publications in English [1], [2], [3], [4] affects authors from L2 background [3], [15] to publish their work and to be recognised among the academic community [10], [11], [12], [13] using this language. The existing literature suggests that social and humanities papers are also predominantly published in English [8], motivating the present study to investigate academic word lists useful for assisting L2 authors in writing their publications and helping them from struggling. Moreover, the literature describes that comprehending the language's formal features [16] to ensure grammar accuracy is crucially required in the publications [3].

The language features crucial for academic writing are EVT and MV. The literature describes that identifying correct EVT is crucial to writing tenses that adhere to the citation style [19] and highly reputable journals [20]. Furthermore, MV is pivotal in amplifying the implications made in the academic text [24]. Besides, the literature suggests that MV frequently appears in the social and

humanities academic texts than in other fields of science [25]. Accordingly, retrieving data from social and humanities is expected to depict EVT forms, how they appear in the existing AWL, and how they could become new social and humanities academic word lists, which can be employed to help L2 authors overcome their language constraints.

The present study demonstrates that EVT retrieved from COCA appears in Coxhead's AWL, ranging from the highest to the lowest sublist, which has two implications. First, it indicates that COCA is an adequate source for identifying verbs from academic text that complement the previous study about tenses distributions in Elsevier journal article abstracts [20] since the EVT found in the present study adheres to tenses used in reputable journals. Second, from the findings, it can be denoted that EVT from the present study appeared in the form of tenses advised tense by the existing literature, i.e., present tense, past tense, and past perfect tense [19]. Consequently, the generated EVT from COCA can be enacted to supplement the literature. When L2 authors write about these tenses, the words collected from this study can be the core words to be used and taught to enrich academic writing. Accordingly, L2 authors can construct meaningful and logical words [18], [32] in social and humanities papers.

Further, the existing literature describes MV as the means to strengthen hypothesis and interpretation [24] that frequently appear in social and humanities [25], indicating two propositions. First, COCA can be utilised to retrieve types of MV in academic writing besides academic essays and journal articles. Second, considering its crucial role in emphasising the L2 author's voice in academic writing, the study depicts that all social and humanities fields of study contain MV, with slightly distinctive frequency and types. It implies that the authors also attempt to display their voices to the readers in this field of study.

Ultimately, the frequency of EVT and MV from three social and humanities fields of study varies. It implies that each field is unique and could have different language styles and a tendency to use particular academic and modal verbs. Accordingly, the findings can be used to prioritise EVT and MV, which could be taught to L2 authors depending on their field of study. Besides, from a practical perspective, the study contributes to the implementation of EVT and MV in academic writing papers, especially for L2 authors. Further, EVT, whether included or excluded from Coxhead's AWL, is paramount to be applied in academic writing classes. For the sake of academic writing, teachers, instructors, or professors can employ these findings to be added to prepare teaching materials.

4 Conclusion

The present study offers an insight into the body of literature, especially corpus linguistics. Further, the study reveals that the data retrieved from COCA depict the appearance of EVT and MV in the social and humanities fields. They are the base form of the lexical verb, past tense of the lexical verb, *-ing* participle of the lexical verb, infinitive, past participle of the lexical verb, *-s* form of the lexical verb, and modal verb with varied frequencies for each item. The generated EVT from COCA also appear in Coxhead AWL and the new social and humanities academic word list, indicating significant potential for future development, such as teaching materials and dictionaries. Furthermore, this study also has some limitations that require attention in future research. Firstly,

the study aims to investigate EVT and MV. Consequently, the findings do not apply to other linguistics features, which demands another investigation. Secondly, the present study only compares EVT with AWL. Hence, there is an opportunity to compare EVT from COCA with other specialised academic word lists proposed by other scholars. Thirdly, the present study employs COCA. Therefore, many other corpora, such as the British National Corpus (BNC) and or specialised corpus, have the potential to be explored.

References

[1] J. Flowerdew and Y. Li, "English or Chinese? The trade-off between local and international publication among Chinese academics in the humanities and social sciences," *J Second Lang Writ*, vol. 18, no. 1, pp. 1–16, 2009, doi: 10.1016/j.jslw.2008.09.005.

[2] C. Gnutzmann and F. Rabe, "'Theoretical subtleties' or 'text modules'? German researchers' language demands and attitudes across disciplinary cultures," *J Engl Acad Purp*, vol. 13, no. 1, pp. 31–40, 2014, doi: 10.1016/j.jeap.2013.10.003.

[3] A. Mauranen, N. Hynninen, and E. Ranta, "English as an academic lingua franca: The ELFA project," *English for Specific Purposes*, vol. 29, no. 3, pp. 183–190, 2010, doi: 10.1016/j.esp.2009.10.001.

[4] U. Römer, "English in Academia: Does Nativeness Matter?," *Anglistik: International Journal of English Studies*, vol. 20, no. 2, pp. 89–100, 2009.

[5] W. Baker, "English as an academic lingua franca and intercultural awareness: Student mobility in the transcultural university," *Language and Intercultural Communication*, vol. 16, no. 3, pp. 437–451, 2016, doi: 10.1080/14708477.2016.1168053.

[6] R. Phillipson, "Lingua franca or lingua frankensteinia? English in European integration and globalisation," *World Englishes*, vol. 27, no. 2, pp. 250–267, 2008.

[7] S. Melo-Pfeifer, "Between multilingual ethos and English pathos: how do multilingual scholars navigate international academia?," *Language and Education*, vol. 38, no. 1, pp. 80–96, 2023, doi: 10.1080/09500782.2023.2244928.

[8] J. C. Huang, "Attitudes of Taiwanese scholars toward English and Chinese as languages of publication," *Asia Pacific Journal of Education*, vol. 31, no. 2, pp. 115–128, 2011, doi: 10.1080/02188791.2011.566983.

[9] R. E. Hamel, "The dominance of English in the international scientific periodical literature and the future of language use in science," *AILA Review*, vol. 20, no. August, pp. 53–71, 2007, doi: 10.1075/aila.20.06ham.

[10] M. J. Curry and T. Lillis, "Multilingual Scholars and the Imperative to Publish in English: Negotiating Interests, Demands, and Rewards," *Tesol Quarterly*, vol. 38, no. 4, pp. 663–688, 2004.

[11] Md. O. Hamid, "English Teachers' Choice of Language for Publication: Qualitative Insights from Bangladesh," *Current Issues in Language Planning*, vol. 7, no. 1, pp. 126–140, 2006, doi: 10.2167/cilp090.0.

[12] M. Petersen and P. Shaw, "Language and disciplinary differences in a biliterate context," *World Englishes*, vol. 21, no. 3, pp. 357–374, 2002, doi: 10.1111/1467-971X.00255.

[13] J. Schluer, "Writing for publication in linguistics: Exploring niches of multilingual publishing among German linguists," *J Engl Acad Purp*, vol. 16, pp. 1–13, 2014, doi: 10.1016/j.jeap.2014.06.001.

[14] F. Pan, "A multi-dimensional comparison of L1 and L2 medical research articles," *Southern African Linguistics and Applied Language Studies*, 2023, doi: 10.2989/16073614.2023.2249040.

[15] C. Pérez-Llantada, "Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage," *J Engl Acad Purp*, vol. 14, pp. 84–94, 2014, doi: 10.1016/j.jeap.2014.01.002.

[16] A. Oshima and A. Hogue, *Introduction to Academic Writing: Second Edition*, 3rd Ed. New York: Pearson Education, 2007.

[17] F. Romano, "Grammatical accuracy in EAP writing," *J Engl Acad Purp*, vol. 41, p. 100773, 2019, doi: 10.1016/j.jeap.2019.100773.

[18] Y. Deng, D. Liu, and S. Wu, "Academic English verbs across disciplines: A corpus study and its implications," *Ampersand*, vol. 9, no. August, p. 100093, 2022, doi: 10.1016/j.amper.2022.100093.

[19] A. C. Winkler and J. R. Metherell, *Writing the Research Paper: A Handbook*, 8th ed. Boston: Wadsworth, 2010.

[20] D. A. Kwary, A. Kirana, and A. F. Artha, "The distribution of verb tenses and modals in journal articles' abstracts," *Southern African Linguistics and Applied Language Studies*, vol. 35, no. 3, pp. 229–244, 2017, doi: 10.2989/16073614.2017.1373366.

[21] E. Hinkel, "The effects of essay topics on modal verb uses in L1 and L2 academic writing," *J Pragmat*, vol. 41, no. 4, pp. 667–683, 2009, doi: 10.1016/j.pragma.2008.09.029.

[22] L. X. Li, "Developmental patterns of English modal verbs in the writings of Chinese learners of English: a corpus-based approach," *Cogent Education*, vol. 9, no. 1, 2022, doi: 10.1080/2331186X.2022.2050457.

[23] C. L. Liardét, "'As we all know': Examining Chinese EFL learners' use of interpersonal grammatical metaphor in academic writing," *English for Specific Purposes*, vol. 50, pp. 64–80, 2018, doi: 10.1016/j.esp.2017.11.005.

[24] A. de Waard and H. Pander Maat, "Verb form indicates discourse segment type in biological research papers: Experimental evidence," *J Engl Acad Purp*, vol. 11, no. 4, pp. 357–366, 2012, doi: 10.1016/j.jeap.2012.06.002.

[25] T. D. Hardjanto, "Hedging Through the Use of Modal Auxiliaries," *Humaniora*, vol. 28, no. 1, pp. 37–50, 2016.

[26] S. Alghazo, M. N. Al Salem, I. Alrashdan, and G. Rabab'ah, "Grammatical devices of stance in written academic English," *Heliyon*, vol. 7, no. 11, pp. 1–8, 2021, doi: 10.1016/j.heliyon.2021.e08463.

[27] P. Baker, "Corpus Methods in Linguistics," in *Research Methods in Linguistics*, L. Litosseliti, Ed., London: Continuum, 2010, pp. 93–113.

[28] D. Biber, "Corpus linguistics and the study of literature," *Sci Study Lit*, vol. 1, no. 1, pp. 15–23, 2011, doi: 10.1075/ssol.1.1.02bib.

[29] H. Lindquist, "Corpus linguistics and the description of English," *Corpus Linguistics and the Description of English*, pp. 1–219, Dec. 2009, doi: 10.2478/ICAME-2020-0006.

[30] T. McEnery and A. Hardie, *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press, 2011.

[31] A. Coxhead, "A New Academic Word List," *TESOL Quarterly*, vol. 34, no. 2, 2000.

[32] H. Ma and M. Qian, "The creation and evaluation of a grammar pattern list for the most frequent academic verbs," *English for Specific Purposes*, vol. 58, pp. 155–169, 2020, doi: 10.1016/j.esp.2020.01.002.

[33] "English Corpora.org." Accessed: Jul. 26, 2023. [Online]. Available: https://www.english-corpora.org/

[34] D. A. Kwary, D. Ratri, and A. F. Artha, "Lexical bundles in journal articles across academic disciplines," *Indonesian Journal of Applied Linguistics*, vol. 7, no. 1, pp. 132–140, 2017, doi: 10.17509/ijal.v7i1.6866.

[35] D. A. Kwary, "A corpus platform of Indonesian academic language," *SoftwareX*, vol. 9, pp. 102–106, 2019, doi: 10.1016/j.softx.2019.01.011.

[36] UCREL Lancaster University, "Free CLAWS web tagger." Accessed: Oct. 11, 2022. [Online]. Available: http://ucrel-api.lancaster.ac.uk/claws/free.html

[37] UCREL Lancaster University, "CLAWS part-of-speech tagger." Accessed: Oct. 11, 2022. [Online]. Available: https://ucrel.lancs.ac.uk/claws/