



Causality in English Academic Writing: A Case of Research Articles in Applied Linguistics and Physical Chemistry*

Sun-Young Oh (Seoul National University)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: Oct. 01, 2021

Revised: Jan. 22, 2022

Accepted: Jan. 27, 2022

Sun-Young Oh
Professor, Department of
English Language Education,
Seoul National University
Tel: 02) 880-7675
Email: sunoh@snu.ac.kr

* I would like to thank Ju Young Min for her help with data retrieval and coding and two anonymous reviewers for useful suggestions on an earlier version of this article.

ABSTRACT

Oh, Sun-Young. 2022. Causality in English academic writing: A case of research articles in applied linguistics and physical chemistry. *Korean Journal of English Language and Linguistics* 22, 40-54.

This study examines how causality is conveyed in English academic writing, with a special focus on the comparison between two contrasting disciplines, i.e., applied linguistics (AL) and physical chemistry (PC). Two academic corpora were compiled with research articles from each discipline, and a total of 135 explicit causative devices were analyzed for their frequency and use in each corpus and compared with the finding from a corpus of general written English (Xuelan and Kennedy 1992, Expressing causation in written English. *RELC Journal* 23(1), 62-80). The results indicate that frequent representation of the relation of cause and effect is one of the defining characteristics of academic prose, irrespective of disciplines. Another common feature of English academic writing, in contrast to general written English, regarding the expression of causality was the heavy reliance on the nominal category. A careful inspection of the data revealed subtle differences between the two disciplines including some preferred causality markers and their divergent phraseology, which are associated with distinctive epistemic conventions of each discipline. These findings are discussed in terms of the nature of academic writing and of hard versus soft disciplines, with some pedagogical implications drawn for English for academic purposes.

KEYWORDS

causality, cause, effect, academic writing, corpus, EAP

1. Introduction

Causality, which refers to the relationship between cause and effect, pervades our daily life in many ways. Leucippus, a Greek philosopher in the 5th century, B.C., is known to suggest that “there is causality in nature” (Zimmerman 1989: 35), and it is “as fundamental as that of time and space” (Bridgman 1927/1961: 80, cited in Darian 2003). It is no wonder that this notion has been explored for an extended period of time in diverse fields of study including philosophy (e.g., Hume 1740/1965, Mackie 1980, Mill 1872/1973, Rescher 1970), psychology (e.g., Cheng and Novick 1992, Downing et al. 1985, Kelley 1973, Van Overwalle and Heylighen 1995), as well as linguistics (e.g., Greenbaum 1969, Halliday and Hasan 1976, Martin 1992, Rutherford 1970, Winter 1982). By definition, causality refers to the relation of “two events or states of affairs if one is understood as the cause of or reason for the other” (Altenberg 1984: 20). Thus, the notion of causality inherently involves two elements, a cause/reason and an effect/result (Xuelan and Kennedy 1992).

One genre of text in which the relation of cause and effect tends to be highly expected is academic writing. Academic writing typically consists of a number of rhetorical functions that occur at different stages of discourse. Representative functions that researchers have identified are numerous including definition, description, exemplification, classification, observation, hypothesizing, prediction, argumentation, comparison/contrast, and cause-effect (Darian 2003, Gilquin and Paquot 2007, Murray and Hughes 2008, Zimmerman 1989). Among these many functions, causality figures prominently, playing a critical role in academic thinking and writing where it is necessary to provide reasons and justifications for methodological decisions, interpretation of observations and/or findings, and claims that are made based on such results. Despite the significance of causality in academic writing, however, it has not received enough attention in the field of English for academic purposes (EAP). Although some information is available as to the variety and frequency of causative devices used in written English (e.g., Altenberg 1984, Xuelan and Kennedy 1992), the concept of cause and effect has rarely been focused upon in EAP literature. Given that the topic of causality “presents many lexical, syntactic, and rhetorical problems” (Darian 2003: 113) for student readers and writers of academic writing, however, it deserves close attention. In addition, it is yet unknown whether there exist any differences in the construction of causality according to disciplines. Research has identified many linguistic and rhetorical disparities across academic disciplines (Becher and Trowler 2001, Cortes 2004, Hyland 2001, 2002, 2004, 2005, 2006, 2015); this suggests the possibility that causality may also be expressed in ways that reflect the epistemological nature of a discipline.

The present study aims to explore the ways in which the concept of causality is realized in English academic writing, focusing on the sub-genre of research articles. In particular, it represents the first attempt to explore the possibility of disciplinary difference in this regard. Given that the literature highlights the distinction between the “harder” and the “softer” sciences in disciplinary variations (Coffin et al. 2003, MacDonald 1994), the study has chosen to compare two significantly divergent disciplines, one from the hard science (physical chemistry) and the other from the soft (applied linguistics). The selection of the two disciplines was informed by “the traditional distinctions between the sciences and engineering as hard knowledge, and the social sciences/humanities as soft disciplines” (Hyland 2004: 29), and also motivated by the interest in the researcher’s own discipline (applied linguistics) and the availability of the experts who can guide the corpus compilation (physical chemistry). The findings of the current study will hopefully add to the existing knowledge on the issue of causality and the characteristics of academic writing, which can help one understand and learn how to convey causality in research articles in ways that are approved and accepted by their disciplinary communities.

2. Literature Review

Traditionally, causality in the English language has been dealt with incidentally in the study of grammar (e.g., Martin 1992, Winter 1982) and cohesion (Halliday and Hasan 1976) or in the investigation of adverbials (e.g., Greenbaum 1969) or subordinate clauses (Rutherford 1970). The first study, to the researcher’s knowledge, which

zoomed in on the concept of causality is Altenberg (1984). He sought to catalog an extensive list of linguistic devices which signal cause and effect in English. Remarkably for the time, Altenberg's (1984) inventory is based upon empirical analysis of authentic data of both spoken and written corpora¹ and includes four main categories: adverbials (e.g., *so, therefore*), prepositions (e.g., *because of, due to*), subordination (e.g., *as, because*), and clausal linkage (e.g., *that's why..., the result was...*).

Another long catalog of English causality markers is provided by Xuelan and Kennedy (1992), who compiled a total of 130 devices from various sources such as standard reference books, English teaching syllabi, and course books. These causative devices were classified into eight grammatical types (i.e., adjective phrases, adverbs, complex prepositions, conjunctions, nouns, prepositions, verbs, and verb phrases). The researchers also reported the frequencies of the causative devices in the one million-word Lancaster-Oslo/Bergen corpus of modern English (LOB). According to their findings, conjunctions and adverbs were the most frequently used categories. Another interesting finding of theirs is that cause/reason is represented more often than effect/result in written English.

The lexical causative devices identified in the aforementioned two studies may be divided into two types. While "regular (traditional) markers" always signal cause (e.g., *because of, causes, due to, results from, as a result of*) and effect or result (e.g., *accordingly, as a result, consequently, for this reason, hence, thus, therefore*), "special cause-and-effect markers" such as *when, since, by, if, and so* may indicate cause and effect but may also convey different relationships (Darian 2003). Causal relations, however, are not signaled in such explicit ways only. There also exist implicit ways of expressing cause and effect, which employ implicit causative verbs, elliptical syntactic patterns, and juxtaposition of phrases, clauses, and sentences (Xuelan and Kennedy, 1992).

More recently, researchers have turned their attention to specific registers or genres in exploring the expression of causality, such as newspapers (e.g., Khoo 1995), science textbooks (e.g., Darian 1996), academic writing textbooks (e.g., Moreno 2003), and history writing (e.g., Achugar and Schleppegrell 2005, Coffin 2004, Martin 2002, Veel and Coffin 1996). Interests have also grown in the differences between English learners and native speakers in the use of causality markers (Flowerdew 1998, Oh 2009, Reynolds 2002). While some of these studies have examined the representation of causality in English for academic purposes in a broad sense, dealing with texts from the discipline of science or history, for example, there has been no systematic attempt so far to investigate the topic with research articles, "the central genre of knowledge production," (Ruiying and Allison 2003) as the data. A study that is the closest to this concern may be Charles (2011). Using corpus techniques with discourse analysis, she investigated the frequency, phraseology, and functions of adverbials of result (*thus, therefore, then, hence, so, and consequently*) in the two corpora of native-speaker theses in politics and materials science. She found distinctive styles of use with each adverbial and accounted for the differences in terms of genre, discipline, context, and function. The study, however, is quite limited in scope, including only six adverbials of result from among a vast array of English causative devices. Besides, theses belong to "the genre of student writing" (Hüttner 2008), with a different purpose, audience, and organizational patterning from research articles.

In view of the increased interest in causality within particular genres and its pivotal role in academic writing, it deserves greater attention in the area of EAP. It is not clear yet whether, and how, the expression of causality in academic writing, especially the genre of research articles, differs from non-academic writing. In addition, one of the primary concerns in EAP, the disciplinary variations, may be worth being explored along with causality. It would be intriguing to observe how the epistemological nature of different disciplines interacts with the concept of causality. To fill these gaps, the present study examines the use of a total of 135 English causative devices in two EAP corpora in comparison with the general written English represented by LOB (Xuelan and Kennedy 1992). The followings are two research questions that the study aims to provide answers to:

- (1) How is causality conveyed in research articles in terms of the frequency and the distribution of grammatical categories of the causative devices and the ratio of cause versus effect?

¹ The corpora consist of 20 conversations (of c 5,000 words each) taken from London-Lund Corpus of Spoken English (LLC), and 50 texts (of c 2,000 words each) from Lancaster-Oslo/Bergen corpus of modern English (LOB).

- (2) Is causality expressed differently in research articles of applied linguistics and physical chemistry in terms of preferred causality devices and their use in context?

3. Method

3.1 Corpus Compilation

To investigate the expression of causality in academic writing and compare it in different disciplines, two corpora were compiled of academic journal articles in the fields of applied linguistics (AL) and physical chemistry (PC). Ten representative journals were selected in each discipline based on the impact factor and consultation with two experts.² From each journal 10 and 17 articles³ in AL and PC, respectively, which were published in the years between 2007 and 2011, have been drawn through stratified random selection. Downloaded texts were preprocessed, through which figures, captions for figures and tables, reference lists, acknowledgments, keywords, and other peripheral languages (such as contact information, affiliation, and address of the authors) were all deleted. The AL corpus and the PC corpus respectively contain 100 and 169 articles, both containing approximately 815,000 words. The composition of the two academic corpora is shown in Table 1:

Table 1. The Academic Corpora

Corpus	Number of texts	Number of words
Applied Linguistics (AL)	100	815,182
Physical Chemistry (PC)	169	814,554

3.2 Analysis

To examine the construction of causality in academic writing, the current study made use of 135 causality devices drawn from previous studies (Flowerdew 1998, Xuelan and Kennedy 1992).⁴ They are categorized into eight different grammatical types: adverbs, adjective phrases, conjunctions, complex prepositions, noun phrases, prepositions, verbs, and verb phrases (see Table 2).

² AL Journals thus selected are *Studies in Second Language Acquisition*, *System*, *Journal of Second Language Writing*, *Applied Linguistics*, *Language Learning*, *Language Learning & Technology*, *TESOL Quarterly*, *ESP*, *Second Language Research*. PC journals are *Journal of Chemical Physics*, *Journal of Physical Chemistry B*, *Phys Chem Chem Phys*, *ChemPhysChem*, *Chemical Physics*, *Chemical Physics Letters*, *Faraday Discussions*, *Langmuir*, *Molecular Physics*, *JPC A*.

³ Initially, 10 articles were selected from each journal in both disciplines, but it was shortly found that the PC articles are much shorter than the AL articles, which rendered the two corpora incomparable in size. Thus, the decision was made to include more articles from PC in order to match the number of words of the two corpora as closely as possible. Later, one PC article was removed out of a total of 170 due to the same concern, i.e., to maximize the comparability of the two corpora in terms of size.

⁴ The 135 markers consist of 130 cause/effect devices identified by Xuelan and Kennedy (1992) and additional 5 devices from Flowerdew (1998).

Table 2. Target Linguistic Devices Expressing Causality

Grammatical Classes	Markers of cause/reason and result/effect
Nouns	reason(s), cause(s), source(s), root(s), mainspring, the whys and wherefores, contributor to, inspiration, <i>effect(s)</i> , <i>result(s)</i> , <i>consequence(s)</i> , <i>outcome(s)</i> , <i>product(s)</i> of, <i>corollary</i> , <i>aftermath</i> , <i>upshot</i>
Conjunctions	because, since, as, now that, inasmuch as, seeing that, seeing as, <i>so that</i> , <i>so...that</i> , <i>such...that</i>
Complex prepositions	due to, because of, in (the) light of, in view of, given, for (that) reason, owing to, thanks to, on account of, on the ground(s) that, by reason of, by virtue of, what with, as a matter of, out of, on the ground(s) of, on the strength of, for reasons of, in consideration of, on that account, on that score, at the bottom of, from reasons of, by courtesy of, <i>as a result of</i> , <i>in consequence of</i> , <i>with the result that</i> , <i>as a consequence of</i> , <i>with the consequence that</i> , <i>by consequence</i>
Prepositions	with, by + noun, by + ing, through, from, for, under
Verbs	cause, produce, bring, inspire, provoke, generate, create, induce, present, pose, arouse, underlie, trigger, give, prompt, incur, raise, compel, engender, yield, evoke, motivate, occasion, precipitate, rouse, stir, incite, awaken, spark, beget, breed, foment, spawn
Verb phrases	result in, contribute to, lead to, bring about, give rise to, make for, account for, bring on, stir up, spark off, give occasion to, <i>result from</i> , <i>come from</i> , <i>arise from</i> , <i>attribute to</i> , <i>ascribe to</i> , <i>derive from</i> , <i>arise out of</i> , <i>put down to</i> , <i>spring from</i> , <i>emerge from</i> , <i>stem from</i> , <i>proceed from</i> , <i>spring out of</i>
Adjective Phrases	responsible for, <i>consequent on/upon</i> , <i>consequential to</i>
Adverbs	why, <i>so</i> , <i>therefore</i> , <i>hence</i> , <i>consequently</i> , <i>then</i> , <i>thus</i> , <i>as a result</i> , <i>as a consequence</i> , <i>thereby</i> , <i>accordingly</i> , <i>in consequence</i>

The steps taken for analysis are as follows. First, the two corpora were searched for each of the 135 target causative devices using the WordSmith (version 5) concordancing program. All the inflectional forms of the devices (e.g., plural forms of nouns, various tense/aspectual forms of verbs) were included in the search. Unlike the unambiguous or “regular” (Darian 2003) causative markers, which express causation irrespective of context (e.g., *outcome*, *because (of)*, *therefore*, *as a result (of)*, *result from*), ambiguous or “special” ones (e.g., *produce*, *through*, *as*, *accordingly*) needed to be examined one by one to include only those which signal causal relations. The judgment was made by two coders including the researcher and guided by the criteria established by Xuelan and Kennedy (1992), according to which items were considered as expressing causality when they answered the following questions, “Why has something happened, is happening or might happen?”, “What caused or causes what (to happen)?” and/or when they are replaceable with unambiguous causative markers. The intercoder reliability calculated on data from the AL corpus was 88%.

For some devices of which the frequency exceeds the pre-specified limit, which was 200 (e.g., *since*, *as*, *given*, *by*, *through*, *for*, *present*, *effect*, *result*, *so*, *then*), 200 tokens were chosen for examination through random selection, and the results were extended to the entire tokens. For items that occur as more than one part of speech (POS) (e.g., *cause*), concordance lists were generated separately according to each POS. Corpora were POS-tagged for this purpose.

The next step was to compute the frequency of occurrences for each causative marker within the two corpora. All the causative devices were categorized into eight grammatical classes and the distribution of these classes, as well as the total frequencies, were compared between the two corpora as well as with LOB (Xuelan and Kennedy 1992). In addition, the devices were classified either into cause/reason or effect/result types based upon Xuelan and Kennedy (1992) (see Table 2 for the classification) and the ratio between the two types was compared across the corpora. To estimate the statistical significance of differences in frequency across corpora, chi-square tests were run wherever relevant. In addition, Cramer’s V was calculated as the measure of effect size. Finally, preferred causative markers (i.e., top 10 most frequent devices) were identified and compared between the two academic corpora, and their concordances were scrutinized in full context to reveal discipline-specific characteristics of lexico-grammatical (e.g., collocations, formulaic phrases) and discourse (e.g., organizational and rhetorical practices) patterns.

4. Results and Discussion

4.1. Causality in Academic Writing

4.1.1 Frequency and distribution of grammatical categories of causality devices

The first step of comparing the expression of causality in academic versus general English writing was to compare the frequencies of causality markers both in total and according to grammatical categories. Table 3 below presents the raw frequencies of occurrence and the relative percentages of the eight grammatical classes of causative devices in the two corpora under study (i.e., AL and PC) in comparison with the corresponding information from the LOB corpus (Xuelan and Kennedy 1992). Table 4 shows the results of chi-squared tests performed to check the significance of differences in frequency between the corpora.

Table 3. Distribution of Causality Markers according to Grammatical Classes

Grammatical Classes	AL		PC		LOB	
	frequency	percentage	frequency	percentage	frequency	percentage
Nouns	3,863	43.1%	2,761	32.3%	930	15.9%
Conjunctions	1,303	14.5%	1,130	13.2%	1,683	28.7%
Prepositions	220	2.5%	112	1.3%	165	2.8%
Complex Prepositions	705	7.9%	1,051	12.3%	634	10.8%
Verbs	456	5.1%	655	7.7%	501	8.5%
Verb Phrases	821	9.2%	1,167	13.6%	294	5%
Adverbs	1,585	17.7%	1,619	18.9%	1,643	28%
Adjective Phrases	18	0.2%	58	0.7%	11	0.2%
Total Causality Markers	8,971	100.0%	8,553	100.0%	5,861	100.0%
Total Words in Corpus	815,182		814,554		1,000,000	
Density	1.10		1.05		0.59	

Table 4. Results of Chi-square Tests with Effect Sizes

	Academic vs General Corpora			AL vs PC		
	χ^2	<i>p</i> -value	Cramer's <i>V</i>	χ^2	<i>p</i> -value	Cramer's <i>V</i>
Nouns	966.08	.000	.203**	216.41	.000	.111*
Conjunctions	666.15	.000	.169*	6.31	.012	.019
Prepositions	17.90	.000	.028	30.77	.000	.042
Complex prepositions	3.04	.081	.011	95.28	.000	.074
Verbs	33.37	.000	.038	48.90	.000	.053
Verb phrases	199.73	.000	.092	87.87	.000	.071
Adverbs	254.07	.000	.104*	4.66	.031	.016
Adjective phrases	7.17	.007	.018	23.12	.000	.036

** medium effect, * small effect

Table 3 reveals some significant disparities between the academic writing corpora (i.e., AL and PC) and the general written (British) English corpus (i.e., LOB). Above all, the overall frequencies of the causative devices are much higher (almost twice as high) in the former than the latter (1.10 and 1.05 versus 0.59 in density). The difference in frequency between these two types of corpora was statistically significant, as indicated by the results

of the chi-square test ($\chi^2(1) = 12970.17, p < .001$, Cramer's $V = .119$). This finding confirms the hypothesis that the relation of cause and effect figures much more frequently and significantly in academic writing. Second, the rank order of the grammatical classes of the causative markers is comparable in the two academic corpora but dissimilar to that of the LOB. In the academic corpora, nouns rank the top, comprising an average of 43.1% of the total tokens in the AL and 32.3% in the PC. The next most frequent are adverbs (17.7% and 18.9%), which are preferred to conjunctions (14.5% and 13.2%) and verb phrases (9.2% and 13.6%), followed by the remaining three other classes (7.9% and 12.3% for complex prepositions, 5.1% and 7.7% for verbs, and 2.5% and 1.3% for prepositions). In the LOB corpus, in contrast, conjunctions are the most frequent grammatical class (28.7%), closely followed by adverbs (28%). Nouns rank third in the LOB, accounting for 15.9%, and the other three classes together explain 37.3%. Compared with the non-academic writers, the research article authors thus rely more on nouns in expressing causality while they make less frequent use of conjunctions or adverbs. The differences between academic and general corpora in the frequency of these three categories were statistically significant, with small to medium overall effect: nouns ($\chi^2(1) = 966.08, p < .001$, Cramer's $V = .203$), conjunctions ($\chi^2(1) = 666.15, p < .001$, Cramer's $V = .169$), and adverbs ($\chi^2(1) = 254.07, p < .001$, Cramer's $V = .104$). The prevalence of nominal causality markers ratifies the characteristics of academic writing that favor nominalization, a finding reported by several previous studies (Biber, Conrad and Reppen 1998, Bilig 2008, Liardét 2013, Ryshina-Pankova 2010).⁵ Nominalization has been regarded as a lexico-grammatical realization of "grammatical metaphor," which is a critical tool in academic writing (Biber and Gray 2013, Halliday 1994). Verb phrases are often used to convey cause and effect relations in the academic corpora (9.2% and 13.6%) but occur less frequently in the LOB corpus (5.0%). Although the result of the chi-square test on this difference was statistically significant with a p-value of .000, it did not have enough practical significance, as indicated by the effect size (Cramer's $V = .092$). The differences between academic and general corpora in frequency of the other categories, i.e., (complex) prepositions, verbs, and adjective phrases, are relatively small and not significant (see Table 4).

Compared with the discrepancies between academic and non-academic writing, the differences between disciplines within the academic discourse, i.e., between AL and PC, seem to be relatively small, at least in terms of the grammatical classes employed to mark causality as well as the total frequencies of the devices. The most noticeable divergence concerns significantly greater employment of nouns in the AL (43.1%) than in the PC (32.3%) ($\chi^2 = 216.41(1), p < .01$, Cramer's $V = .111$) although the latter is still a larger amount compared with the LOB (15.9%), as discussed above. In contrast, the categories of complex prepositions (12.3% vs. 7.9%) and verbs (phrases) (21.3% vs. 14.3%) are somewhat more frequent in the PC than in the AL albeit not with practical significance (see Table 4 for details). These patterns of differences and similarities are illustrated in Figure 1 below (The frequencies are normalized to 100,000 words for easier comparison).

⁵ For example, verbs and adjectives such as *develop* and *able* are often nominalized in academic texts into *development* and *ability*, respectively.

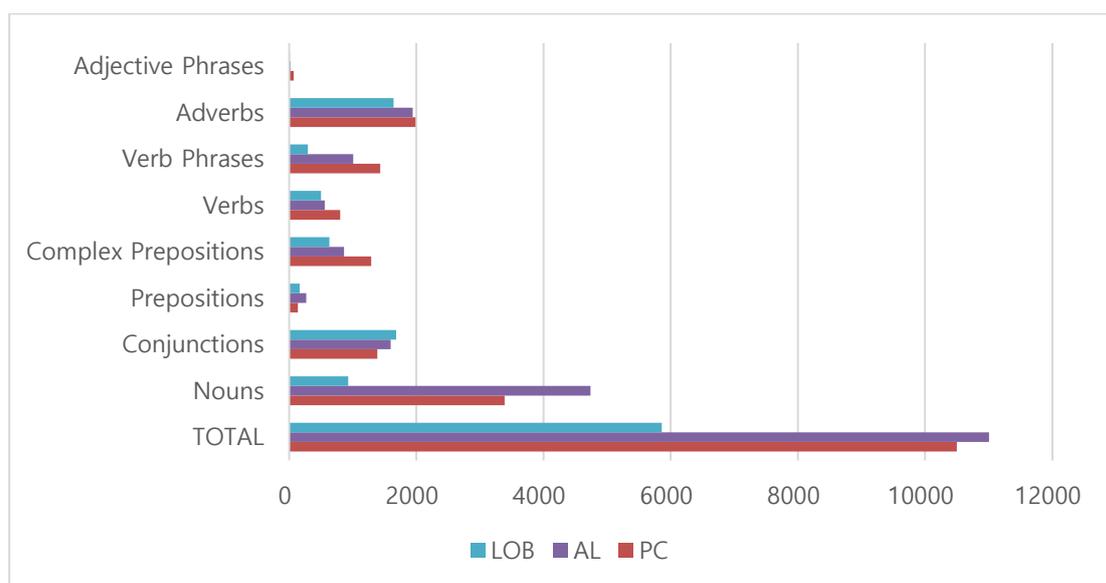


Figure 1. Grammatical Categories of Causality Markers across Corpora (per 100,000 words)

The current finding regarding the distribution of grammatical categories in academic writing is consistent with Biber and Gray (2016), who compared major word classes across the three registers of novels, newspaper prose, and academic prose. They found that nouns (and adjectives) are the most frequent in academic prose whereas verbs (and adverbs) are most common in fiction, with an in-between tendency found for newspaper prose. As noted, both of the two academic writing corpora in the present study similarly showed the strongest reliance on nouns in expressing causality, albeit to a differing degree. To be specific, nouns were far more frequent in the AL, which is a type of soft discipline, whereas verbs and verb phrases were more common in the PC, a hard discipline. This finding in fact is in line with a recent study by Kaidan et al. (2021), who investigated the use of nominalization in the corpus of research articles in physics and applied linguistics. According to the researchers, nominalization, which is predominant in academic writing, exhibits disciplinary disparities. Specifically, they found a higher frequency of nominalization in applied linguistics research articles than in physics counterparts, and attributed the finding to “the more abstract nature of discourse in this [applied linguistics] field as an instance of soft fields and the tendency among writers to create abstraction by maintaining conciseness in their respective discourse” (*ibid*: 13). In addition, they note the difference in “the nature of the two disciplines, with applied linguistics dealing with more abstract topics ... than physics” (*ibid*: 9). This observation also appears to hold for the current study examining a comparable database. The disciplinary variation in the expression of causality will further be dealt with in Section 4.2.

4.1.2 Distribution of cause/reason versus effect/result devices

Inspired by a previous study that reported the relative proportion of cause/reason versus effect/result in the general written English (Xuelan and Kennedy 1992), the current study also examined the relative percentages of these two different types of causality devices in the two academic corpora (i.e., AL and PC). The results are presented in Table 5, in parallel with Xuelan and Kennedy’s (1992) finding based on LOB.

Table 5. Ratio of Cause/Reason versus Effect/Result Markers

Types	AL		PC		LOB	
	frequency	percentage	frequency	percentage	frequency	percentage
Cause/reason	3,734	41.6%	3,940	46.1%	925	71.7%
Effect/result	5,237	58.4%	4,613	53.9%	364	28.3%
Total	8,971	100.0%	8,553	100.0%	1,289	100.0%

As can be easily noticed, the two academic corpora, AL and PC, are not very different from each other in the relative proportions of the two kinds of causality markers. In both of the corpora, markers of effect/result account for over half of the total (58.4% and 53.9% in AL and PC, respectively) whereas devices of cause/reason occur 41.6% and 46.1%. The difference in terms of percentiles between the two types of markers is greater in AL (16.8%) compared with PC (7.8%). This result stands in contrast to the previous finding on general written English, i.e., LOB, where the cause/reason devices occur much more frequently (71.7%) than effect/result markers (Xuelan and Kennedy 1992). The difference between the academic corpora and the general corpus was statistically significant ($\chi^2(1) = 378.51, p < .001$, Cramer's $V = .142$); the two academic corpora, however, were not practically different ($\chi^2(1) = 35.11, p < .001$, Cramer's $V = .045$).

According to Xuelan and Kennedy (1992), two-thirds (66.9%) of 130 causative markers that they examined express cause/reason. Based on this observation and a similar percentage (71.7%) of their relative frequencies in LOB, the researchers argue for “a tendency for the language to more often mark the cause or reason rather than the effect in a causal relationship” (*ibid*: 72). As the current study shows, however, this tendency does not apply to academic writing, where result/effect is expressed more frequently than cause/reason. This finding is remarkable when we consider that the percentage of the markers that express cause/reason (out of the total causality markers) remains the same in this study as (69.6%) – in fact even slightly higher than – that in Xuelan and Kennedy (1992). It appears to reflect the interests of the academic discourse community in the effects of a particular variable(s) that are zoomed in on in the research. Authors of research articles typically pose some questions or issues that are deemed critical yet unresolved in their field of study and attempt to seek answers to those questions/issues through their own research. A significant portion of the articles is thus devoted to presenting and discussing the results of the study, especially the effects of research variables. It is also true, on the other hand, that academic works naturally entail understanding and proposing solutions to certain problems. Analyzing problems always involves the identification and interpretation of the causes and reasons. What seems more significant than the relative proportion of cause/reason versus result/effect may thus be the discovery that the prevalence of both types of causality is one of the hallmarks of academic writing, which differentiates it from general written English.

4.2 Comparison between AL and PC

One way of exploring the possible differences between the two academic disciplines is to compare causative markers that occur with high frequencies and the unique patterns of their use. Table 6 displays the ten most common causative devices in the AL and the PC corpus, with the standardized number of occurrences (per 100,000 words) and the corresponding percentages.

Table 6. Top Ten Most Frequent Causality Devices

Rank	Types	AL		Types	PC	
		frequency	percentage		frequency	percentage
1	<i>result(s)</i>	1722	19.2%	<i>result(s)</i>	1508	17.6%
2	<i>effect(s)</i>	1512	16.9%	<i>effect(s)</i>	1046	12.2%
3	<i>because</i>	888	9.9%	<i>due to</i>	793	9.3%
4	<i>thus</i>	557	6.2%	<i>thus</i>	521	6.1%
5	<i>therefore</i>	408	4.5%	<i>therefore</i>	463	5.4%
6	<i>due to</i>	325	3.6%	<i>lead to</i>	426	5.0%
7	<i>reason(s)</i>	299	3.3%	<i>since</i>	401	4.7%
8	<i>lead to</i>	245	2.7%	<i>because</i>	347	4.1%
9	<i>result in</i>	240	2.7%	<i>as</i>	330	3.9%
10	<i>outcome(s)</i>	196	2.2%	<i>result in</i>	293	3.4%
Subtotal		6392	71.3%		6128	71.6%
TOTAL		8971	100.0%		8553	100.0%

The AL and PC corpora have eight causative devices in common on the list (i.e., *result*, *effect*, *because*, *thus*, *therefore*, *due to*, *lead to*, *result in*). The other two items unique to each corpus are nouns (*reason* and *outcome*) in AL and conjunctions (*since* and *as*) in PC. *Reason* and *outcome* rank 17th and 57th in PC, whereas *since* and *as* rank 12th and 15th in AL, respectively. It thus seems that *outcome* is particularly more closely associated with AL than PC. This appears to be at least partly ascribable to the discipline-specific characteristics of the AL corpus, where students' learning outcomes are very often in focus. In fact, the noun *learning* was the most frequent L1 collocate of *outcome(s)*, followed by *literacy*, *academic*, *developmental*, etc. Similarities of the two corpora, however, apparently outweigh the differences: The top ten most frequent causative devices account for about 70% of the total tokens (71.3% in AL and 71.6% in PC).

When the top 20 causality devices were compared, the AL list included 4 nouns, zero verb, and 3 verb phrases whereas the PC inventory had 3 nouns, 3 verbs, and 4 verb phrases. In particular, *cause*, *induce*, and *yield* were three verbs that appeared with very high frequency in the PC, but not in the AL. An examination of the concordances from the PC indicates that these verbs typically convey causality involved in concrete chemical processes commonly reported in the PC articles. See examples (1) through (5) below:

- (1) Proportional increase of functionalizing molecules **causes** a change of the amount of π -conjugated bonds broken due to the back-bonding mechanism.
- (2) This irradiation process **causes** the formation of a single damage track through the films.
- (3) The extra hydrogen bond in SS-2 **induces** significant changes in geometry at each amino acid residue.
- (4) Solvation also **induces** a peak shift, from 2.45 Å separation to 2.75 Å.
- (5) The fluorescence decay curves analyzed either at 381 nm or at 485 nm **yielded** single exponential decays with the same decay parameters.

Detailed analysis of concordances of preferred causality markers further revealed patterns of use that are unique to one or the other of the two academic corpora. Above all, the noun *reason* frequently collocates with ordinal numbers (i.e., *first*, *second*, *third*) in the AL corpus. This is a type of "frame markers," which "signal text boundaries or elements of schematic text structure" (Hyland 2005: 51). By using such sequencing devices, the writer can order and arrange his/her arguments regarding causality issues, thereby guiding the readers to predict and follow the text including them more easily. See Figure 2 below:

the focus and orientation of our research. The **first reason** was the contrasting explanations and claim pre-task and extended pre-task conditions. The **first reason** may be that this measure of fluency improved up presentations for three main reasons. The **first reason** concerned time-consuming cooperation and negotiation (2001b, 2005, 2007) Cognition Hypothesis. The **second reason** was a scarcity of task-based research studies repair (e.g., Lyster and Ranta, 1997). The **second reason** is that the analysis focused on the repair data was oral, not written, is a potential **second reason** for the lack of significance in the analysis attractive as their English language competitors. A **second reason** for the use of English language textbooks or syntactic, rather than lexical, nature. A **second reason** for the importance of our findings is that here who are from the same social context. A **third reason** for examining English tense morphemes has home and were attending all-English schools. A **third reason** for the discrepancy might lie in the broad

Figure 2. Concordances of *reason* from the AL Corpus

Another noticeable pattern unique to the AL corpus is the recurrent occurrence of the formulaic phrase *there is reason to*.

of their chosen fields. As noted earlier, **there is reason to** question whether this potential benefit tween vocabulary instruction and writing, **there is reason to** believe that vocabulary instruction may be book for those interested in religion. **There is reason to** suggest that the preference for reader arners (Abrahamsson and Hyltenstam 2008), **there is reason to** believe that aptitude may constitute a n and production in literate L1 speakers, **there is reason to** believe that the same would occur with incorrect, construction. In other words, **there is reason to** view this as a syntactic error, coupled arch (Hyltenstam et al. 2009), we believe **there is reason to** question the hypothesis that predicts a h on exceptionality in language learning. **There is reason to** believe that polyglots, who belong to t actual effects of such a choice are, and **there is reason to** be skeptical about the positive gains i

Figure 3. Concordances of *there is reason to* from the AL Corpus

As shown by the concordances in Figure 3, this phrase is characteristically used with “mental verbs” (Biber et al. 1999) denoting cognitive states (e.g., *believe*, *question*, *view*), with which the writer signals his/her degree of certainty or doubt in making a claim. In terms of semantic categories or strength of epistemic commitment signaled, these verbs are classified as “probability” (as opposed to “certainty” or “possibility”) epistemic items (Holmes 1982, 1983, 1988, McEnery and Kifile 2002). In other words, with these epistemic verbs writers signal that the proposition is probably true (rather than it is certainly or possibly true). Such use of probability markers has been noted as an effective rhetorical strategy in argumentation, which with its less aggressive style tends to allow disagreeing or different views on the part of the readers (Oh and Kang 2013). It contributes to involving the readers and expanding the dialogic space for alternate viewpoints in the interaction (White and Sano 2006). This type of claim-making that relies on negotiations with the members of a shared discourse community, as will be discussed later, is more characteristic of soft disciplines than of hard sciences.

Related to the point just made, the pattern of the phrase *this is due to*, which is quite strong in epistemic strength, is often used in the PC (see Figure 4), but rarely found in the AL corpus. Instead, *due to* is almost always hedged in the AL with modal verbs such as *may*, *could*, or *might*, as illustrated in Figure 5.

ily focused on RNA instead of DNA. Perhaps **this is due to** many known RNA hairpin structures and thei o that observed in the case of slit pores. **This is due to** the fact that water molecules which direct ssure, and (iv) a smaller hysteresis loop. **This is due to** the favourable adsorption of water molecuol 1000 1C a sharp Q0 peak occurs at 62 ppm. **This is due to** the formation of forsterite (Mg₂SiO₄), and tical using anything but small basis sets. **This is due to** the fact that computer costs meaning the C s smoothed using a density representation. **This is due to** the fact that the evolution of density wit we further discuss below, we consider that **this is due to** the fact that in the formation of the full d full relativistic computations show that **this is due to** the simultaneous relativistic expansion of

Figure 4. Concordances of *due to* from the PC Corpus

ment, there is a 15% likelihood that this **could be due to** chance alone. The Tukey test shows a significant difference between written and aural tests. This **could be due to** instruction because the participants in the retention of sigmatic forms for novel verbs **could be due to** regularity, or to frequency, or to a combination of retention from the less familiar passage **may be due to** greater difficulty experienced by the weak and last training sessions (A and C). This **may be due to** increasing time pressure, as the available linguistic knowledge between adjective and noun forms **may be due to** properties of the input. First, some adjectives acknowledged that the lack of correlation **may be due to** small sample sizes. However, our suggestion of another reason for the difference in results **may be due to** the difference in the nature of the recast of specific processing patterns that **may be due to** the L1. Future research should assess the nouns and verbs: Adverbial phrases: (3) This **may be due to** the amorphous nature of the as-deposited form in the case of the No Repetition condition, it **may be due to** the difference in topic. The topic of the writing process could be shown. This **might be due to** the choice we made to include five specific nouns for L2 proficiency on L2 text quality **might be due to** the way in which L2 proficiency was measured.

Figure 5. Concordances of *due to* from the AL Corpus

The noted discrepancy between the two corpora in the use of the causative device *due to* seems to support the finding that the soft-knowledge disciplines (i.e., humanities and social science) tend to make far more frequent use of interactional metadiscourse markers such as hedges and boosters than hard sciences (Hyland 2005). Hyland attributes this finding to the observation that “the soft-knowledge fields are typically more interpretive ... than the hard sciences and their forms of argument rely more on a dialogic engagement and more explicit recognition of alternative voices” (2005: 145). Given greater variability in research context and outcomes with less certainty of the evidence for claims and confidence on the part of the writers in the soft fields, Hyland argues, it seems natural to find that there are far more negotiations and interactions between the writers and the target readers. In discussing the relation of cause and effect based on the findings of his/her research, authors of the AL research articles may well depend more on interactional metadiscourse devices such as modal verbs to hedge their arguments than those of PC articles who share “relatively unmediated real-world phenomena” (*ibid*: 145).

5. Conclusion

The current study has aimed to investigate how causality is expressed in English academic writing in comparison with general written English (the first research question), and whether causality marking differs according to academic disciplines (the second research question). The analysis has shown that academic discourse, irrespective of disciplines, is characterized by rich expression of causality, making heavy use of both cause/reason and effect/result markers. In particular, the strong preference for nouns over the other grammatical categories was characteristic of the construction of causality in research articles, verifying the prominence of nominalization in academic discourse (Biber, Conrad and Reppen 1998, Bilig 2008, Liardet 2013, Ryshina-Pankova 2010). This finding is in sharp contrast to that from general written English, where conjunctions and adverbs are most favored in expressing cause and effect (Xuelan and Kennedy 1992). Regarding the second research question, disciplinary differences between applied linguistics and physical chemistry turned out to be rather small. Overall, the two disciplines were alike in their expression of causal relations for example in preference for nominal categories and a greater focus on effect/result than cause/reason. Still, however, there existed subtle differences as exemplified by the higher deployment of nouns in AL than PC and the phraseology of some frequent causality devices (e.g., *reason* and *due to*), which are associated with distinctive epistemic practices of each discipline (i.e., weaker and stronger commitment to propositions in AL and PC, respectively). These findings were discussed in terms of the nature of academic discourse and the discrepancy between soft and hard disciplines such as the greater degree of interpretation and dialogic interaction with the readers in the former than the latter.

The findings of this study may suggest some pedagogical implications regarding the teaching and learning of academic English writing. Novice students who wish to become full-fledged members of a disciplinary discourse

community need to learn how to convey cause and effect in their writing because it is one of the most important rhetorical functions therein. They are more likely to succeed in persuading readers when they convey their messages “in ways which appeal to appropriate community recognized relationships,” or according to “the epistemic conventions of a discipline” (Hyland 2006: 21). The current findings regarding both commonalities and disparities of the two academic disciplines in causality expression lay a foundation on which students can make linguistic and rhetorical choices suitable for the interaction with other members of their respective disciplinary community.

English learners’ recurrent use of a limited variety of causative devices such as the conjunctive adverbial *so* has been repeatedly reported in previous studies (Lee 2004, Oh 2009, Yoon 2006). Learners would be helped by being exposed to the full gamut of causative devices and a variety of ways in which causality is signaled in academic discourse. Familiarizing them with a category of causative devices that are not very common in general English but favored in academic writing (i.e., nouns) seems particularly helpful. This does not need to be done in the form of a decontextualized list of causative devices. Rather, it would be pedagogically more useful for learners to read authentic academic texts that contain such markers. To be more effective, those markers may be highlighted by means of input enhancement techniques (such as typing causative devices in bold or italics) so that learners can identify them more easily. In the classroom, students may additionally be required to analyze cause and effect relationships signaled by such devices. It is also necessary to alert them to different as well as common characteristics between general English and academic writing in the construction of causal meanings.

The current study contributes to developing the curriculum and syllabi as well as teaching materials for EAP classes by providing information regarding causality expression, which is drawn from analysis of authentic academic texts. Given the inadequacy of native speaker intuition regarding the use of causative devices in academic writing (Flowerdew 1998), findings from this study may serve a useful pedagogical purpose of teaching how to convey causality fittingly in academic writing. Considering the pervasiveness and significance of causality in academic discourse and its potential difficulty especially for non-native speakers, it deserves special attention in EAP pedagogy.

The current study has only dealt with explicit causative devices, but causative relation can be expressed implicitly as well. Implicit signaling of causality may exploit, for example, non-explicit causative verbs (e.g., *destroy*, *make*) and elliptical syntax (e.g., *to V* phrases, *V-ed* phrases, *V-ing* phrases) (Xuelan and Kennedy 1992). It has been noted that in the discipline of history, the construction of causation is often achieved in implicit ways (Achugar and Schleppegrell 2005). Future studies that focus upon implicit marking of causative relations at inter-sentential, as well as intra-sentential levels, are expected to uncover a broader spectrum of linguistic and rhetorical strategies utilized for signaling causality and the similarities and differences across disciplines in this regard. This will help EAP learners to achieve a fuller understanding of academic texts and eventually be equipped with a variety of explicit and implicit ways in which the rhetorical function of causality is fulfilled in their academic discipline. It would also be worthwhile to explore the relationship between causality and epistemic modality in academic writing. As hinted in this study, the degree of certainty with which writers commit to the truth of the causal relations may display interesting disciplinary variations.

References

- Achugar, M. and M. J. Schleppegrell. 2005. Beyond connectors: The construction of cause in history textbooks. *Linguistics and Education* 16, 298-318.
- Altenberg, B. 1984. Causal linking in spoken and written English. *Studia Linguistica* 38(1), 20-69.
- Becher, T. and P. Trowler. 2001. *Academic Tribes and Territories*. Buckingham: Open University Press.
- Biber, D. and B. Gray. 2016. *Grammatical Complexity in Academic English: Linguistic Change in Writing*. Cambridge: Cambridge University Press.
- Biber, D., S. Conrad and R. Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*.

- Cambridge: Cambridge University Press.
- Biber, D., S. Johanson, G. Leech, S. Conrad and E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow, UK: Longman.
- Bridgman, P. 1961 [1927]. *The Logic of Modern Physics*. New York: Macmillan.
- Cheng, P. W. and L. R. Novick. 1992. Covariation in natural causal induction. *Psychological Review* 99(2), 365-382.
- Coffin, C. 2004. Learning to write history: The role of causality. *Written Communication* 21(3), 261-289.
- Coffin, C., M. J. Curry, M.J. S. Goodman, A. Hewings, T. Lillis and J. Swann. 2003. *Teaching Academic Writing: A Toolkit for Higher Education*. London: Routledge.
- Cortes, V. 2004. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes* 23, 397-423.
- Darian, S. 1996. Cause and effect in a corpus of science textbooks. *ESP Malaysia* 4, 65-83.
- Darian, S. 2003. *Understanding the Language of Science*. Austin, TX: The University of Texas Press.
- Downing, C. J., R. J. Sternberg and B. H. Ross. 1985. Multicausal inference: Evaluation of evidence in causally complex situations. *Journal of Experimental Psychology: General* 114(2), 239-263.
- Flowerdew, L. 1998. Integrating 'expert' and 'interlanguage' computer corpora findings on causality: Discoveries for teachers and students. *English for Specific Purposes* 17(4), 329-345.
- Gilquin, G. and M. Paquot. 2007. Spoken features in learner academic writing: Identification, explanation and solution. In M. Davies, P. Rayson, S. Hunston and P. Danielsson, eds., *Proceedings of the Fourth Corpus Linguistics Conference, University of Birmingham, 27-30 July 2007*.
- Greenbaum, S. 1969. *Studies in English Adverbial Usage*. London: Longman.
- Halliday, M. A. K. and R. Hasan. 1976. *Cohesion in English*. London: Longman.
- Holmes, J. 1982. Expressing doubt and certainty in English. *RELC Journal* 13(2), 9-28.
- Holmes, J. 1983. Speaking English with the appropriate degree of conviction. In C. Brumfit, ed., *Learning and Teaching Languages for Communication: Applied Linguistics Perspective*, 100-121. London: CILT.
- Holmes, J. 1988. Doubt and certainty in ESL textbooks. *Applied Linguistics* 9, 20-44.
- Hume, D. 1965. *An Abstract of a Treatise of Human Nature*. Hamden, Connecticut: Archon Books. (Original work published 1740).
- Hüttner, J. 2008. The genre(s) of student writing: Developing writing models. *International Journal of Applied Linguistics* 18(2), 146-165.
- Hyland, K. 2001. Humble servants of the discipline? Self-mentions in research articles. *English for Specific Purposes* 20, 207-226.
- Hyland, K. 2002. Options of identity in academic writing. *ELT Journal* 56(4), 351-358.
- Hyland, K. 2004. *Disciplinary Discourses: Social Interactions in Academic Writing*. Michigan: University of Michigan Press.
- Hyland, K. 2005. *Metadiscourse: Exploring Interaction in Writing*. New York: Continuum.
- Hyland, K. 2006. Disciplinary differences: Language variation in academic discourses. In K. Hyland and M. Bondi, eds., *Academic Discourse Across Disciplines*, 17-45. Bern: Peter Lang.
- Hyland, K. 2015. Genre, discipline and identity. *Journal of English for Academic Purposes* 19, 32-43.
- Kaidan, Z. H., A. Jalilifar and A. Don. 2021. On the significance of disciplinary variation in research articles: Perspectives from nominalization. *Cogent Education* 8(1), 1890872. DOI: 10.1080/2331186X.2021.1890872
- Kelley, H. H. 1973. The process of causal attribution. *American Psychologist* 28(2), 107-128.
- Khoo, C. S. G. 1995. *Automatic Identification of Causal Relations in Text and Their Use for Improving Precision in Information Retrieval*. Doctoral dissertation, Syracuse University, Syracuse, NY, USA.
- Lee, E. 2004. A corpus-based analysis of the Korean EFL learners' use of conjunctive adverbials. *English Teaching* 59(4), 283-301.
- Liardét, C. L. 2013. An exploration of Chinese EFL learner's deployment of grammatical metaphor: Learning to

- make academically valued meanings. *Journal of Second Language Writing* 22(2), 161-178.
- MacDonald, S. P. 1994. *Professional Academic Writing in the Humanities and Social Sciences*. Carbondale and Edwardsville, IL: Southern Illinois University Press.
- Mackie, J. L. 1980. *The Cement of the Universe: A Study of Causation*. Oxford: Oxford University Press.
- Martin, R. 1992. *English Text: System and Structure*. Amsterdam, the Netherlands: Benjamins.
- Martin, R. 2002. Writing history: Construing time and value in discourses of the past. In M. J. Schleppegrell and M. C. Colombi, eds., *Developing Advanced Literacy in First and Second Languages: Meaning with Power*, 87-118. Mahwah, NJ: Lawrence Erlbaum Assoc.
- McEnery, T. and N. Kifle. 2002. Epistemic modality in argumentative essays of second-language writers. In J. Flowerdue, ed., *Academic Discourse*, 182-195. London: Longman.
- Mill, J. S. 1973. A system of logic: Ratiocinative and inductive. Book III: Of induction. In J. M. Robsonm, ed., *Collected Works of John Stuart Mill* (vol. VII). Toronto: University of Toronto Press (Original work published in 1872).
- Moreno, A. 2003. Matching theoretical descriptions of discourse and practical applications to teaching: The case of causal metatext. *English for Specific Purposes* 22, 265-295.
- Murray, N. and G. Hughes, G. 2008. *Writing up Your University Assignments and Research Projects: A Practical Handbook*. New York: Open University Press.
- Oh, S. 2009. Korean college students' use of English demonstratives in argumentative essays. *English Teaching* 64(1), 1-29.
- Oh, S. and S. Kang. 2013. The effect of English proficiency on Korean undergraduates' expression of epistemic modality in English argumentative writing. *Journal of Asia TEFL* 10(4), 97-132.
- Rescher, N. 1970. *Scientific Explanation*. New York: Free Press.
- Ruiying, Y. and Allison, D. 2003. Research articles in applied linguistics: Moving from result to conclusions. *English for Specific Purposes* 22, 365-385.
- Rutherford, W. E. 1970. Some observations concerning subordinate clauses in English. *Language* 46, 97-115.
- Ryshina-Pankova, M. 2010. Toward mastering the discourses of reasoning: Use of grammatical metaphor at advanced levels of foreign language acquisition. *The Modern Language Journal* 94(2), 181-197.
- Van Overwalle, F. J. and F. P. Heylighen. 1995. Relating covariation information to causal dimensions through principles of contrast and invariance. *European Journal of Social Psychology* 25, 435-455.
- Veel, R. and C. Coffin. 1996. Learning to think like an historian: The language of secondary school history. In R. Hasan and G. Williams, eds., *Literacy in Society*, 191-231. Harlow, Essex: Addison Wesley Longman.
- White, P. R. and M. Sano, M. 2006. Dialogistic positions and anticipated audiences: A framework for stylistic comparisons. In K. Aijmer and A. Simon-Vandenberg, eds., *Pragmatic Markers in Contrast*, 189-214. Amsterdam: Elsevier.
- Winter, E. 1982. *Towards a Contextual Grammar of English*. London: George Allen and Unwin.
- Xuelan, F. and G. Kennedy. 1992. Expressing causation in written English. *RELC Journal* 23(1), 62-80.
- Yoon, H. 2006. A corpus-based analysis of connectors in Korean students' essay writing. *Korean Journal of Applied Linguistics* 22(2), 159-178.
- Zimmerman, F. 1989. *English for Science*. Englewood Cliffs, New Jersey: Prentice-Hall.

Examples in: English
 Applicable Languages: English
 Applicable Level: Tertiary