Expressing cause and effect in hard and soft scientific discourse: A corpus-based analysis

Javier Pérez-Guerra[1], **Elizaveta Smirnova**[2] **& Elena Kostareva**[2]
[1]Universidade de Vigo (Spain) & [2]HSE University (Perm, Russia)
jperez@uvigo.gal, easmirnova@hse.ru & ekostareva@hse.ru

Abstract

This study is a quantitative analysis of the use of cause/effect expressions in two corpora that comprise research articles in four hard and four soft disciplines, selected to represent a broad cross-section of academic discourse. Linguistic expressions of cause/effect are hypothesised to diverge in hard and soft sciences, and the differences pertain not only to key linguistic units but also to syntactic patterns. With the ultimate goal of facilitating the production of discipline-specific materials that can effectively address the needs of learners of different sciences, this study investigates the productivity of linguistic units representative of cause/effect expressions in hard and soft scientific disciplines, as identified in the Louvain EAP Dictionary, as well as the main grammatical patterns where the former are attested. Results show that cause expressions are more common in soft sciences, while the expressions of effect are used in similar proportion in the two categories. As for lexical strategies, soft sciences tend to use more nouns to express cause/effect, while hard sciences rely on prepositions and conjunctions to a larger extent.

Keywords: academic writing, cause, effect, hard science, soft science, disciplinary variation, English for Academic Purposes.

Resumen

La expresión de la causa y el efecto en el discurso de las ciencias duras y blandas: un estudio basado en corpus

Este estudio ofrece un análisis cuantitativo del uso de expresiones de causa/efecto en dos corpus compuestos por artículos de investigación en cuatro

disciplinas de ciencias *duras* y cuatro disciplinas *blandas* seleccionadas para que resulten representativas del discurso académico. Se suele asumir que las expresiones lingüísticas de causa/efecto son diferentes en las ciencias duras y blandas y que estas diferencias afectan tanto a unidades lingüísticas como a patrones sintácticos. Con el objetivo final de facilitar la producción de materiales específicos de una disciplina que sean eficaces para las necesidades de los estudiantes de diferentes ciencias, este estudio investiga la productividad de las unidades lingüísticas representativas de las expresiones de causa/efecto en las disciplinas científicas duras y blandas, tal y como se identifican en el *Louvain EAP Dictionary*, así como los principales patrones gramaticales en que se atestiguan dichas expresiones. Los resultados muestran que las expresiones de causa son más comunes en las ciencias blandas, mientras que las expresiones de efecto se utilizan con proporciones similares en las dos categorías. En cuanto a las estrategias léxicas, las ciencias blandas tienden a utilizar más sustantivos para expresar causa/efecto, mientras que las ciencias duras se basan en mayor medida en preposiciones y conjunciones.

Palabras clave: escritura académica, causa, efecto, ciencia dura, ciencia blanda, variación disciplinar, inglés para fines académicos.

1. Introduction

The expression of cause/effect is a pivotal rhetorical function in academic discourse (Flowerdew, 1998; Jordan, 1999; Gilquin & Paquot, 2007). Cause/effect is realised as a relationship established between two events, where the first is considered to be a reason for the second one and the second event is the result of the first one (Gopalan & Devi, 2017). Given their essential role in textual inference (Linderholm et al., 2000), cause/effect devices may serve to characterise academic work and prove useful in the organisation of (academic) scientific discourse (Paquot, 2010), thus facilitating effective academic communication.

Previous studies have investigated functional (i.e. communicative) differences between cause/effect expressions (van Dijk, 1977; Winter, 1977, 1982; Sanders & Spooren, 2015), their (semi-)automated detection (Marshman, 2004; Marshman & L'Homme, 2006; Chukharev-Hudilainen & Saricaoglu, 2016; Gopalan & Devi, 2017; Cao et al., 2018) and their cognitive categorisation (Altenberg, 1984; Pander Maat & Sanders, 2001; Sanders & Spooren, 2015) in different genres and discourses (see Section 2). Since academic knowledge is characterised as "the outcome of a process of getting

people to believe things" (Hyland & Jiang, 2018, p. 19), reporting research findings in a particular disciplinary context demands close attention to their contextualisation, to make the reader "accept a particular observation as a worthwhile contribution". In this vein, this study focuses on the crucial role played by cause/effect expressions in building arguments in academic discourse.

Since academic discourse shows large variability in terms of word frequencies, rhetorical moves and collocational patterns (Hyland, 2008), this investigation hypothesises that the instantiation of cause/effect relations diverges across academic disciplines and, from this stance, explores the way in which cause/effect expressions are used specifically in research articles in the so-called 'hard' and 'soft' scientific disciplines. The earliest attested use of the label 'hard' science dates back to the nineteenth century when it was employed by Winkworth et al. (1858) in *Journal of the Society of Arts*. Storer (1967) used the terms 'hard' and 'soft' to compare scientific fields on the basis of perceived methodological rigour, exactitude and objectivity. In a nutshell, the applied, empirical, experimental and natural sciences (e.g. physics, biology, astronomy, mathematics, technology) are considered 'hard', whereas the social sciences and humanities (e.g. psychology, sociology, political science, philosophy, linguistics, literature, history) are considered 'soft'. So-called hard sciences mainly report quantitative facts and tackle "the analysis of observable experience to establish empirical uniformities" (Hyland, 1999, p. 114), and consequently may opt for content/physical causality and (semi-)automatic effects (see, for example, Sweetser, 1990), whereas in soft sciences, which have "less control of variables and greater possibilities for diverse outcomes" (Hyland, 2005, pp. 187-188), reasoning by humans and argumentation, i.e. epistemic causality (Sweetser, 1990), may predominate. Study of deviations in the way in which cause/effect relations are materialised in different disciplines will contribute to research into disciplinary variation and the peculiarities of professional academic writing and have practical implications for EAP (English for Academic Purposes) instruction.

This paper is organised as follows. Section 2 outlines previous work on disciplinary variation and the instantiation of cause/effect linguistic units in academic discourse and presents the research questions. Section 3 describes the data, method of analysis and research items. Section 4 is devoted to the analysis of the data. Section 5 contains the discussion and interpretation of the results, and Section 6 reports the main findings and conclusions. The

appendices provide a list of journals used for the analysis and a list of the patterns evincing cause/effect expressions that are explored in the paper.

2. Literature review and research questions

This section summarises previous research on cause/effect linguistic devices across discourses and disciplines, which justifies the empirical study carried out in the ensuing sections, in a corpus of hard and soft academic papers.

Earlier studies on cause/effect relations focused on the functional differences of cause/effect markers and their stylistic peculiarities (see, for example, Rutherford, 1970; Kac, 1972; Sager et al., 1980; Mel'čuk et al., 1995; Flowerdew, 1998), while more recent studies have highlighted the role of cause/effect expressions in different types of discourses and tackled their automated extraction and analysis (Marshman, 2004; Marshman & L'Homme, 2006; Sanders & Spooren, 2015; Chukharev-Hudilainen & Saricaoglu, 2016; Gopalan & Devi, 2017; Cao et al., 2018). The functional effects of specific cause/effect markers, investigated by, among others, Rutherford (1970), Kac (1972) and van Dijk (1977), revealed that different causal relations (i.e., subordinate clauses containing the cause of the action expressed in the main clause and subordinate clauses which express the reason why the main clause is formulated; see Gruiţă, 1983) are encoded in different grammatical forms and are strongly dependent on communicative contexts. Other studies (Halliday & Hasan, 1976; Krogsrud, 1980; Winter, 1977, 1982; Whalley, 1981) focused on the functions of the markers in the written medium, and on their contribution to the organisation and cohesion of discourse. Also, Li (2014, p. 12) demonstrated the connection between cause/effect expressions and the degree of subjectivity of the discourse – see specifically Li et al. (2013, p. 105), on the role of subjectivity in the use of Mandarin Chinese causal connectives.

To narrow down the scope of this piece of research, other contributions are more closely aligned with the methodological perspective adopted in this paper. Firstly, in Abraham's (1991) study on the use of *because* and *because of* in different registers (spoken and written language) and genres (fiction mysteries, biographies and scientific writings) in English, it was found that *because* tends to encode new information, whereas *because of* significantly precedes given information, except in the scientific samples, where *because of* was used to convey new information, more often than given information.

Degand (2000, p. 694), who also explored the choice between causal prepositional phrases and causal clauses in Dutch, concluded that given information, inferable from the surrounding discourse, may be presented in compacted (phrasal) form, whereas new information is typically materialised in a developed, explicit grammatical structure, adopting a clausal design on many occasions. Secondly, as far as disciplinary variation is concerned, Hyland studied the productivity of specific linguistic strategies such as self-mentions (Hyland, 2001), the use of directives in academic writing (Hyland, 2002), and the expression of the writers' positions and addresses to the readers (Hyland, 2005) in various disciplines. As regards the soft/hard dichotomy, Hyland (2005) showed that greater elaboration is demanded in humanities, whereas hard-knowledge research opts for the application of precise expressive strategies. In this vein, one may hypothesise that hard sciences, assumed to be more objective and impersonal, are likely to employ more conjunctions and prepositions when expressing causality; whereas soft sciences might opt for epistemic modality, that is, connected with people's "judgments about the factual status of the proposition" (Palmer, 2001, p. 8), by using more cause/effect nouns and verbs. Thirdly, Groom (2005) dealt with variation in grammatical patterns across history and literary-criticism discourses through the analysis of the corpus data. He proved that genre/discipline-specific practices determine variation in a number of grammatical patterns (for example, verb+adjective+*that* as in *it is clear that the problem*... versus verb+adjective+*to*-infinitive as in *it is important to compare*…), so it is reasonable to assume that different disciplines may express causality through different patterns as well. The necessity to compile pervasive discipline-based lists of grammatical patterns was also emphasised by Ma and Qian (2020, p. 162) in their study on frequent academic verbs. In a similar vein, this paper provides a list of patterns in which cause/effect devices are attested with the purpose of exploring disciplinary variation in their use.

The studies mentioned here pave the way for a fine-grained exploration of cause/effect linguistic realisations across academic disciplines and for the study of cause/effect relations materialised by specific lexical and grammatical strategies in academic texts, under the hypothesis that they diverge in hard and soft sciences as a consequence of the different degrees of impersonality and objectivity in such discourses. Two research questions (RQ) will be addressed in what follows:

- RQ1: What are the differences in the frequency of cause/effect linguistic units in research articles in hard and soft sciences?
- RQ2: What are the differences as regards the grammatical patterns in cause/effect expressions employed by hard and soft sciences?

3. Data and method

This section describes the corpus and the linguistic units that serve as the empirical basis of the analysis carried out in Section 4.

3.1. Corpus description

The analysis was conducted on two corpora of randomly selected research articles from peer-reviewed journals representing a broad cross-section of academic discourse. The first corpus comprises hard-science texts in chemistry, physics, mathematics and engineering. The other corpus includes soft-science texts in business studies, history, linguistics and political science. All the articles were published in leading academic journals, indexed in Scopus Quartile 1, in 2016-2020 (see Appendix A). Within each discipline we took a random sample of articles to ensure that disciplinary subcorpora are balanced as regards token numbers. In more detail, the random selection involved browsing the journals' websites and picking an approximately equal number of papers from each of them for every discipline to ensure a balanced corpus with regards to the number of tokens. After the selection of the articles, raw texts were prepared for further analysis. To that purpose, tables, formulae, graphs, charts, metadata and reference lists were removed from the manuscripts. Table 1 provides information about the size of corpora.

	Discipline	Number of texts	Word totals
Hard sciences	Chemistry	34	197,806
	Physics	44	200,206
	Mathematics	28	199,380
	Engineering	34	198,926
	Total	*140*	*796,318*
Soft sciences	Business	10	197,956
	Linguistics	10	200,997
	History	10	199,394
	Political science	11	202,040
	Total	*41*	*800,387*

Table 1. Corpus size and composition.

3.2. Linguistic units under consideration

This investigation relies on the exploration of lexical and constructional or pattern-based devices revealing cause/effects relations. As regards the former, the cause/effect linguistic units selected for the analysis were taken from the Louvain EAP Dictionary (LEAD) (https://leaddico.uclouvain.be), a web-based dictionary of EAP for non-native speakers of English (Granger & Paquot, 2010, 2015; Paquot, 2012). This resource provided exhaustive lists of academic words grouped by their function. The dictionary is grounded on the so-called Academic Keyword List (Paquot, 2010) of 930 items, that is, words that were identified to be especially frequent in academic texts and quite uncommon in other text types. In Paquot's (2010, p. 29) words, these lexical items refer to the activities that "characterise academic work, organise scientific discourse, and build the rhetoric of academic texts, and so be granted the status of academic vocabulary". Our investigation is based on the list of 34 cause/effect expressions provided by LEAD. In order to adopt an objective criterion for the selection of the key lexical items, we restricted our queries to the expressions in LEAD's Academic Keywork List, which, as already mentioned, was compiled on the basis of the frequency and functional distinctiveness of the items – for example, the verbs *cause* and *effect* themselves are not included in the Academic Keyword List because of their significant frequency in constructions other than the cause/effect ones. Even though cause/effect processes are logically connected, they are realised in texts by different linguistics means; therefore, in this study we treat cause/effect devices separately. The linguistic units under consideration from the LEAD dictionary are given in Table 2.

	Nouns	Conjunctions	Verbs	Prepositions
CAUSE	*cause factor ground origin reason root source*	*because given that since*	*contribute*	*because of due to following given in view of on account of owing to*
EFFECT	*consequence effect implication outcome result*	*so so that*	*result lead*	*accordingly consequently hence naturally thereby therefore thus*

Table 2. Linguistic units expressing cause/effect used in the study.

It should be noted here that we have looked into all the forms of the linguistic units in the study. For example, the inclusion of the noun *cause* implies the analysis of singular *cause* and also of plural *causes*; likewise, as regards the verb *contribute*, forms such as *contribute*, *contributes*, *contributed* and *contributing* have been counted in the analysis.

3.3. Data analysis

The corpus data were processed with the concordance function of AntConc (Anthony, 2014). The results were manually disambiguated, and those that did not convey cause/effect were removed from the database. To give an example, instances such as (1), in which the key noun *origin* is not evincing cause but refers to 'the point of intersection of the axes in Cartesian coordinates' (*Oxford English Dictionary*), were discarded.

> (1) the cross-section of C, is either the sum of two distinct l-dimensional subspaces of Rl+k meeting only at the origin or the sum of four distinct rays in R1+k meeting only at the origin, depending on whether C ∈ P≤n−2 or C ∈ Cn−1, respectively.

In line with Hunston and Francis (2000, p. 37), who define 'pattern' as "all the words and structures which are regularly associated with the word and which contribute to its meaning", in this study we hypothesise that the selection of specific grammatical patterns by the linguistic units previously described is meaningful, and thus contributes to the linguistic characterisation of scientific academic discourse. In order to detect the main grammatical patterns in which the nominal and verbal cause/effect key items were attested, we analysed the concordance lines of the items in the corpus. Function words such as cause/effect conjunctions and prepositions were not considered in this respect. The nominal and the verbal items that demonstrated pattern variability across the disciplines were, respectively, the cause nouns *origin*, *factor*, *cause*, *reason* and *source*, the effect nouns *consequence*, *effect*, *implication*, *outcome* and *result*, the cause verb *contribute,* and the effect verb *lead*. The concordance lines were right-sorted alphabetically via the AntConc concordance tool, which eased their manual analysis and the subsequent detection of the patterns and the actual variants. The list of relevant patterns, given in Appendix B, is based on the taxonomy of 'grammar patterns' identified by the *Collins COBUILD English Dictionary* (https://www.collinsdictionary.com/grammar-pattern).

To test the statistical significance of the differences in the use of the cause/effect markers in the hard and soft sciences, the chi-squared test was used. The chi-squared test has a number of advantages, namely, it has been proven to be more sensitive in comparison with, for example, t-test, since it does not require the normal distribution of the data (McEnery & Wilson, 2001). Following Brezina (2018, p. 112), we posit that the chi-squared test is appropriate for our study, since we deal with one linguistic (the cause/effect expressions) and one explanatory (hard/soft science) variable. Individual frequencies lower than 5 examples, which cannot support statistical validation, have been discarded in this analysis. Table 3 and Figures 1 to 3 show the normalised frequencies of the key grammatical categories conveying cause in the hard- and soft-science datasets. Preliminary statistical significance was conventionalised as follows: '***' when $p \leq .001$, '**' when $p \leq .01$ and '*' when the significance of the variation is reported by a p-value $\leq .05$, and we have fixed our *a priori* α criterion (i.e., the significance level threshold in our hypothesis testing or, in other words, the maximum acceptable probability of rejecting the null hypothesis) for significant individual item-by-item variation only at the level of 0.001 ('***' in the ensuing figures).

4. Results

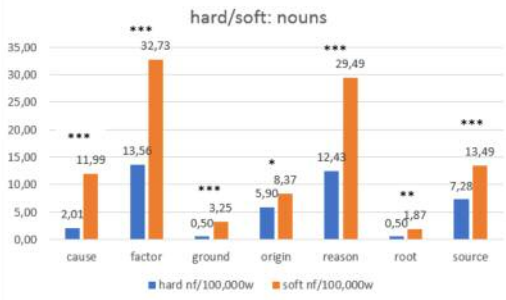
This section deals with the analysis of the productivity of the key lexical items identified in LEAD as potential features characterising cause/effect expressions in hard- and soft-science discourse (Section 4.1), and of the variation evinced by the grammatical patterns in which the lexical items are attested (Section 4.2).

4.1. Cause/effect linguistic units

In this section we report the results of the analysis of the distribution of the key words in cause/effect expressions that were identified as potential proxies for the hard/soft distinction.

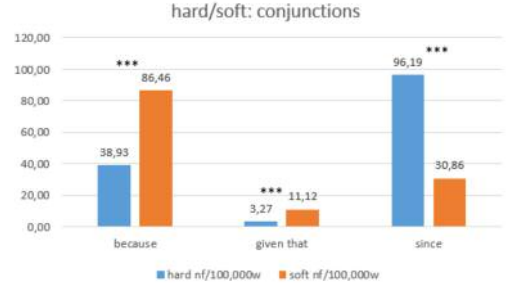
Cause

As regards the linguistic units evincing cause, Table 3 displays the raw and normalised frequencies (per 100,000 words) of the key nouns, conjunctions, verbs and prepositions, as well as the chi-squared and p values of the variation hard versus soft sciences.



	hard	soft
	tokens	tokens
*cause**	16	96
*factor**	108	262
*ground**	4	26
*origin**	47	67
*reason**	99	236
*root**	4	15
*source**	58	108
Total	**336**	**810**

Figure 1. Normalised (and raw) frequency of cause nouns employed in the hard- and soft-science papers.

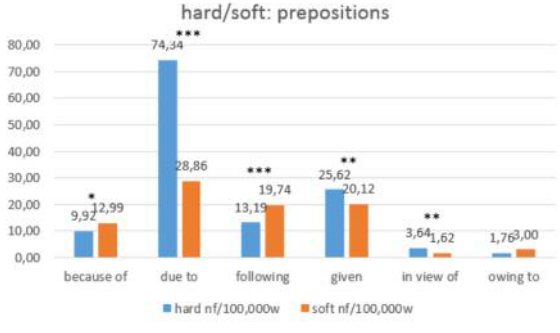


	hard	soft
	tokens	tokens
because	310	692
given that	26	89
since	766	247
Total	**1102**	**1028**

Figure 2. Normalised (and raw) frequency of cause conjunctions employed in the hard- and soft-science papers.

	hard	hard	soft	soft		
	tokens	nf/100,000w	tokens	nf/100,000w	chi-square	p-value
*contribut**	93	11.679	232	28.986	58.731	<0.00001 ***

Table 3. Normalised (and raw) frequency of the cause verb employed in the hard- and soft-science papers.



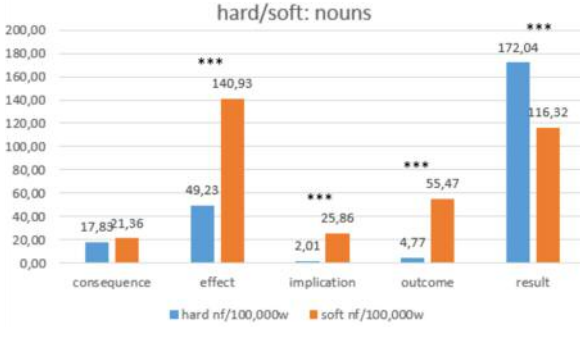
	hard	soft
	tokens	tokens
because of	79	104
due to	592	231
following	105	158
given	204	161
in view of	29	13
on account of	1	1
owing to	14	24
Total	**1024**	**692**

Figure 3. Normalised (and raw) frequency of the cause prepositions employed in the hard- and soft-science papers.

The analysis of the productivity for the linguistic units associated with *cause* relations leads to the following findings. Firstly, the frequency of the key nouns conveying cause is significantly higher in the soft-science dataset than in the hard-science one. Secondly, in light of the frequencies of the key conjunctions reported by LEAD, one cannot conclude that this grammatical category serves as a linguistic proxy for the hard- versus soft-science characterisation. In detail, *because* and *given that* are significantly more frequent in the soft-science papers, and *since* is specifically pervasive in the hard-science articles. Thirdly, the frequency of the only key verb, i.e. *contribute*, is significantly higher in the soft-science papers. Fourthly, the key prepositions exhibiting the most significant variation in the hard- and the soft-science datasets are *due to* and *given*, their productivity being higher in the hard-science texts. Summing up, of the linguistic units in LEAD that contribute to the characterisation of academic discourse conveying cause relations, the data reveal that prepositions are characteristic of hard sciences, whereas the key nouns and the verb (*contribute*) are especially frequent in soft-science papers.

Effect

Table 4 shows the raw and normalised frequencies (per 100,000 words) of the key nouns, conjunctions, verbs and adverbs in the effect dataset, as well as the chi-squared and *p* values of the variation hard versus soft sciences. Figures 4 to 7 provide the normalised frequencies of these key grammatical categories in the hard- and soft-science subcorpora.

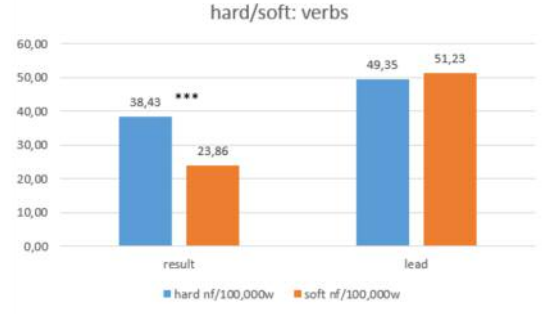


	hard	soft
	tokens	tokens
*consequence**	142	171
*effect**	392	1128
*implication**	16	207
*outcome**	38	444
*result**	1370	931
Total	**1958**	**2881**

Figure 4. Normalised (and raw) frequency of the effect nouns employed in the hard- and soft-science papers.

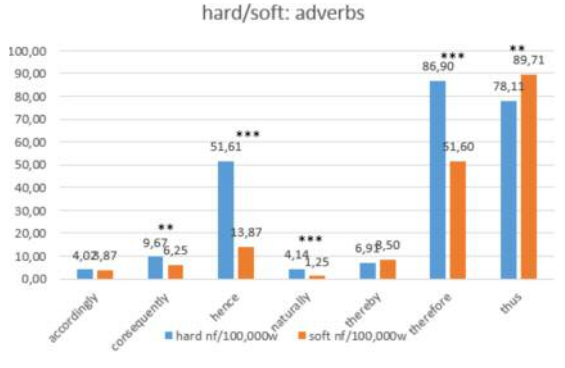
	hard	soft
	tokens	tokens
so	297	266
so that	190	60
Total	**487**	**326**

Figure 5. Normalised (and raw) frequency of the effect conjunctions employed in the hard- and soft-science papers.



	hard	soft
	tokens	tokens
*result**	306	191
*lead**	393	410
Total	**699**	**601**

Figure 6. Normalised (and raw) frequency of the effect verbs employed in the hard- and soft-science papers.



	hard	soft
	tokens	tokens
accordingly	32	31
consequently	77	50
hence	411	111
naturally	33	10
thereby	55	68
therefore	692	413
thus	622	718
Total	**1922**	**1401**

Figure 7. Normalised (and raw) frequency of the effect adverbs employed in the hard- and soft-science papers.

The quantitative analysis of the linguistic units expressing effect leads to the following results. Firstly, the frequency of all effect nouns but *consequence* and *result* is significantly higher in the soft-science dataset than in the hard-science one. Secondly, the frequencies of the two key conjunctions, i.e. *so* and *so that*, are higher in the hard-science texts, even though the difference in the frequency of the former is not statistically significant. Thirdly, the frequencies of the effect verbs demonstrate different patterns: while *result* is significantly more common in hard sciences, *lead* is more often used in the soft disciplines, but on this occasion the difference is not statistically significant. Fourthly, as regards the adverbs in effect expressions, *consequently*, *hence*, *naturally* and *therefore* are significantly more common in the hard-science papers, whereas *thus* demonstrates significantly higher frequencies in soft sciences. Overall, the analysis of the data revealed that in soft sciences, effect tends to be expressed via nouns, whereas conjunctions and adverbs are characteristic of hard sciences. The key verbs have not proved to be clear indicators of the hard/soft distinction.

Table 4 summarises the variation between the hard- and the soft-science subcorpora as regards the distribution and keyness of the central categories nouns and linking items such as prepositions and conjunctions, given that verbs and adverbs have shown misleading dispersion. The chi-squared test applied to the figures in Table 4 indicates that the variation is highly significant $\chi^2(2)= 332.38$, $p\leq.0001$, well below the α criterion of .017 demanded by Bonferroni correction in this multiple comparison involving 7 noun and 10 preposition/conjunction levels, and an *a priori* α of .001 for individual comparison.

cause/effect	hard	soft
noun	2,294	3,691
function words	2,613	2,046

Table 4. Cause/effect categorial variation in soft/hard science (raw frequencies).

In the next stage, we looked at the distribution of the linguistic units expressing cause/effect across disciplines. The results are presented in Table 5 (normalised frequencies per 100,000 words in brackets).

		Hard					Soft				
	Features	Chemistry	Engineering	Maths	Physics	Total	Business	History	Linguistics	Political science	Total
Cause	Nouns	120 (60.67)	76 (38.21)	38 (19.06)	102 (50.95)	**336 (45.96)**	170 (85.88)	238 (119.36)	199 (99.01)	203 (100,48)	**810 (101.20)**
	Conjunctions	168 (84.93)	138 (69.37)	559 (280.37)	237 (118.38)	**1102 (138.39)**	244 (123.26)	163 (81.75)	272 (135.33)	349 (172.74)	**1028 (128.44)**
	Verbs	19 (9.61)	15 (7.54)	21 (10.53)	38 (18.98)	**93 (11.68)**	77 (38.90)	71 (35.61)	48 (23.88)	36 (17.82)	**232 (28.99)**
	Prepositions	279 (141.05)	359 (180.47)	149 (74.73)	237 (118.38)	**1024 (128.59)**	164 (82.85)	178 (89.72)	155 (77.12)	195 (96.52)	**692 (86.46)**
	Total	**586 (296.26)**	**588 (295.59)**	**767 (384.69)**	**614 (306.69)**	**2555 (324.62)**	**655 (330.89)**	**650 (326.44)**	**674 (335.34)**	**783 (387.56)**	**2762 (345.09)**
Effect	Nouns	358 (180.99)	571 (287.04)	524 (262.81)	505 (252.24)	**1958 (245.88)**	704 (355.63)	258 (129.39)	482 (239.80)	1437 (711.25)	**2881 (359.95)**
	Conjunctions	19 (9.61)	80 (40.22)	236 (118.37)	152 (75.92)	**487 (61.16)**	60 (30.31)	42 (21.06)	112 (55.72)	112 (55.43)	**326 (40.73)**
	Verbs	184 (93.02)	223 (112.10)	102 (51.16)	190 (94.90)	**699 (87.78)**	222 (112.15)	99 (49.65)	102 (50.75)	178 (88.10)	**601 (75.09)**
	Adverbs	321 (162.28)	483 (242.80)	608 (304.95)	510 (254.74)	**1922 (241.36)**	451 (227.83)	249 (124.88)	401 (199.51)	300 (148.49)	**1401 (175.04)**
	Total	**882 (445.90)**	**1357 (682.16)**	**1470 (737.29)**	**1357 (677.80)**	**5066 (636.18)**	**1437 (725.92)**	**648 (324.98)**	**1097 (545.78)**	**2027 (1003.27)**	**5209 (650.81)**

Table 5. Cause/effect expressions across disciplines: raw and normalised frequencies (brackets).

Table 5 shows that the soft-science database contains significantly more linguistic units expressing cause than the hard-science subcorpus ($p \leq .01$, $\chi^2=6.94$). In detail, the large number of conjunctions involved in cause expressions in the hard-science texts is due to the mathematics samples, where this grammatical category is much more frequent than in the other hard disciplines ($p \leq .01$, χ^2 ranging from 131.3295 when compared to physics, to 252.8911 with respect to engineering). By contrast, nouns are more common in soft sciences except in the history texts (history versus the other soft sciences: $p \leq .01$, χ^2 ranging from 16.6951 when compared to business, to 65.0608 with respect to political science). As already mentioned, the only verb expressing cause is, in general, more common in the soft samples, except in the political science articles, where its frequency is lower than in the other soft sciences and is closer to that in, for examples, physics. Finally, the use of prepositions in cause expressions is significantly pervasive in hard sciences as compared to the soft samples, their frequency in the mathematics texts being closer to that in the soft disciplines, with half as many instances as in the hard-science dataset.

Table 5 also shows that the overall frequencies for the linguistic units in the effect expressions are quite alike in the hard and the soft disciplines, the difference not being statistically significant. The key nouns are more common in the soft sciences mainly due to their incidence in the political science papers, where their frequency is especially higher when compared to the other soft disciplines. The key conjunctions are significantly pervasive in the hard disciplines versus the soft ones, especially in the mathematics samples (mathematics versus other hard sciences: χ^2 ranging from 18.5166 when compared to physics, to 182.8332 with respect to chemistry; $p \leq .01$). The key verbs in the effect expressions are less frequent in soft sciences, specifically in the history and the linguistics papers, where they occur half as often than in the other disciplines.

These results thus confirm our hypothesis about the dominance of cause/effect nouns and verbs in soft sciences, and the pervasiveness of conjunctions and prepositions in hard sciences. This can be explained by way of the different types of causality (content/physical vs. epistemic) in the disciplines (see Section 5).

4.2 Cause/effect patterns

This section tackles the relevance of the patterns headed by the linguistic units selected by LEAD. Table 6 provides the frequencies and percentages (between brackets) per lexical item and discipline of the patterns listed in Appendix B. The percentages are computed by considering all occurrences of the noun or verb in the papers; to give an example, the corpus encompasses 20 tokens of the noun *origin*, with 8 of them (40%) appearing in the pattern 'det origin of n/-ing v-link', 4 (20%) in 'v det origin of n', and 3 (15%) in 'v-link of origin', totalling 15 (75%) cause/effect patterns with *origin* ('Total use in patterns') out the 20 occurrences ('Total occurrences') of this lexical item. The table also provides the mean value of the percentages of head nouns/verbs employed in the patterns (mean of 'Total use in patterns').

	Pattern	Hard					Soft				
		Chemistry	Engineering	Maths	Physics	Total	Business	History	Linguistics	Political science	Total
Origin	**Total occurrences**	**20**	**7**	**1**	**19**	**47**	**5**	**37**	**1**	**24**	**67**
	det origin of n/-ing v-link	8 (40)	4 (57)	1 (100)	2 (11)	**15 (32)**	0	0	0	1 (4)	**1 (1)**
	v det origin of n	4 (20)	1 (14)	0	4 (21)	**9 (19)**	3 (60)	6 (16)	0	1 (4)	**10 (15)**
	v-link of origin	3 (15)	0	0	2 (11)	**5 (11)**	1 (20)	1 (3)	1 (100)	0	**3 (4)**
	Total use in patterns	**15 (75)**	**5 (71)**	**1 (100)**	**8 (43)**	**29 (62)**	**4 (80)**	**7 (19)**	**1 (100)**	**2 (8)**	**14 (20)**
Factor	**Total occurrences**	**54**	**26**	**10**	**18**	**108**	**50**	**44**	**98**	**70**	**262**
	det factor that/which v	2 (4)	1 (4)	5 (50)	5 (28)	**13 (12)**	10 (20)	4 (9)	9 (9)	14 (20)	**37 (14)**
	det factor in n/-ing	0	3 (11)	5 (50)	5 (28)	**13 (12)**	8 (16)	17 (39)	6 (6)	5 (7)	**36 (14)**
	Total use in patterns	**2 (4)**	**4 (15)**	**10 (100)**	**10 (56)**	**26 (24)**	**18 (36)**	**21 (48)**	**15 (15)**	**19 (27)**	**73 (28)**
Cause	**Total occurrences**	**5**	**11**	**0**	**0**	**16**	**16**	**49**	**14**	**17**	**96**
	det cause of n v	5 (100)	0	0	0	**5 (31)**	4 (25)	13 (27)	1 (7)	4 (23)	**22 (23)**
	det cause for n	0	1 (9)	0	0	**1 (6)**	0	0	1 (7)	1 (6)	**2 (2)**
	Total use in patterns	**5 (100)**	**1 (9)**	**0**	**0**	**6 (37)**	**4 (25)**	**13 (27)**	**2 (14)**	**5 (29)**	**24 (25)**
Reason	**Total occurrences**	**15**	**27**	**22**	**35**	**99**	**49**	**72**	**58**	**57**	**236**
	det reason for n/-ing	3 (20)	10 (37)	6 (27)	3 (9)	**22 (22)**	9 (18)	12 (17)	11 (19)	6 (10)	**38 (16)**
	for det reason	8 (53)	11 (41)	10 (45)	15 (43)	**44 (44)**	17 (35)	19 (26)	21 (36)	21 (37)	**78 (33)**
	reason to-inf	1 (7)	0	1 (5)	3 (9)	**5 (5)**	2 (4)	10 (14)	7 (12)	13 (23)	**32 (14)**
	det reason why clause v-link	2 (13)	0	4 (18)	1 (3)	**7 (7)**	6 (12)	6 (8)	3 (5)	10 (18)	**25 (11)**
	Total use in patterns	**14 (93)**	**21 (78)**	**21 (95)**	**22 (64)**	**78 (78)**	**34 (69)**	**47 (65)**	**42 (72)**	**50 (88)**	**173 (74)**
Source	**Total occurrences**	**26**	**3**	**5**	**24**	**58**	**43**	**16**	**20**	**29**	**108**
	det source of n	13 (50)	3 (100)	2 (40)	7 (29)	**25 (43)**	31 (72)	6 (37)	8 (40)	22 (76)	**67 (62)**
	n as det source of n	1 (4)	0	1 (20)	1 (4)	**3 (5)**	8 (19)	4 (25)	2 (10)	2 (7)	**16 (15)**
	Total use in patterns	**14 (54)**	**3 (100)**	**3 (60)**	**8 (33)**	**28 (48)**	**39 (91)**	**10 (62)**	**10 (50)**	**24 (83)**	**83 (77)**
Consequence	**Total occurrences**	**30**	**7**	**61**	**44**	**142**	**54**	**47**	**24**	**46**	**171**
	as a consequence of n	10 (33)	0	11 (18)	5 (11)	**26 (18)**	3 (6)	6 (13)	0	5 (11)	**14 (8)**
	as a consequence	4 (13)	1 (14)	12 (20)	10 (23)	**27 (19)**	7 (13)	3 (6)	3 (13)	9 (20)	**22 (13)**
	v-link det consequence of n	8 (27)	4 (57)	9 (15)	4 (9)	**25 (18)**	3 (5)	5 (11)	2 (8)	4 (9)	**14 (8)**
	det consequence of n/-ing v	2 (7)	2 (29)	4 (6)	3 (7)	**11 (8)**	7 (13)	2 (4)	3 (13)	2 (4)	**14 (8)**
	v det consequence for n	0	0	2 (3)	0	**2 (1)**	11 (20)	5 (11)	3 (13)	2 (4)	**21 (12)**
	Total use in patterns	**24 (80)**	**7 (100)**	**38 (62)**	**22 (50)**	**91 (64)**	**31 (57)**	**21 (45)**	**11 (47)**	**22 (48)**	**85 (49)**
Effect	**Total occurrences**	**119**	**111**	**10**	**152**	**392**	**241**	**73**	**182**	**632**	**1128**
	have det effect on n	32 (27)	10 (9)	1 (10)	9 (6)	**52 (13)**	41 (17)	3 (4)	6 (3)	57 (9)	**107 (9)**
	det effect of n/-ing on n	32 (27)	42 (38)	2 (20)	60 (40)	**136 (35)**	143 (59)	19 (26)	52 (29)	111 (18)	**325 (29)**

	Total use in patterns	64 (54)	52 (47)	3 (30)	69 (46)	188 (48)	184 (76)	22 (30)	58 (32)	168 (27)	432 (38)
Implication	Total occurrences	0	2	2	12	16	57	17	43	90	207
	det implication v-link	0	0	1 (50)	0	1 (6)	2 (4)	0	0	6 (7)	8 (4)
	det implication of n	0	1 (50)	0	3 (25)	4 (25)	23 (40)	5 (29)	3 (7)	27 (30)	58 (28)
	det implication for n	0	1 (50)	0	1 (8)	2 (13)	20 (35)	3 (18)	6 (14)	9 (10)	38 (18)
	Total use in patterns	0	2 (100)	1 (50)	4 (33)	7 (44)	45 (79)	8 (47)	9 (21)	42 (47)	104 (50)
Outcome	Total occurrences	10	23	0	5	38	156	33	33	222	444
	det outcome of n	8 (80)	5 (22)	0	2 (40)	15 (39)	19 (12)	11 (33)	3 (9)	24 (11)	57 (13)
	det outcome v-link	2 (20)	2 (9)	0	2 (40)	6 (16)	9 (6)	1 (3)	2 (6)	11 (5)	23 (5)
	Total use in patterns	10 (100)	7 (31)	0	4 (80)	21 (55)	28 (18)	12 (36)	5 (15)	35 (16)	80 (18)
Result	Total occurrences	170	449	451	300	1370	191	88	212	440	931
	as a result of n	14 (8)	6 (1)	0	3 (1)	23 (2)	20 (10)	10 (11)	18 (9)	6 (1)	54 (6)
	link-v det result of n	18 (11)	8 (2)	3 (1)	1 (03)	30 (2)	3 (2)	8 (9)	8 (4)	4 (1)	23 (2)
	in result of n	0	2 (04)	0	0	2 (0.1)	0	0	0	0	0
	as a result	7 (4)	22 (5)	1 (0.2)	5 (2)	35 (3)	33 (17)	4 (5)	5 (2)	21 (5)	63 (7)
	det result link-v	15 (8)	18 (4)	33 (7)	18 (6)	84 (6)	5 (3)	7 (8)	5 (2)	30 (7)	47 (5)
	Total use in patterns	54 (31)	56 (12.4)	37 (8.2)	27 (9.3)	174 (13.1)	61 (32)	29 (33)	36 (17)	61 (14)	187 (20)
Contribute	Total occurrences	19	15	21	38	93	77	71	48	36	232
	contribute to n	18 (95)	15 (100)	18 (86)	31 (82)	82 (88)	56 (72)	61 (86)	35 (73)	32 (89)	184 (79)
	contribute to-inf	1 (5)	0	0	0	1 (1)	0	3 (4)	0	0	3 (1)
	Total use in patterns	19 (100)	15 (100)	18 (86)	31 (82)	83 (89)	56 (72)	64 (90)	35 (73)	32 (89)	187 (80)
Lead	Total occurrences	85	130	38	140	393	154	57	60	139	410
	lead to n/-ing	80 (94)	129 (99)	30 (79)	138 (99)	377 (96)	128 (83)	46 (81)	50 (83)	111 (80)	335 (82)
	lead n to-inf	5 (6)	1 (1)	2 (5)	2 (1)	10 (2)	26 (17)	11 (19)	10 (17)	28 (20)	75 (18)
	lead towards n/-ing	0	0	6 (16)	0	6 (2)	0	0	0	0	0
	Total use in patterns	85 (100)	130 (100)	38 (100)	140 (100)	393 (100)	154 (100)	57 (100)	60 (100)	139 (100)	410 (100)
	No. of different patterns	*27*	*25*	*24*	*28*	*35*	*30*	*30*	*28*	*31*	*33*
	Mean proportion of patterns	*66*	*64*	*58*	*50*	*55*	*61*	*50*	*46*	*48*	*48*

Table 6. Cause/effect patterns employed in the hard- and the soft-science papers.

Table 6 shows that even though the number of pattern types is practically identical in the hard- and the soft-science datasets, their distribution is uneven across disciplines. Firstly, overall, the hard sciences appear to be more 'patterned', i.e. employ more grammatical patterns, as regards the expression of cause/effect in view of the higher percentages of most of the linguistic constructions investigated here in comparison with those corresponding to the

soft sciences. Secondly, as regards the trends that are unequivocally revealed by the data, all the patterns involving the nouns *factor*, *implication*, *result* and *source* are more common in the soft disciplines. Thirdly, the pattern *det origin of n/-ing v-link* is hardly attested in the subcorpus of soft sciences, while the patterns *v det consequence for n*, *det implication v-link* and *det implication for n* are very rare in the hard-science papers. Fourthly, Table 6 evinces a number of idiosyncratic findings. To give a few examples, *as a consequence of n* is not found either in the engineering or in the linguistics samples; *lead n to-inf* is rare in engineering, mathematics and physics; *to contribute to-inf* is also infrequent in the data, and is only rarely found in chemistry and history. Fifthly, the data reveal correlations between lexical items and grammatical patterns. For example, most occurrences of the nouns *reason* and *consequence* are attested in the patterns depicted in Appendix B, while the noun *result* occurs in patterns to a much lesser extent. Finally, the data demonstrate that the variation evinced by the proportions of unit per pattern is significant across disciplines. For example, even though the range of patterns disclosed by the engineering data is limited, the proportion of the linguistic units expressing cause/effect attested in the patterns is large in this discipline. By contrast, the variety of patterns employed in the political science dataset is relatively large, and the proportion of cause expressions used in patterns only accounts for approximately 48% of the tokens in this discipline.

5. Discussion

This study has given empirical support to the characterisation of scientific discourse on the basis of the frequency of the linguistic items and patterns involved in expressions of cause/effect. As regards the first research question (RQ1: What are the differences in the frequency of cause/effect linguistic units in research articles in hard and soft sciences?), the results have revealed, first, that overall, cause expressions are more common in soft disciplines, while the proportion of effect expressions is quite alike in hard and soft sciences. However, causality expressions realised by means of function words (conjunctions and/or prepositions) are more frequent in the hard sciences. This finding accords both with the objective style of the hard-science texts, where statements are recurrently justified and nuanced, and with the epistemic and subjective nature of the soft-science discourse (see, for example, Biber, 1988), where the weight of the author in such texts does not require probatory strategies to the same extent as in the hard sciences.

Second, the analysis of the linguistic units revealed that soft sciences rely on nouns in linguistic strategies of cause/effect to a larger extent than the hard disciplines, where functions words are more significant in this respect. The extensive use of nouns expressing cause/effect in soft sciences is in keeping with Biber and Gray (2016, p. 135), who also attested an overall increase in the use of nouns in science prose. Relying on the literature, we hypothesised that the objective and impersonal style of the hard-science discourse is in keeping with the frequent use of linguistic strategies that provide arguments and justify the statements. From this perspective, expressions introduced by linking items such as conjunctions (mostly in adverbial clauses) and prepositions (commonly introducing modifiers of nominal categories) are optimal ways of contributing towards the expression of causality and effect in these hard-science texts. By contrast, the epistemic character of the soft-science discourse reveals a more speaker-centric assertive style, where linking functional items introducing 'probes' such as adverbials or modifiers are not pervasive, which justifies the frequency of grammatical categories such as nouns. Also, soft sciences might be more responsive to the use of nouns expressing cause/effect because soft-knowledge domains implicate less control of variations and thus more possibilities for varied outcomes (Hyland, 2005). To give an example, the noun *implication*, which occurs on 223 occasions in the corpus, expresses epistemic causality in 202 instances, as in (2) below.

> (2a) The key pedagogical *implication* of this study is that a simple act of copying novel words, while processing meaningful L2 input, may significantly boost quality of lexical knowledge (AL 2016-4)
>
> (2b) This section tests the second observable implication of the theory – when lineage groups join village political institutions it increases the likelihood of land expropriation (WP 2016-1)

In response to the second research question (RQ2: What are the differences as regards the grammatical patterns in cause/effect expressions employed by hard and soft sciences?), hard and soft sciences have been shown to prefer different patterns, some of which are discipline- (and even subdiscipline-) specific. This might be related to Hyland's (2008) claim that the use of discipline-specific language identifies writers as experts in their own discourse communities.

As regards the discipline-specific differences, although the mathematics samples were found to employ cause/effect expressions more frequently

than the other disciplines, the range of patterns was found to be more limited in the former. This can be explained by the fact that most of the patterns under consideration here are based on head nouns and verbs, and the mathematics texts opt for conjunctions and adverbs of effect more frequently than other hard sciences. In this vein, Burton and Morgan (2000, p. 435) claim that in this discipline "the apparent absence of the author from the text fits with positivist epistemologies in which the mathematician's role is subordinate to that of the mathematics itself", following, in Davis and Hersh's (1981, p. 36) words, "a purely mechanical procedure" – see (3) and (4) for an illustration.

> (3) In fact, df has maximal rank over Nsm *so that* f is a totally geodesic immersion with the induced flat metric there and in particular N0 = N00 (JDG 2016-3)
>
> (4) *Thus* a chain of indices as above gives rise to an (n + 1)/2-tuple of integer partitions δ(j) (CM 2016-2)

Finally, the finding that the proportion of cause expressions associated with grammatical patterns was larger in the hard-science dataset than in the soft-science one is in line with Hyland's (2008, p. 10) characterisation of technical subjects as being "routinely patterned" and formulaic. By contrast, the high productivity of cause/effect key nouns in the political science samples, alongside the small proportion of patterns with some key nouns, may be argued to be connected with the non-formulaic style of academic discourse in this discipline. Indeed, the high frequency of cause/effect nouns in political science is in line with Holmes' (1997) claim that results are presented in a more extended way in social sciences than in other disciplines.

6. Conclusion

To conclude, the findings in this study have revealed that the soft sciences are characterised by a larger number of cause expressions realised mainly by means of nouns which can be explained by the epistemic nature of the soft-science domain. It has also been found that the hard and soft sciences tend to employ different patterns of cause/effect devices with the hard sciences being more patterned than the soft ones.

The previous findings may shed light on the way in which English cause/effect expressions could be taught in academic disciplines and on the

design of discipline-specific language learning materials that might meet the needs of learners of different scientific fields more effectively. As pointed out by Becher (1989, p. 197), every science has its own "activities, knowledge structures, social structures, and communication patterns", which provide a complex and "purpose-driven rhetorical field for language learning and use". Since most teaching materials in EAP are not discipline-specific (Bennett, 2010; Boulton, 2012), EAP/ESP students would benefit from specific language resources (see, in this respect, Hardy & Römer, 2013, p. 184). To give an example, instruction addressed to learners majoring in mathematics could focus on the use of specific conjunctions in cause/effect expressions, whereas learners of political science discourse might benefit from being trained specifically in the use of cause prepositions and effect nouns. The teaching/learning process can also take advantage of Data Driven Learning (DDL) approaches, as suggested in, for example, Johns (1991, 1997), in which language learners are given access to linguistic data in the form of concordance lines in an attempt to meet their learning needs.

As far as the limitations of the current study are concerned, first, even though the significance of our results has been statistically verified, we are aware of the limited size of the corpus. Second, a look into additional disciplines is needed so as to provide a fuller picture of disciplinary variation in the use of cause/effect expressions, as well as a look into other science communication genres beyond the research article. In fact, Hyland (2004. p. 30) contends that the classification of sciences into hard and soft is not able to capture the complexity of disciplinary variation to the fullest. Third, the list of key cause/effect expressions analysed in this paper is based exclusively on the academic word list in LEAD, which could be extended by adding other less formal linguistic units that could be representative of (soft-/hard-science) academic discourse. Fourth, the distribution of cause/effect markers across articles has not been considered in the framework of this research, however, it should be noted that there might be differences in introductions and results sections, for instance. Investigating cause/effect expressions in various sections of a research article that are seen as self-contained rhetorical units could be a subject for further studies. To give an example, a fine-grained analysis of maths texts might reveal that certain cause/effect linguistic strategies are strongly associated with mathematical formulae.

Article history:
Received 01 May 2023
Received in revised form 31 October 2023
Accepted 01 November 2023

References

Abraham, E. (1991). Why 'because'? The management of given/new information as a constraint on the selection of causal alternatives. *Text - Interdisciplinary Journal for the Study of Discourse, 11*(3), 323-339. https://doi.org/10.1515/text.1.1991.11.3.323

Altenberg, B. (1984). Causal linking in spoken and written English. *Studia Linguistica, 38*(1), 20-69.

Anthony, L. (2014). *AntConc* (Version 3.4.4) [Computer Software]. Waseda University. https://www.laurenceanthony.net/software/antconc/

Becher, T. (1989). *Academic tribes and territories: Intellectual enquiry and culture of the disciplines.* The Society for Research into Higher Education & Open University Press.

Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers* (Vol. 10). University of Michigan Press. https://doi.org/10.3998/mpub.371534

Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.

Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press.

Boulton, A. (2012). Corpus consultation for ESP: A review of empirical research. In A. Boulton, S. Carter-Thomas & E. Rowley-Jolivet (Eds.), *Corpus-informed research and learning in ESP: Issues & application* (pp. 261-291). John Benjamins. https://doi.org/10.1075/scl.52

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press. https://doi.org/10.1017/9781316410899

Burton, M., & Morgan, C. (2000). Mathematicians writing. *Journal for Research in Mathematics Education, 31*(4), 429-453. https://doi.org/10.2307/749652

Cao, M., Sun, X., & Zhuge, H. (2018). The contribution of cause-effect link to representing the core of scientific paper – The role of semantic link Network. *PloS One, 13*(6), e0199303. https://doi.org/10.1371/journal.pone.0199303

Chukharev-Hudilainen, E., & Saricaoglu, A. (2016). Causal discourse analyzer: Improving automated feedback on academic ESL writing. *Computer Assisted Language Learning, 29*(3), 494-516. https://doi.org/10.1080/09588221.2014.991795

Davis, P. J., & Hersh, R. (1981). *The mathematical experience.* Birkhauser.

Degand, L. (2000). Causal connectives or causal prepositions? Discursive constraints. *Journal of Pragmatics, 32*(6), 687-707. https://doi.org/10.1016/S0378-2166(99)00066-1

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. https://doi.org/10.1016/S0889-4906(97)00014-8

Gilquin, G., & Paquot, M. (2007). Spoken features in learner academic writing: Identification, explanation and solution. In M. Davies, P. Rayson, S. Hunston & P. Danielsson (Eds.), *Proceedings of the Fourth Corpus Linguistics Conference CL2007* (pp. 27-30). University of Birmingham.

Gopalan, S., & Devi, S. L. (2017). Cause and effect extraction from biomedical corpus. *Computación y Sistemas, 21*(4), 749-757. https://doi.org/10.13053/cys-21-4-2854

Granger, S., & Paquot, M. (2010). Customising a general EAP dictionary to meet learner needs. In S. Granger & M. Paquot (Eds.), *Lexicography in the 21st century: New challenges, new applications. Proceedings of ELEX2009* (pp. 87-96). Presses universitaires de Louvain.

Granger, S., & Paquot, M. (2015). Electronic lexicography goes local: Design and structures of a needs-driven online academic writing aid. *Lexicographica – International Annual for Lexicography, 31*(1), 118-141. https://doi.org/10.1515/lexi-2015-0007

Groom, N. (2005). Pattern and meaning across genres and disciplines: An exploratory study. *Journal of English for Academic Purposes*, *4*(3), 257-277. https://doi.org/10.1016/j.jeap.2005.03.002

Gruiţă, M. (1983). On causal connectives in English and Romanian. *Revue Roumaine de Linguistique, 28*(1), 53-62.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.

Hardy, J. A., & Römer, U. (2013). Revealing disciplinary variation in student writing: A multi-dimensional analysis of the Michigan Corpus of upper-level student papers (MICUSP). *Corpora, 8*(2), 183-207. https://doi.org/10.3366/cor.2013.0040

Holmes, R. (1997). Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes*, *16*(4),

321-337. https://doi.org/10.1016/S0889-4906(96)00038-5

Hunston, S., & Francis, G. (2000). *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. John Benjamins. https://doi.org/10.1075/scl.4

Hyland, K. (1999). Disciplinary discourses: writer stance in research articles. In C. N. Candlin & K. Hyland (Eds.), *Writing: texts, processes and practices*. Routledge. https://doi.org/10.4324/9781315840390

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

Hyland, K. (2002). Directives: Argument and engagement in academic writing. *Applied Linguistics, 23*(2), 215-239. https://doi.org/10.1093/applin/23.2.215

Hyland, K. (2004). Disciplines and discourses: Social interaction in the construction of knowledge. In A. Arbor (Ed.), *Disciplinary discourses: Social interactions in academic writing* (pp. 193-214). University of Michigan Press.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies, 7*(2), 173-192. https://doi.org/10.1177/1461445605050365

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21. https://doi.org/10.1016/j.esp.2007.06.001

Hyland, K., & Jiang, J. J. (2018). 'In this paper we suggest': Changing patterns of disciplinary metadiscourse. *English for Specific Purposes, 51*, 18-30. https://doi.org/10.1016/j.esp.2018.02.001

Johns, T. F. (1991). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal, 4*, 1-16.

Johns, T. F. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100-115). Longman. https://doi.org/10.4324/9781315842677

Jordan, R. R. (1999). *Academic writing course: Study skills in English.* Pearson Education.

Kac, M. B. (1972). Clauses of saying and the interpretation of *because*. *Language, 48,* 626-632.

Krogsrud, B. (1980). Factors motivating the position of finite adverbial clauses introduced by *as*, *because*, *since.* MA thesis, Oslo University.

Li, F. (2014). *Subjectivity in Mandarin Chinese: the meaning and use of causal connectives in written discourse*. PhD dissertation. Utrecht University.

Li, F., Evers-Vermeul, J., & Sanders, T. J. M. (2013). Subjectivity and result marking in Mandarin. *Chinese Language and Discourse, 4*(1), 74-119. https://doi.org/10.1075/cld.4.1.03li

Linderholm, T., Everson, M. G., Van Den Broek, P., Mischinski, M., Crittenden, A., & Samuels, J. (2000). Effects of causal text revisions on more- and less-skilled readers' comprehension of easy and difficult texts. *Cognition and Instruction, 18*(4), 525-556. https://doi.org/10.1207/S1532690XCI1804_4

Louvain EAP Dictionary (LEAD). https://leaddico.uclouvain.be/

Ma, H., & Qian, M. (2020). The creation and evaluation of a grammar pattern list for the most frequent academic verbs. *English for Specific Purposes, 58,* 155-169. https://doi.org/10.1016/j.esp.2020.01.002

Marshman, E. (2004). The cause-effect relation in a French-language biopharmaceuticals corpus: Some lexical knowledge patterns. *The Computational and Computer Assisted Terminology Workshop, LREC 2004* (pp. 40-43). http://lrec.elra.info/proceedings/lrec2004/ws/ws9.pdf#page=44

Marshman E., & L'Homme M. C. (2006). Disambiguating lexical markers of cause and effect using actantial structures and actant classes. In H. Picht (Ed.), *Modern approaches to terminological theories and applications* (pp. 261-285). Peter Lang.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics* (2nd ed.). Edinburgh University Press.

Mel'čuk, I., Clas, A., & Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire.* AUPELF-UREf & Éditions Duculot.

Oxford English Dictionary. https://www.oed.com.

Palmer, F. R. (2001). *Mood and modality* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139167178

Pander Maat, H., & Sanders, T. (2001). Subjectivity in causal connectives: An empirical study of language in use. *Cognitive Linguistics*, *12*(3), 247-274. https://doi.org/10.1515/cogl.2002.003

Paquot, M. (2010). *Academic vocabulary in learner writing*. Continuum.

Paquot, M. (2012). The LEAD dictionary-cum-writing aid: An integrated dictionary and corpus

tool. In S. Granger, & M. Paquot (Eds.), *Electronic lexicography* (pp. 163-186). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199654864.003.0009

Rutherford, E. (1970). Some observations concerning subordinate clauses in English. *Language, 46,* 97-115.

Sager, J. C., Dungworth, D., & MacDonald, P. F. (1980). *English special languages: Principles and practice in science and technology.* Oscar Brandstetter Verlag KG.

Sanders, T. J., & Spooren, W. P. (2015). Causality and subjectivity in discourse: The meaning and use of causal connectives in spontaneous conversation, chat interactions and written text. *Linguistics, 53*(1), 53-92. https://doi.org/10.1515/ling-2014-0034

Storer, N. W. (1967). The hard sciences and the soft: Some sociological observations. *Bulletin of the Medical Library Association, 55*(1), 75-84.

Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620904

van Dijk, T. A. (1977). *Text and context.* Longman.

Winkworth, T., Tredgold, T., & Glynn, J. (1858). *The Journal of the Society of Arts, 6*(310), 697-706.

Winter, E. (1977). A clause-related approach to English texts. *Instructional Science, 6,* 1-92.

Winter, E. (1982). *Towards a contextual grammar of English.* George Allen & Unwin.

Whalley, P. (1981). A partial index of text complexity involving the lexical analysis of rhetorical connectives. *ALLC Journal, 2,* 55-60.

Javier Pérez-Guerra is a Professor in the English Department of Universidade de Vigo, Spain, where he coordinates the Language Variation and Textual Categorisation research group. His areas of specialisation are information packaging in the clause, multidimensional approaches to register variation, grammatical variation in diachrony and the impact of performance preferences and ease of processing on the design of grammars.

Elizaveta Smirnova is currently an Associate Professor in the Foreign Languages Department at the HSE University, Perm, Russia. Her research interests include but are not limited to corpus linguistics, functional grammar, academic discourse, and teaching academic writing. Her works have been published in *Lingua, Journal of English as a Lingua Franca, Pragmatics and Society*.

Elena Kostareva is an Associate Professor in the Foreign Languages Department at the HSE University in Perm, Russia. Her research interests are discourse analysis, with emphasis on EFL and ELT situations and cultural issues, English for specific purposes, professional communication, and cognitive linguistics.

Appendix A. Journals

Science type	Disciplines	Journals
Hard	Chemistry	*Cell Chemical Biology (CCB)* *Chem* *Chemical Science (CS)* *Trends in Analytical Chemistry (TrAC)*
	Engineering	*Automatica (Auto)* *Materials Characterisation (MC)* *International Journal of Engineering Science (IJES)* *Engineering (Engin)*
	Mathematics	*Compositio Matematica (CM)* *The Journal of Differential Geometry (JDG)* *Acta Mathematica (ActaM)* *Applied Mathematics and Computation (AMC)*
	Physics	*Physics Letters B (PL)* *Reviews in Physics (RP)* *European Physical Journal C (EPJ)* *Nuclear Physics B (NPh)*
Soft	Business	*The Journal of Management (JM)* *The Journal of Management Studies (JMS)* *Academy of Management Annals (AMA)* *Journal of Business Research (JBR)*
	History	*Contemporary European History (CEH)* *The Journal of Modern History (JMH)* *Journal of Global History (JGH)* *History of the Family (HF)*
	Linguistics	*Applied Linguistics (AL)* *Lingua (Ling)* *Modern Language Journal (MLJ)* *Language in Society (LS)*
	Political science	*Political Analysis (PA)* *World Politics (WP)* *American Journal of Political Science (AJPS)* *British Journal of Political Science (BJPS)*

Appendix B. List of patterns and examples

ORIGIN

[det origin of n v-link]:

> *the origin of the coherence is* more complicated than "pure" vibrations or electronic dynamics… (Chem 2016-1)

[v-link of origin]:

> the coherence observed *is purely of vibrational origin* without the involvement of other nearby excited states… (Chem 2016-1)

[v det origin of n]:

> Such a reputational logic has been used *to explain the origins and escalation of conflicts*, from the Peloponnesian War to the Vietnam War… (WP 2016-4)

FACTOR

[det factor that/which v]:

> …I treat the level of hierarchy as an *exogenous factor that determines* the propensity of a country for

civil war… (WP 2016-3)

[det factor in n/-ing]:

…phonetic cues as contrastive stress have been pointed out as *another factor in determining* the interpretation of pronouns (Ling 2016-1)

Cause

[det cause of n/-ing]:

We tried to further verify whether ceramide accumulation is *the main cause of* apoptosis by treating the cells with different ceramide synthesis inhibitors (CCB 2016-4)

[det cause for n]:

This alone is *a cause for concern* and gives us reason to view the results of some of the analyses with caution (PA 2016-1)

Reason

[det reason for n/-ing]:

The reason for the lack of UfaA activity is unclear (CCB 2016-1)

[for det reason]:

For this reason, we opted to expedite our preliminary [15N]-incorporation studies… (Chem 2016-3)

[v reason to-inf]:

…there *are reasons to believe* that many tumor-specific carbohydrate antigens including T antigen and PSA have a number of attributes (CCB 2016-3)

[det reason why clause v-link]:

The main reason why Assumption ((cid:63)) is not known to hold is that potential semi-stable reduction is not known… (CM 2017-1)

Source

[det source of n]:

This analysis presumes that *the source of the constraints* on have-cliticisation is ultimately UG, namely, the TP Projection Principle (Ling 2016-4)

[n v as det source of n]:

For smaller firms, in contrast, acquisition will be more critical as *a source of new technologies…* (JM 2016-2)

Consequence

[as a consequence of n]:

Moreover, the finding that 16:1n-9 derives from oleic acid, a major cellular fatty acid, does not mean the former is merely produced *as a consequence of an excess* of the latter (CCB 2016-5)

[as a consequence]:

As a consequence, many research groups have tried to exploit FASN as a target for cancer (CCB 2016-4)

[v-link det consequence of n]:

This comparison demonstrates that the features in Figures 3A and S3 *are not a consequence of spectral filtering* (Chem 2016-1)

[det consequence of n/-ing v]:

The consequences and adaptive responses of acute or chronic inhibition of essential enzymes such as FASN are not fully understood (Chem 2016-1)

[v det consequence for n]:

…cultural differences between alliance partners reduce trust and lead to coordination difficulties that *have negative consequences for alliance performance* (JMS 2016-1)

Effect

[have det effect on n]:

Temperature *has many different effects on limestone* (MC 2017-4)

[det effect of n/-ing on n]:

Gónzalez-Gómez et al. studied *the effects of thermal degradation on the compression strength* (MC 2017-4)

Implication

[det implication v-link]:

The implication is that the benefits of raising m above five will not outweigh the costs in terms of extra computation time (PA 2016-1)

[det implication of n]:

Another implication of institutional complementarity is that an MNE's response to an institutional setting in a host country is not a single decision (JMS 2016-5)

[det implication for n]:

Besides *this implication for relational fairness theory and research*, the moderator role that we found for self-efficacy for voice is also interesting for the self-efficacy literature (JM 2016-3)

Outcome

[det outcome of n]:

These perturbations do not appear to affect *the outcome of the experiments* (CCB 2016-2)

[det outcome v-link]:

The outcome is a similarity map for additives given in Fig. 3 (Ling 2016-5)

Result

[as a result of n]:

…workers with concealable or invisible disabilities may have unique interpersonal experiences *as a result of their ability* to use more identity management strategies in the workplace (JM 2016-5)

[link-v det result of n]:

The straight and ragged boundaries *are the result of coherent crystallographic twinning* of the primary dendrite trunk (MC 2016-1)

[in result of n]:

> …closer look at the literature devoted to this area of science su□ces to reveal a scarcity of complete data related to the analysis of initiation and propagation of microcracks and cracks developing *in result of the impact of complex stress states* on concrete (MC 2017-2)

[as a result]:

> *As a result,* successful diamination of unactivated alkenes typically requires adequate electron-withdrawing, protecting groups to suppress undesired strong ligation… (Chem 2017-3)

[det result link-v]:

> *This result is* undoubtedly well-known to experts, and the proof was explained to the author by Chenyang Xu (JDG 2016-5)

Contribute

[contribute to n]:

> Two major factors *have contributed to the lack* of development of asymmetric alkene diamination with protection-free alkylamine (Chem 2017-3)

[contribute to-inf]:

> Consequently, both improvements can *contribute to boost* the energy density of AORFBs (Chem 2017-1)

Lead

[lead to n/-ing]:

> Careful pretesting *should lead to more successful manipulations* of emotion (PA 2016-2)

[lead n to-inf]:

> A state's behavior in one dispute (yielding to Hitler's demands at Munich, for example) *leads potential adversaries and allies to make* inferences about its likely behavior in future disputes (WP 2016-4)

[lead towards n/ing]:

> These conditions *lead towards a determining system* which could be solved in several cases (AMC 2018-1)