

## Using corpus linguistics to investigate mathematical explanation

Juan Pablo Mejía-Ramos, Lara Alcock, Kristen Lew, Paolo Rago, Chris Sangwin, and Matthew Inglis

**Abstract:** In this chapter we use methods of *corpus linguistics* to investigate the ways in which mathematicians describe their work as explanatory in their research papers. We analyse use of the words explain/explanation (and various related words and expressions) in a large corpus of texts containing research papers in mathematics and in physical sciences, comparing this with their use in corpora of general, day-to-day English. We find that although mathematicians do use this family of words, such use is considerably less prevalent in mathematics papers than in physics papers or in general English. Furthermore, we find that the proportion with which mathematicians use expressions related to ‘explaining why’ and ‘explaining how’ is significantly different to the equivalent proportion in physics and in general English. We discuss possible accounts for these differences.

Key words: corpus linguistics, mathematical language, mathematical explanation.

### I. Corpus linguistics

Corpus linguistics is a methodological approach that involves analysing large collections of naturally occurring texts, known as corpora. Its methods can be used to investigate many types of linguistic questions. Before reporting our study on the notion of explanation in mathematics and physics research papers, we briefly outline the basic concepts of this approach. This outline of corpus linguistics falls into three parts, each focusing on an important stage of conducting a corpus analysis: assembling a corpus, processing raw text to render it suitable for analysis, and deciding upon an analytical approach.

#### 1.1 Assembling a corpus

A corpus is simply a large collection of machine-readable texts designed to represent some broader body of natural language. In theory, any text could be considered a corpus, but the term is normally reserved for a set of texts carefully sampled to be representative of a larger body of language. For example, while we might analyse the complete *Diary of Samuel Pepys* with a view to understanding linguistic features of Pepys’s writing, we would not consider it a corpus representative of the writing of 17th century England: such a generalisation would be problematic because we would not know whether a particular linguistic feature was characteristic of the period’s writing generally, or only of Pepys’s writing.

The desire for generalization arises because corpus linguists are most commonly interested in understanding the properties of some broad body of language, such as broadsheet newspaper articles or political speeches. As it would be difficult to collect the text of every political speech ever made, a first consideration is how to obtain as *representative* a sample as possible (Biber, 1993). This gives rise to important issues of sampling that parallel those of traditional empirical research. Just as experimental psychologists ideally seek to sample participants randomly from their population of

interest, corpus linguists ideally seek to sample texts randomly from the wider set of texts to which they would like to generalize. The population is referred to as the ‘sampling frame’, and sometimes genuine random sampling can occur: given access to an appropriate archive, it would be possible to randomly select 10% of all newspaper articles published in a given time period. But corpus linguists are often interested in a less accessible sampling frame, which makes it difficult to randomly sample. For example, in the current investigation we would like to generalize to all research-level mathematical writing. But randomly sampling from this population would be difficult, as some writing in the population is inaccessible. In such situations we must instead appeal to the representativeness of our corpus.

One common approach to ensuring adequately representative sampling is to use a bibliographic index. For instance, researchers might define the sampling frame to be every text included in a particular list of published texts. The Lancaster-Oslo/Bergen (LOB) corpus, designed to be representative of general written British English, took this approach and used the British National Bibliography and Willings’ Press Guide as indices (Johansson et al. 1978). Alternatively, it is possible to sample participants rather than texts. For instance, the British National Corpus (a comprehensive collection of 100 million words of spoken and written English, designed to represent a cross-section of current English usage) contains a spoken section where participants – selected using demographic sampling techniques – were asked to record their day-to-day spoken interactions for several days (Crowdy 1993). In either case, representativeness might be further ensured via hierarchical or stratified sampling approaches. A sample might be composed of 10% of texts from one sub-category, 10% from another, and so on. The Brown and LOB corpora both adopted this approach in an attempt to be representative of American and British English respectively. Each contains 500 texts sampled from 15 categories (e.g. press reportage, popular lore, general fiction, science fiction, learned and scientific writing).

A second consideration when assembling a corpus is size. The required size depends on the linguistic feature being studied: if the feature is relatively rare, then a much larger corpus will be needed. Biber (1993), for instance, gave the relative frequencies of various linguistic features in a particular corpus, noting that conditional subordination occurred 2.5 times per 1000 words, whereas prepositions occurred 111 times. Clearly, this means that a larger corpus is required to study conditional subordinations than prepositions. Fortunately, creating extremely large samples of texts has recently become considerably easier, and corpora have been constructed based on webpages (e.g. various corpora based on Wikipedia articles), on television subtitles (e.g. the SUBTLEX-UK corpus; Van Heuven et al. 2014), and on parliamentary proceedings (e.g. the Hansard corpus; The SAMUELS Consortium, 2015).

A third consideration concerns dispersion. This refers to how evenly distributed a linguistic feature is across texts in a corpus. If a feature appears in only few texts, perhaps written by only a few authors, then this calls into question the generalisability of any claimed results, even if the corpus is reasonably representative in general (McEnery and Wilson, 2001).

In the study we describe below, our decisions concerning the sampling frame meant that we were able to assemble a large corpus of mathematical texts (approximately 31m words) as well as a control corpus of physics texts (approximately 59m words).

We then addressed considerations of representativeness by following good practice from empirical research in psychology: we assembled two further corpora of mathematics and physics texts of approximately the same size and from the same source, thus allowing us to replicate all our analyses on a new dataset. This should enable the reader to feel confident that we did not conduct a large number of analyses and report only those which gave statistically significant results (*cf.* John et al. 2012, and Simmons et al. 2011, on *p*-hacking).

### *1.2 Processing the corpus*

Assembling a novel corpus, or selecting an existing corpus, is only the first stage of a corpus linguistics research project. Often it is necessary to process texts in some way before proceeding with the analysis. In general contexts it is often important to annotate a corpus with tags relevant to the researchers' questions. This might involve grammatical tagging (often called part-of-speech, or POS tagging) where each word in the corpus is tagged with a label that categorizes it in some way. For instance, it may be useful to know if a word is an adjective, noun, or adverb (and so on). One common way of tagging a corpus is to append the tag after each word (e.g. replacing "cutting" with "cutting\_NN" where the "NN" represents the code 'noun, singular or mass'). Leech (2013) pointed out several important features an annotation scheme should have. First, it should always be possible to remove the tags and revert to the raw corpus. Second, the tags should be extricable from the corpus if necessary (e.g. it should be possible to count the number of nouns in a corpus). Third, the annotation scheme should be carefully documented. Leech also emphasized that the quality of the annotation should be documented. If, for example, a computer-based POS tagger is used (e.g. TagAnt), it might be appropriate to manually check a sample of the corpus and record the agreement percentage. Understandably, automated computer-based POS tagging is a complex process (for a review see Garside et al. 2013).

POS tagging has been successfully applied to mathematical corpora. For instance, Dawkins et al. (2018) used a corpus of university-level mathematics textbooks and presented a comparative analysis of the use of 'is' in mathematical and non-mathematical English. Because they were interested in advanced scientific texts that contain complex mathematical notation, this raised particular issues about the processing of LaTeX-encoded mathematical symbols. Our approach to this issue is discussed in the methods section below.

### *1.3 Analysing the corpus*

Once a corpus has been successfully assembled and processed, the next step is to choose an analysis technique that addresses the research question. Of course, there are many possible analytical approaches; here we give a brief overview of only the most common.

Often it is possible to answer research questions by simply studying the frequency with which certain words or phrases occur. Mejía-Ramos and Inglis (2011), for instance, used this approach to analyse 'semantic contamination' from day-to-day English into mathematical language. Semantic contamination refers to the phenomenon in which the meanings of words in natural language 'leak' into a different linguistic register (e.g. Monaghan, 1991). Mejía-Ramos and Inglis compared the frequency of the verb and noun forms of the word 'proof' (i.e. 'prove' and

‘proof’) in the specialist (business, medical, legal proceedings, etc.) and informal (conversations, popular radio, etc.) sections of the British National Corpus. They found that the verb form was significantly more common in informal than in formal language, and derived the hypothesis that ‘proof’ was more likely to be associated with the notions of formal validity, whereas ‘prove’ was more likely to be associated with the less formal notion of conviction. In two subsequent experimental studies, they found that changing a question from “does the argument prove the claim?” to “is the argument a proof of the claim?” did indeed change students’ responses in the direction predicted.

In the study reported here, our interest in the relative frequency of a particular set of words (which we will refer to as ‘explain words’) in mathematical writing means that our main analysis too involved defining which words fell into our category and counting their occurrences. By also conducting the same analysis on a different corpus (of physics papers) we were then able to compare the frequencies in two subgenres of research papers. Clearly, when comparing frequencies in corpora of different sizes it is necessary to adopt a frequency rate measure; the number of occurrences per million words is typically used. Counting the (relative) frequencies of a category of words in two corpora generates a two-by-two contingency table, with the two corpora as rows and the hits and non-hits (words in the corpus that are and are not in the category) as columns. This permits use of a chi-squared test or Fisher’s exact test to assess statistically whether the relative frequencies differ significantly between the two corpora. Our analysis is reported below.

As well as counting words or categories of words, it is also possible to study the frequencies of more complex linguistic features. For instance, one might be interested in producing a list of the most frequently occurring ‘n-grams’ or ‘lexical bundles’ – collections of words of a given length. Or one might be interested in producing a frequency list of ‘clusters’ – collections of words of a given length that contain a given word. Herbel-Eisenmann, Wagner and Cortes (2010) used the notion of a lexical bundle to study common interaction patterns in mathematics classrooms. Using a corpus formed from transcripts of interactions in secondary mathematics lessons, they found that there are particular types of lexical bundles involved in teacher/student interactions that allow the communication of feelings, attitudes, value judgements and assessments. By comparing their findings with other corpora (of, for instance, university classes) the researchers were able to argue that their findings were particular to the secondary mathematics context.

An alternative way to compare corpora is to identify keywords: words that occur disproportionately often in one corpus compared to another, but that have not been identified a priori by the researcher. These can be identified using chi-squared statistics in a similar manner to the approach described earlier. For instance, we can compare British and American English by identifying the keywords in the LOB corpus compared to the Brown corpus (organized by chi-squared values, these are: London, labour, I, sir, Mr, she, towards, Britain, British and centre). Similarly, we can identify keywords in the Brown corpus compared to the LOB corpus (program, toward, state, states, center, defense, federal, labor, York and American). By identifying the keywords in one corpus with respect to another, researchers can begin to understand differences between the bodies of language represented by each of the corpora.

More qualitative ways to analyse corpora can help researchers to understand how given words are used. For instance, most corpus linguistics software packages allow examination of ‘concordances’ or ‘key words in context’ (KWIC). The packages generate lists containing every occurrence of a given word – the search term – with context on either side (perhaps 80 characters to either side of the occurrence). By carefully studying a concordance, or a randomly selected subset of a concordance, researchers can begin to develop categories capturing how the word is used. This can subsequently support further quantitative analysis, especially if multiple equivalent corpora are available (e.g. a concordance analysis can be conducted on one corpus and then a quantitative analysis can be used on the other to triangulate).

Similarly, packages can also permit examination of words that systematically co-occur. For instance, ‘back’ and ‘front’ are often found close to each other, and corpus linguists would say that ‘back’ is a collocate of ‘front’ (and vice versa). More formally, two words are collocates if there is an above-chance co-occurrence of them within some given span (perhaps plus or minus five words). Collocates can be identified by constructing a word frequency list of all words within a five-word window around the search term, and comparing it to the overall word frequency list of the corpus. Words that disproportionately occur around the search term are its collocates (various statistical criteria can be used to formalize what ‘disproportionately’ means). Understanding the collocates of a given word can help reveal its meaning, and perhaps uncover implicit associations that it has with other words or ideas (Hunston, 2002; Sinclair, 1991). In particular, studying the collocates of a word can identify a word’s ‘semantic prosody’, the “consistent aura of meaning with which a form is imbued by its collocates” (Louw, 1993, p. 157). Baker, Gabrielatos, Khosravini, Krzyżanowski, McEnery and Wodak (2008) used this method to study representations of refugees and asylum seekers in British newspapers, finding that collocates were often words that may negatively stereotype refugees and asylum seekers. For example, their collocate analysis showed how references to refugees and asylum seekers were often accompanied by quantification via water metaphors (e.g. pour, flood, stream), which “tend to dehumanize [refugees and asylum seekers], constructing them as an out-of-control, agentless, unwanted natural disaster.” (p. 287) In other words, the words ‘refugee’ and ‘asylum seeker’ have negative semantic prosody in British newspapers.

The outline of corpus linguistic research methods given here is necessarily basic. Readers interested in conducting corpus analyses might wish to consult McEnery and Wilson’s (2001) excellent textbook for further information. We now turn to our investigation of the notion of explanation in mathematics and physics research papers.

## **II. Mathematical Explanation**

Explanations are important, nowhere more so than in education. Teachers routinely offer instructional explanations as part of classroom practice, answering implicit or explicit questions posed by their students or themselves (e.g. Leinhardt, 2001; Treagust and Harrison, 1999). Also important are self-explanations generated by learners with the aim of increasing their own understanding (Rittle-Johnson et al. 2017). Encouraging students to self-explain can be a highly effective pedagogical strategy: in the context of university level mathematics, Hodds et al. (2014) found that students prompted to explain a mathematical proof to themselves attained comprehension one standard deviation better than peers in a control group (*cf.* Chi et

al.1989, Fonseca and Chi 2011). Finally, student-generated explanations can be used in educational assessment, particularly when one wishes to focus on the depth of students' conceptual understanding (e.g. Bisson et al. 2016; Knuth et al. 2006). But what are explanations, especially in mathematics?

Philosophers of science have devoted considerable attention to the question of what it means for A to explain B. Many accounts rely on either statistical associations or causal mechanisms. For instance, Salmon (1971, 1984) suggested that A explains B if B is consistently correlated with A or if there is a causal history that connects B and A (*cf.* Hempel and Oppenheim, 1948). So one can say that buying shoes in the wrong size explains why one's feet hurt because there is a causal connection between the two events. But, while causal and statistical accounts work well in scientific contexts, they fail in mathematics (Colyvan 2011, Mancosu 2001). Mathematical concepts are not related causally because there is no temporal order: the fact that the square root of 2 is irrational is not located at a particular time. Nor are they related statistically: mathematical facts take no probabilities other than 0 or 1. Consequently, scientific accounts of explanation do not seem to apply to mathematics.

But if mathematical explanations are not scientific explanations, then what are they? This question has generated significant philosophical interest. A small number of philosophers regard the lack of causal and correlational relations as reason to deny that mathematical explanations exist (Resnik and Kushner 1987, Zelcer 2013), arguing that there is little empirical evidence to suggest that explaining is central to mathematicians' practice. However, others vociferously dispute this (e.g. Colyvan 2011, Weber and Frans, 2016). The dispute seems to turn on the extent to which practicing mathematicians use the notion of explanatory value in their own work. For example, Steiner (1978) claimed that "mathematicians routinely distinguish proofs that merely demonstrate from proofs which explain" (p.135). In contrast, Resnik and Kushner (1987) claimed that mathematicians "rarely describe themselves as explaining" (p. 151). Hafner and Mancosu (2005) responded by stating that "[c]ontrary to what Resnik and Kushner claim (p. 151), mathematicians *often* describe themselves and other mathematicians as explaining" (p. 223, emphasis in the original). Hafner and Mancosu (2005) supported this claim by presenting several examples of what they called *explanatory talk* in mathematical practice: passages of research mathematics papers in which the authors explicitly describe themselves or some piece of mathematics as *explaining* a given "mathematical phenomenon". While this evidence is not sufficient to settle the disagreement, the specific cases discussed by Hafner and Mancosu have been interpreted in significantly different ways:

"I believe that detailed case studies, such as those by Hafner and Mancosu (2005), *decisively* refute Resnik's and Kushner's [claim]" (Lange 2009, p. 203, our emphasis).

"Though philosophers have lately been pointing out some exceptions, the examples tend to be rather exotic (e.g., in Hafner and Mancosu 2005). There has been no systematic analysis of standard and well-discussed texts illustrating any pattern of mathematical explanations." (Zelcer 2013, 179-180).

Clearly – and contrary to Lange's (2009) suggestion – it is impossible to decisively refute a claim that a given event is rare by identifying one or more instances of the event occurring. Instead, a systematic analysis of the type suggested by Zelcer (2013)

is needed. Attempts in this direction have been made, in scientific fields if not in mathematics. Overton (2013), for instance, analysed all regular articles published in the journal *Science* in a one-year period (a total of 781 papers and approximately 1.6 million words). He searched for all ‘explain words’ (defined to be: *explain, explains, explained, explaining, explainable, explanation, explanations, unexplained, unexplainable, explicate, explicates, explicated, explicable, inexplicable*) and compared their frequencies to those of words of other types. Overton found that approximately 45% of the 781 papers contained at least one “explain” word (with an average of 0.96 “explain” words per article), and he concluded that: “The numbers for “explain” are perhaps surprisingly low if scientific journals are vehicles for explanations. [...] The observed frequencies of “explain” words suggests that explanation is only moderately important in science.” (p. 1387). This low frequency of ‘explain words’ in articles in science—a field in which explanation is widely regarded as playing a central role—might warrant Zelcer’s (2013) scepticism of the predominance of explanatory talk in mathematics – a field in which even the existence of explanation is debated. One goal of this chapter is to shed light on this decades-old dispute among philosophers concerning the frequency with which mathematicians describe themselves or their mathematical work as explaining other mathematics. Like Overton (2013), we do this by analysing large collections of text.

A second goal of this chapter is to explore the types of explanations mathematicians discuss in their explanatory talk. To date, analyses of mathematical explanations tend to differentiate between explanations of other mathematics (mathematics X explains mathematics Y, or X is an explanatory proof of theorem Y), and explanations of physical phenomena (mathematics X explains physical phenomenon Y). Colyvan (2011) referred to these as *intra-mathematical* and *extra-mathematical* explanations respectively. Hafner and Mancosu (2005) further differentiated between two uses of intra-mathematical explanations: those that are “instructions” on how to master the tools of the trade, explaining how to employ mathematical techniques, and those that “call for an account of the mathematical facts themselves, the reason why” (p. 217). While Hafner and Mancosu considered the latter to be a “deeper” use of mathematical explanation, others have emphasized the importance of the former in mathematical practice. For instance, Rav (1999) emphasized the mathematical methodologies and problem solving strategies/techniques contained in proofs, and insisted that one of the main reasons mathematicians read proofs is to glean this mathematical know-how:

“Proofs are for the mathematician what experimental procedures are for the experimental scientist: in studying them one learns of new ideas, new concepts, new strategies—devices which can be assimilated for one's own research and be further developed.” (p. 20)

This claim is consistent with empirical evidence from both small-scale interview studies and large-scale surveys asking mathematicians about their practice (Weber and Mejía-Ramos, 2011, Mejía-Ramos and Weber, 2014). But neither this claim nor the distinctions upon which it relies have been examined at scale in written mathematical research papers.

In this chapter, we address this issue, reporting a study that employs methods of corpus linguistics to address the following specific questions:

1. To what extent do mathematicians describe themselves (or their mathematical work) as explaining mathematical phenomena in their research papers, and how does this compare with descriptions of explanation in physics discourse and in general, day-to-day discourse?
2. How does the extent to which mathematicians describe themselves as explaining compare with the extent to which they describe themselves as engaging in related mathematical activities (such as *proving* theorems)?
3. To what extent do mathematicians describe themselves as explaining *why* a certain mathematical statement is true, as compared with explaining *how* to do something in mathematics?

### III. Methods

#### 3.1 Collecting the Texts

For our study, we needed a large sample of mathematics research papers, together with two comparison corpora: a large sample of research papers from another discipline, and corpora of general, day-to-day English. For our comparison disciplinary corpus, we collected physics research papers, and for our day-to-day English corpus we used both the British National Corpus (BNC) and the larger Corpus of Contemporary American English (COCA)<sup>1</sup>. To assemble our corpora of mathematics and physics research papers, we adopted two largely pragmatic criteria:

- 1) Text should be in LaTeX format to enable consistent processing (discussed below).
- 2) Text should be published non-commercially and freely available online.

Based on these criteria, we used research papers uploaded to the ArXiv (<https://arxiv.org/>). The ArXiv is an online repository of electronic preprints of scientific papers in mathematics, physics, astronomy, computer science, quantitative biology, quantitative finance, and statistics; it is one of the main repositories that mathematicians and physicists around the world use to share their work. We downloaded the bulk source files (mostly TeX/LaTeX) containing all papers uploaded to the ArXiv in the first eight months of 2009 (which provided us with a large enough sample of more than 30,000 research papers), then converted the source code to plain text for use with standard corpus analysis software (all analyses reported in this paper were performed using CasualConc, version 2.0.3).

#### 3.2 Processing the Texts

Converting mathematical language into a form that can be processed using the standard corpus linguistics software presents a challenge. Most professional mathematics is written using the TeX/LaTeX<sup>2</sup> markup language, not plain text. Our

---

<sup>1</sup> COCA contains more than 560 million words of spoken, fiction, popular magazines, newspapers, and academic texts.

<sup>2</sup> TeX was developed in the 1970s (Knuth, 1979) to enable digital typesetting of structured documents containing mathematics. Most professional mathematicians still write using TeX or the subsequent LaTeX markup language. TeX/LaTeX is written as plain text documents that include control codes to structure the document and codes to typeset the special symbolism used in mathematics. The system consciously separated the encoding of the document from the processing and production of the human readable (e.g. printed) output.



first goal was therefore to create a method of converting LaTeX source code to plain text in a way that preserved the natural sentence structure of the language, but removed non-linguistic features of the source code (the code “`\textbf{text}`” for bold text, for instance). We constructed scripts to achieve this, converting LaTeX to analysis-ready plain text.<sup>3</sup>

Another important question for the would-be creator of a mathematical corpus concerns how to deal with inline mathematical notation. For instance, a typical mathematical sentence might be “Let  $f: X \rightarrow Y$  be a bijection.” How should “ $f: X \rightarrow Y$ ” appear in a plain text corpus? One approach would be to leave the LaTeX source code intact and analyse the code as if it were natural language. The difficulty with this option is that there are several different ways in which one could encode “ $f: X \rightarrow Y$ ” in LaTeX. For instance, “`\$f:X\rightarrow Y\$`” and “`\(f:X\rightarrow Y\)`” produce identical output, and “`\$f\,:X\longrightarrow Y\$`” differs only stylistically (the spacing is wider and length of the arrow slightly longer). We therefore felt that retaining the LaTeX codes would be unhelpful for the majority of questions a researcher would wish to answer using a mathematical corpus. A second option would be to delete all mathematical code entirely, and record the example above as “Let be a bijection”. We rejected this option because failing to preserve the logical structures of sentences would influence certain analyses (those that investigate the collocation of words, for instance). Instead we opted to replace all occurrences of inline mathematics with the string “`inline_math`” (although this choice of string can be altered by users of our scripts if desired).

### 3.3 Analyzing the Corpus

With the source files processed, we sorted the articles using their primary subject classification (mathematics, physics, etc.) to assemble our two disciplinary corpora: one containing the processed text of all mathematics papers and the other containing the processed text of all physics papers. As noted above, one benefit of working with these large datasets is that a researcher can partition a large corpus into smaller samples that remain sufficiently large to conduct statistical analyses. This provides samples for both exploratory and confirmatory analyses: the researcher can perform initial analyses on one sample and then test whether the corresponding findings replicate when the same analyses are performed on a different sample. With this in mind, we split each disciplinary corpus into two smaller samples based on the month in which the papers had been uploaded: for each discipline, the first sample contained the papers uploaded in January-April 2009, and the second sample contained the papers uploaded in May-August 2009.

## IV. Results

Table 1 presents the number of physics and mathematics papers uploaded to the ArXiv in January-April and May-August of 2009, together with the number of words<sup>4</sup> in each set of papers. We notice that in those eight months, researchers uploaded

---

<sup>3</sup> These scripts are freely available for research purposes at <https://github.com/sangwinc/arXiv-text-extractor>.

<sup>4</sup> A word here is any string of characters between spaces. As discussed above, for these analyses we opted to replace all occurrences of inline mathematics with the string “`inline_math`” and count it as one word. For instance, the string “Let  $f: X \rightarrow Y$  be a bijection” in a paper, coded in LaTeX by the authors as “Let `\$f:X\rightarrow Y\$` be a bijection”, would have been translated to text as “Let `inline_math` be a bijection” and coded as having 5 words.

approximately 2.4 times as many physics papers as mathematics papers. We also notice that, on average, physics papers contained around 5,000 words, whereas mathematics papers contained around 6,200 words.

	January-April 2009		May-August 2009	
	#papers	#words	#papers	#words
Mathematics	5,087	30,892,695	4,970	31,289,569
Physics	11,787	58,859,660	12,370	62,807,075

Table 1. Number of papers and words in the physics and mathematics corpora

We used the January-April sample to address each of our research questions and the resulting analyses are presented in Sections 4.1-4.3. In Section 4.4, we briefly present the analyses of the replication using the May-August sample.

#### 4.1 Absolute and relative frequency of ‘explanatory talk’

Following Overton (2013), we defined ‘explain words’ to be 18 words linguistically related to the word *explain*<sup>5</sup>:

**‘Explain words’:** explain, explains, explained, explaining, explainable, explanation, explanations, explanatory, unexplained, unexplainable, explicate, explicates, explicated, explicating, explicable, inexplicable, explication, explications.

Table 2 shows the frequencies of ‘explain words’ in our corpus of 5087 mathematics papers and 11787 physics papers uploaded between January and April of 2009. ‘Explain words’ occurred a total of 4910 times in the mathematics papers (around 159 times per million words), an average of 0.97 times per paper, with 1898 of mathematics papers (approximately 37%) in this sample containing at least one ‘explain word’. This provides an existence proof of explicit explanatory talk in this corpus. In order to assess whether this is a large or small frequency, we conducted the same analysis on the physics corpus.

In comparison, ‘explain words’ showed up 21345 times in the corresponding set of physics papers (around 363 times per million words), an average of 1.81 times per paper, with 6499 of these papers (roughly 55%) containing at least one ‘explain word’. Thus, the number of ‘explain words’ per million in the physics papers is around 2.28 times that of the mathematics papers.

---

<sup>5</sup> Our aim was to be consistent with Overton’s (2013) analysis. However we decided to include the words *explanatory*, *explication*, *explications*, and *explicating* given their close relation to some of the 14 words in Overton’s (2013) original analysis (e.g. the original list included *explicate*, but not *explication*). However, as Table 2 shows, these additional words did not appear frequently in these corpora of research papers. Indeed, only five ‘explain words’ (*explain*, *explained*, *explanation*, *explains*, and *explaining*) make up more than 95% of all appearances of ‘explain words’ in each of these two corpora.

	Mathematics		Physics	
	frequency	per million	frequency	per million
explain	1827	59.14	7768	131.97
explained	1690	54.71	6513	110.65
explanation	498	16.12	3564	60.55
explains	484	15.67	1601	27.20
explaining	175	5.66	914	15.53
explanations	119	3.85	675	11.47
explanatory	51	1.65	62	1.05
unexplained	22	0.71	177	3.01
explication	13	0.42	4	0.07
explicated	10	0.32	15	0.25
explicate	6	0.19	5	0.08
explicating	5	0.16	0	0.00
unexplainable	4	0.13	8	0.14
explications	4	0.13	2	0.03
explainable	1	0.03	23	0.39
explicates	1	0.03	1	0.02
explicable	0	0.00	9	0.15
inexplicable	0	0.00	4	0.07
Total	4910	158.92	21345	362.63

Table 2. Frequency and frequency per million words of ‘explain words’ appearing in the January-April mathematics and physics papers

Figure 1 compares across corpora, displaying frequencies per million words of ‘explain words’ in the mathematics papers, the physics papers, and the two general English corpora. We note that while the number of ‘explain words’ per million in the physics papers (362.63) is roughly 1.4 times higher than that in COCA (250.97) or the BNC (260.01), the frequency of ‘explain words’ in these day-to-day English corpora is still around 1.6 times higher than that of the mathematics papers (158.92).

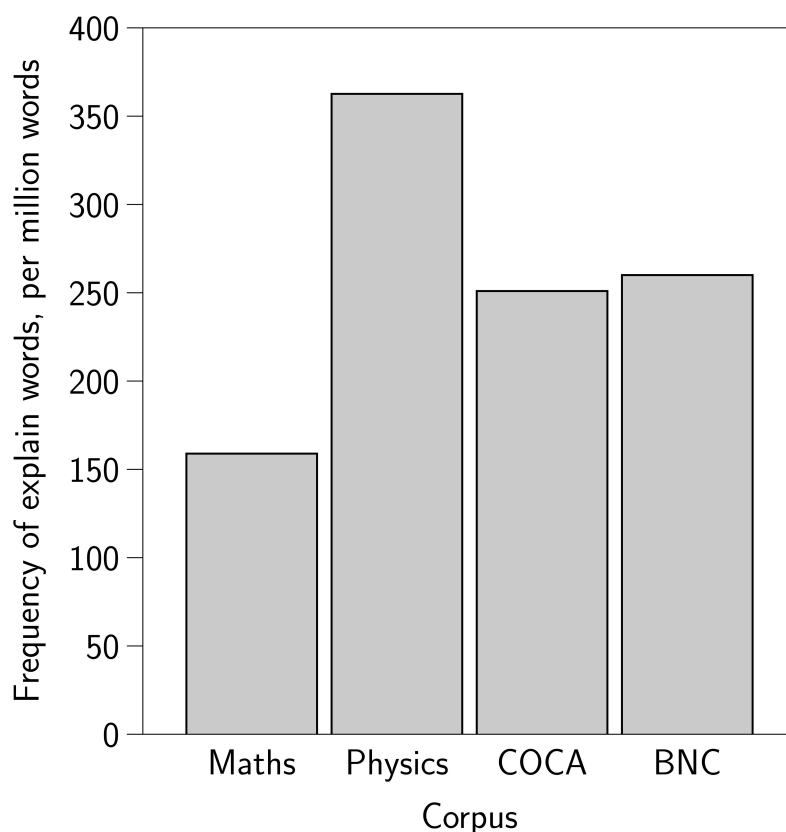


Figure 1. Frequency of ‘explain words’ (per million words) in the mathematics, physics, COCA, and BNC corpora.

#### 4.2 Explanation versus related notions

To assess the extent to which the observed frequencies of ‘explain words’ were high or low *within* mathematical discourse, we compared them against the frequencies of words related to other intuitively important mathematical activities. Table 3 presents the frequencies of words linguistically related to the notions of conjecturing, defining, modeling, proving, showing, and solving.

**‘Conjecture words’:**<sup>6</sup> *conjecture, conjectured, conjectures, conjectural, conjecturally, conjecturing.*

**‘Define words’:** *defined, define, definition, defines, definitions, defining, definable, undefined, redefine, redefined, definability, redefinition, definably, redefining, welldefined, definedness, interdefinable, predefined, redefinitions, interdefinability, redefines, definitional, definitionally, undefinability, undefinable.*

**‘Model words’:** *model, models, modeled, modeling, modelled, modelling, countermodel, submodel, submodels,*

<sup>6</sup> For each group, words are listed in order of frequency, with the most frequent words in the group listed first. The italicized words in each group make up 95% of all instances of words from that group appearing in the mathematics papers uploaded in the first four months of 2009.

modelized, modelization, modelisation, modelize, modelizing, countermodels, premodel, remodeled.

**‘Prove words’:** *proof, prove, proved, proves, proofs, proving, proven, provable, reprove, disprove, provability, provably, reproved, disproved, unprovable, unproven, reproving, disproving, reproves, prover, unproved, subproof, disproof, disproven, disproves, reproven.*

**‘Show words’:** *show, shows, shown, showed, showing.*

**‘Solve words’:** *solution, solutions, solve, solving, solvable, solved, solves, resolvent, solvability, subsolution, resolved, resolving, supersolution, resolve, solver, resolvents, unsolved, resolves, solvers, nonsolvable, supersolutions, subsolutions, unresolved, nonsolvable, unsolvable, cosolvable, equisolvable, supersolvable, unsolvability.*

	frequency	per million	per paper	in #papers	in %papers
define	124129	4018.07	24.40	4838	95%
prove	111838	3620.21	21.99	4710	93%
show	59359	1921.45	11.67	4691	92%
solve	53013	1716.04	10.42	3073	60%
model	23658	765.81	4.65	2013	40%
conjecture	8362	270.68	1.64	1413	28%
explain	4910	158.94	0.97	1898	37%

Table 3. Frequencies of words related to *explaining, conjecturing, defining, modeling, proving, showing, and solving* in the January-April mathematics papers. The last two columns provide the number of papers containing at least one word in that group and the percentage of such articles

Measured against these other frequencies, mathematicians used ‘explain words’ rather infrequently. For instance, mathematicians used ‘explain words’ in their papers approximately 12 times less frequently than ‘show words’ and nearly 23 times less frequently than ‘prove words’.

So far, our study of explanatory talk has investigated the use of ‘explain words’. This approach has the virtue of focusing on unambiguously explicit discussion of explanation, but could potentially leave unnoticed a significant amount of explanatory talk (i.e. mathematicians describing themselves or their work as explaining some mathematical phenomenon). For instance, the main case of explanatory talk discussed by Hafner and Mancosu (2005) highlights how a mathematician described one of his proofs as providing “the true reason why” a given mathematical phenomenon was the case. Clearly, this is a case of explanatory talk that does not use ‘explain words’. However, a difficulty of expanding our investigation to expressions that do not use ‘explain words’ is that these alternative expressions may not actually indicate the presence of explanatory talk. For instance, Overton (2013) argued that the use of words such as ‘because’ (which he found to be ubiquitous in his corpora) may not really indicate the presence of scientific explanations (p. 1387). Fortunately, Hafner

and Mancosu (2005) identified eight expressions that they found to be commonly used in the literatures of mathematics and philosophy of mathematics to describe the search for mathematical explanations. Table 4 presents these expressions along with the specific concordance search we made to investigate their prevalence in the mathematics and physics papers, and the frequencies with which these alternative expressions appeared. The total number of occurrences of these expressions is only about 10% of the total number of ‘explain words’ in each set of papers. Furthermore, there were disproportionately more occurrences of these expressions in physics than in mathematics, so this analysis does not materially affect the findings based on ‘explain words’ only. We suggest, therefore, that Hafner and Mancosu (2005) may have overestimated how common these expressions are in mathematics research papers. We are also left wondering to what extent such *common* alternative expressions exist.

Alternative expression	Concordance search <sup>7</sup>	Mathematics	Physics
"the deep reasons"	deep* reason*	5	16
"an understanding of the essence"	understand* of the essence	0	0
	understand* the essence	0	5
"a better understanding"	better understand*	161	767
"a satisfying reason"	satisfy* reason	0	0
"the reason why"	reason* why	312	924
"the true reason"	true reason*	3	1
"an account of the fact"	an account of the fact	0	0
"the causes of"	cause* of	16	609
Total		497	2322

Table 4. Frequencies of alternative expressions related to explanatory talk in the January-April mathematics and physics papers

### 4.3 Explaining why versus explaining how

To compare mathematicians’ propensity to describe themselves as *explaining why* a certain mathematical statement is true (Hafner and Mancosu’s “deep” explanation) with their propensity to describe themselves as *explaining how* to do something in mathematics (related to Rav’s notion of mathematical know-how), we created a concordance to identify every instance in which an *explain word* was followed immediately by the words *why* or *how*. We did this by searching the concordance for \*expla\* why and \*expla\* how, and checking that all results were indeed uses of ‘explain words’. We then repeated the process with the corpus of physics papers.

When taken together, the total of \*expla\*-why and \*expla\*-how expressions were roughly as common in math papers as they were in physics papers, with approximately 22 of these expressions showing up per million words in each set of papers; they formed a relatively small subset of the wider use of ‘explain words’ (roughly 14% and 6% of *explain word* usage in mathematics and physics, respectively). However, as shown in Table 5, the distributions of these two different types of expressions in the mathematics and physics papers differed significantly

<sup>7</sup> In concordance searches an asterisk can be used as a wildcard to find words (or expressions with words) that contain a particular string of characters, but with potentially different beginnings or endings. For instance, a search for “deep\* reason\*” would find the expressions “deeper reasons” and “deep reasoning”.

(Fisher's exact test,  $p < .001$ ). Mathematicians used nearly twice as many \*expla\*-how expressions as \*expla\*-why expressions; physicists used between two and three times as many \*expla\*-why expressions as \*expla\*-how expressions. Furthermore, general English is more similar to physics than to mathematics in use of these expressions. Figure 2 shows that the frequency of \*expla\*-why expressions in general English is around 15 per million words, and that \*expla\*-how expressions occur roughly half as frequently.

	Mathematics		Physics	
	frequency	per million	frequency	per million
*expla* why	247	7.99	952	16.17
*expla* how	458	14.83	353	6.00
Total	705	22.82	1305	22.17

Table 5. Frequencies and frequencies per million words of 'explain words' immediately followed by the words *why* or *how* in the January-April mathematics and physics papers.

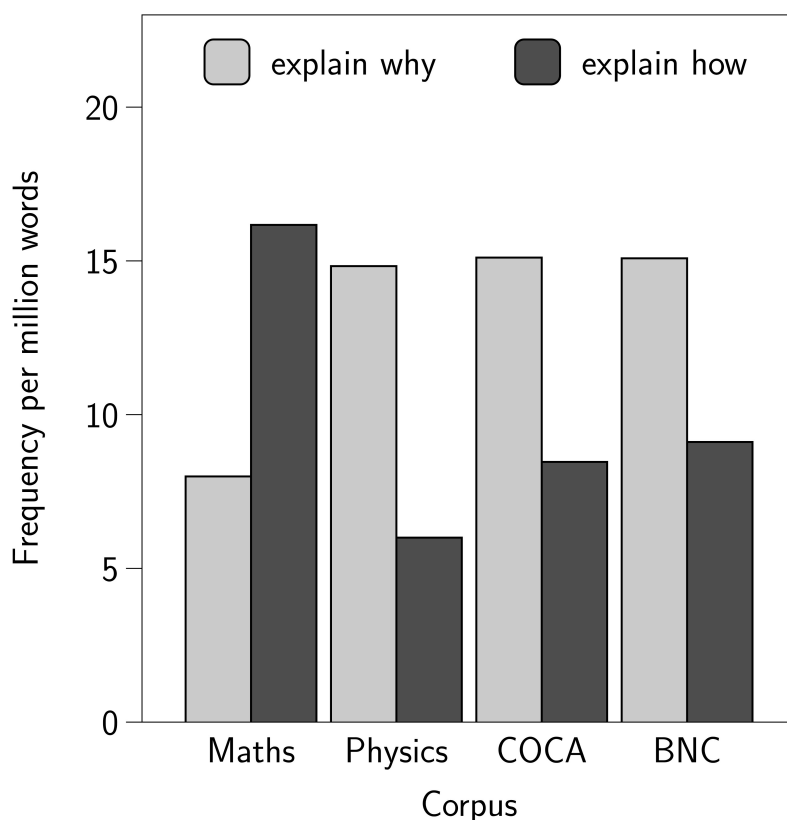


Figure 2. Frequencies of 'explain words' immediately followed by the words *why* or *how* in the mathematics, physics, COCA, and BNC corpora.

#### 4.4 Replication using the May-August 2009 papers

To replicate our analyses, we used the May-August 2009 papers from the ArXiv. Table 6 presents the frequencies of 'explain words' appearing in the May-August set of mathematics and physics papers, while Table 7 presents the frequencies of 'explain words' immediately followed by the words *why* or *how* in these sets of papers.

Table 6 reveals the same pattern of frequencies as Table 2. Indeed, the same five ‘explain words’ (explain, explained, explanation, explains, and explaining) made up 95% of all ‘explain words’ in each set of papers. Furthermore, the number of ‘explain words’ per million was very similar in the two sets of papers (158.94 and 163.63 in mathematics; 362.64 and 351.60 in physics), with around 2.15 times more ‘explain words’ per million in the physics papers than in the mathematics papers.

	Mathematics		Physics	
	frequency	per million	frequency	per million
explain	1881	60.12	7974	126.96
explained	1841	58.84	6596	105.02
explanation	537	17.16	3788	60.31
explains	525	16.78	1694	26.97
explaining	166	5.31	954	15.19
explanations	98	3.13	740	11.78
explanatory	36	1.15	78	1.24
unexplained	19	0.61	159	2.53
explication	1	0.03	5	0.08
explicated	6	0.19	12	0.19
explicate	6	0.19	12	0.19
explicating	0	0.00	5	0.08
unexplainable	0	0.00	1	0.02
explications	2	0.06	5	0.08
explainable	1	0.03	24	0.38
explicates	0	0.00	0	0.00
explicable	0	0.00	24	0.38
inexplicable	1	0.03	12	0.19
Total	5120	163.63	22083	351.60

Table 6. Frequency and frequency per million words of ‘explain words’ appearing in the May-August mathematics and physics papers.

	Mathematics		Physics	
	frequency	per million	frequency	per million
*expla* why	277	8.85	970	15.44
*expla* how	526	16.81	464	7.39
Total	803	25.66	1434	22.83

Table 7. Frequency and frequency per million words of ‘explain words’ immediately followed by the words *why* or *how* in the May-August mathematics and physics papers.

Similarly, the frequencies presented in Table 7 are consistent with those in Table 5: the numbers of \*expla\*-why and \*expla\*-how expressions per million were similar in each discipline (25.66 in mathematics; 22.83 in physics), but there were significantly different distributions of these types of expressions in the two sets of papers (Fisher’s exact test,  $p < .001$ ). Again, mathematicians used nearly twice as many \*expla\*-how expressions as \*expla\*-why expressions, and physicists used a little over twice as many \*expla\*-why expressions as \*expla\*-how expressions.



## V. Discussion

Our analysis of explanatory language in a large sample of mathematics papers allows us to offer empirically justified contributions to philosophical debates. We relate our findings now to various points raised in our opening sections.

First, our findings do not support the often-made claim in the philosophy of mathematics that explanatory talk is prevalent in mathematical writing. Indeed, mathematics research papers contain less than half the amount of ‘explain words’ in physics research papers, and less than two-thirds the amount in general English. Even within the subject, mathematicians discuss explanation less than other practices such as solving problems and proving theorems. Nevertheless, our data are also inconsistent with Zelcer’s (2013) claim that mathematicians “rarely” talk about explanation, whereas this is “the standard vocabulary” of science: we found around 160 ‘explain words’ per million in mathematics and 360 per million in physics. So, although discussion of explanation is less common in mathematics, it is far from non-existent. Philosophers who appeal to mathematical practice to justify the importance of studying mathematical explanation will find succour in our data.

Second, our data shed light on the types of explanation discussed by mathematicians. We found that when mathematicians engage in explanatory talk, they seem more often interested in explaining how to do something in mathematics than in explaining why things are the way they are. In both physics and general English we found the opposite. This is particularly interesting given the concern philosophers of mathematics devote to intra-mathematical explanations of the form  $X$  explains why  $Y$  (where  $X$  and  $Y$  are mathematical assertions), and particularly to the notion of explanatory proofs in which proof  $X$  explains why theorem  $Y$  is true (Colyvan 2011; Steiner 1978). Perhaps this concern has been inherited from more traditional study of scientific explanation, where scientists wish to answer why-questions about the real world and where, according to our findings about physics, this is reflected in their written discourse. Our findings suggest that a focus on explaining why may be misguided for those interested in explanation in the discourse of professional mathematicians. Indeed, as suggested by Rav (1999), it seems that when it comes to proofs and explanations, mathematicians communicate more in terms of learning how to solve problems than in terms of learning why mathematical results hold true.

Of course, as with any empirical work, one must be careful about several of the inferential jumps made in this kind of analysis. First, while the ArXiv may well be the world’s largest, most widely used repository of mathematical preprints and postprints, it nevertheless represents a specific type of mathematical discourse. Our work thus leaves open the possibility that studies of mathematical discourse in settings such as conversational or other digital communications could lead to contrasting findings. Perhaps, for instance, mathematicians are more willing to discuss explanations in general and answers to why questions in particular when communicating in “live” verbal settings. Second, we have analysed these research papers for a potentially limited type of explanatory talk, requiring the use of ‘explain words’ or a limited number of alternative, related expressions. While this was an obvious place to start, it is certainly possible that analysing other expressions related to mathematical explanation might alter our results. These limitations indicate clear avenues for future empirical research on mathematical explanation.

## Acknowledgements

This project was funded by the British Academy and The Leverhulme Trust. This work was presented at the 20th Annual Conference on Research on Undergraduate Mathematics Education (San Diego, 2017) and the 25th International Congress of History of Science and Technology (Rio, 2017) and we are grateful to the audiences for their valuable suggestions. We also thank the editor of the present volume and an anonymous referee.

## Suggested reading

Baker, P., C. Gabrielatos, M. Khosravinik, M. Krzyżanowski, T. McEnery and R. Wodak (2008), ‘A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press’, *Discourse & Society*, 19: 273-306.

Garside, R., G. Leech and T. McEnery (2013), *Corpus Annotation: Linguistic Information from Computer Text Corpora*, 3rd edn, Abingdon, UK: Routledge.

McEnery, A. M. and A. Wilson (2001), *Corpus Linguistics: An introduction*, Edinburgh University Press.

Overton, J. A. (2013), ‘“Explain” in scientific discourse’, *Synthese*, 190: 1383–1405.

## References

Baker, P., C. Gabrielatos, M. Khosravinik, M. Krzyżanowski, T. McEnery and R. Wodak (2008), ‘A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press’, *Discourse & Society*, 19: 273-306.

Biber, D. (1993), ‘Representativeness in corpus design’, *Literary and Linguistic Computing*, 8(4): 243-257.

Bisson, M-J., C. Gilmore, M. Inglis and I. Jones (2016), ‘Measuring conceptual understanding using comparative judgement’, *International Journal of Research in Undergraduate Mathematics Education*, 2 (2): 141-164

CasualConc (Version 2.0.3) [Computer Software], available from <https://sites.google.com/site/casualconc/>

Chi, M. T. H., M. Bassok, M. W. Lewis, P. Reimann and R. Glaser (1989), ‘Self-explanations: How students study and use examples in learning to solve problems’, *Cognitive Science*, 13: 145–182.

Colyvan, M. (2011), *An Introduction to the Philosophy of Mathematics*, Sydney: University of Sydney.

Crowdy, S. (1993), ‘Spoken corpus design’, *Literary and Linguistic Computing*, 8 (4): 259-265.

- Dawkins, P., M. Inglis and N. Wasserman (2018), 'The use(s) of 'is' in mathematics', *22nd Annual Conference on Research on Undergraduate Mathematics Education*, San Diego, CA.
- Fonseca, B. A., & Chi, M. T. (2011). Instruction based on self-explanation. In R. E. Mayer & P. A. Alexander (Eds.) *Handbook of research on learning and instruction*, (pp. 296-321). New York: Routledge.
- Garside, R., G. Leech and T. McEnery (2013), *Corpus Annotation: Linguistic Information from Computer Text Corpora*, 3rd edn, Abingdon, UK: Routledge.
- Hafner, J. and P. Mancosu (2005), 'The Varieties of Mathematical Explanation', in P. Mancosu et al. (ed), *Visualization, Explanation and Reasoning Styles in Mathematics*, 215–250, Berlin: Springer.
- Hempel, C. G. and P. Oppenheim, (1948), 'Studies in the Logic of Explanation', *Philosophy of Science*, 15 (2): 135-175.
- Herbel-Eisenmann, B., D. Wagner and V. Cortes (2010), 'Lexical bundle analysis in mathematics classroom discourse: The significance of stance', *Educational Studies in Mathematics*, 75: 23-42.
- Hodds, M, L. Alcock and M. Inglis (2014), 'Self-explanation training improves proof comprehension', *Journal for Research in Mathematics Education*, 45: 98-137.
- Hunston, S. (2002), *Corpora in Applied Linguistics*, Cambridge: Cambridge University Press.
- Johansson, S., G. N. Leech and H. Goodluck (1978), *Manual of information to accompany the Lancaster-Oslo/Bergen Corpus of British English, for use with digital computer*. Department of English, University of Oslo.
- John, L. K., G. Loewenstein and D. Prelec (2012), 'Measuring the prevalence of questionable research practices with incentives for truth telling', *Psychological Science*, 23 (5): 524-532.
- Knuth, D. E. (1979), *TEX and METAFONT: New Directions in Typesetting*, Bedford, MA: American Mathematical Society and Digital Press.
- Knuth, E., A.C. Stephens, N. M. McNeil and M. W. Alibali (2006), 'Does understanding the equal sign matter? Evidence from solving equations', *Journal for Research in Mathematics Education*, 37 (4): 297-312.
- Lange, M. (2009), 'Why proofs by mathematical induction are generally not explanatory', *Analysis*, 69 (2): 203-211.
- Leech, G. (2013), 'Introducing corpus annotation', in R. Garside, G. Leech and T. McEnery (eds), *Corpus Annotation: Linguistic Information from Computer Text Corpora*, 3rd ed., 1-18, Abingdon, UK: Routledge.
- Leinhardt, G. (2001). Instructional explanations: A commonplace for teaching and location for contrast. In V. Richardson (Ed.), *Handbook of Research on Teaching*

- (4th ed., pp. 333–357). Washington, DC: American Educational Research Association.
- Louw, B. (1993), ‘Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies’, in M. Baker, G. Francis and E. Tognini-Bonelli (eds), *Text and Technology: In Honour of John Sinclair*, 157–76, Philadelphia/Amsterdam: John Benjamins.
- McEnery, A. M. and A. Wilson (2001), *Corpus Linguistics: An Introduction*, Edinburgh University Press.
- Mancosu, P. (2001), ‘Mathematical Explanation: problems and prospects’, *Topoi*, 20: 97-117.
- Mejía-Ramos, J. P. and M. Inglis (2011), ‘Semantic contamination and mathematical proof: Can a non-proof prove?’, *Journal of Mathematical Behavior*, 30: 19-29.
- Mejía-Ramos, J. P. and K. Weber (2014). ‘Why and how mathematicians read proofs: Further evidence from a survey study’, *Educational Studies in Mathematics*, 85 (2): 161-173.
- Monaghan, J. (1991), ‘Problems with the language of limits’, *For the Learning of Mathematics*, 11 (3): 20–24.
- Overton, J. A. (2013), ‘“Explain” in scientific discourse’, *Synthese*, 190: 1383–1405.
- Rav, Y. (1999), ‘Why do we prove theorems?’, *Philosophia Mathematica*, 7 (3): 5-41.
- Resnik, M. D. and D. Kushner (1987), ‘Explanation, independence and realism in mathematics’, *The British Journal for the Philosophy of Science*, 38 (2): 141-158.
- Rittle-Johnson, B., A. Loehr and K. Durkin (2017), ‘Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles’, *ZDM Mathematics Education*, 49: 599-611.
- Salmon, W. (ed) (1971), *Statistical Explanation and Statistical Relevance*, Pittsburgh: University of Pittsburgh Press.
- Salmon, W. (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.
- Simmons, J. P., L. D. Nelson and U. Simonsohn (2011), ‘False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant’, *Psychological Science*, 22 (11): 1359-1366.
- Sinclair, J. (1991), *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.
- Steiner, M. (1978), ‘Mathematical Explanation’, *Philosophical Studies*, 34: 135–151.
- TagAnt (2014, Version 1.2.0) [Computer Software], available from <http://www.laurenceanthony.net/>

- Treagust, D. F., & Harrison, A. G. (1999). The genesis of effective scientific explanations for the classroom. In J. Loughran (Ed.) *Researching Teaching: Methodologies and Practices for Understanding Pedagogy* (pp. 28-43). London: Falmer Press.
- The SAMUELS Consortium (MNOP). The SAMUELS Project. United Kingdom AHRC and ESRC. <http://www.glasgow.ac.uk/samuels>.
- Van Heuven, W. J., P. Mandera, E. Keuleers and M. Brysbaert (2014), 'SUBTLEX-UK: A new and improved word frequency database for British English' *The Quarterly Journal of Experimental Psychology*, 67 (6): 1176-1190.
- Weber, E. and J. Frans (2016), 'Is mathematics a domain for philosophers of explanation?', *Journal for General Philosophy of Science*, 48 (1): 125-142.
- Weber, K. and J. P. Mejía-Ramos (2011), 'Why and how mathematicians read proofs: An exploratory study', *Educational Studies in Mathematics*, 76: 329-344.
- Zelcer, M. (2013), 'Against Mathematical Explanation', *Journal for General Philosophy of Science*, 44: 173-192.