The Language of Proofs: A Philosophical Corpus Linguistics Study of Instructions and Imperatives in Mathematical Texts

Fenner Stanley Tanswell & Matthew Inglis

Published version available on request or can be accessed online here:

https://link.springer.com/referenceworkentry/10.1007/978-3-030-19071-2_50-1

Abstract

A common description of a mathematical proof is as a logically structured sequence of assertions, beginning from accepted premises and proceeding by standard inference rules to a conclusion, which is the theorem to be proved. Does this description match the language of proofs as mathematicians write them in their research articles? In this chapter, we use methods from corpus linguistics to look at the prevalence of imperatives and instructions in mathematical preprints from the arXiv repository. We find thirteen verbs that are used most often to form imperatives in proofs, and that these show up significantly more often within proofs than in the surrounding mathematical writing. We also show that there are many more verbs used to form a diverse selection of instructions in proofs. These findings are at odds with the view of proofs as sequences of assertions. Instead, we argue in favour of the recipe model of proofs: that proofs are like recipes, giving instructions for mathematical actions to be carried out.

1. Introduction

One of the central principles of the philosophy of mathematical practices is to embrace, analyse and understand the real (and often messy) features of mathematics as a living subject. A project like this is necessarily interdisciplinary, relying on history, sociology, anthropology, psychology, cognitive science, education, argumentation studies, etc. to identify and study the salient features of mathematics. Here, we will use methods from *corpus linguistics* to study the language used in modern mathematics, and especially proofs.

Despite the central importance of attending to the realities of practices, our understanding of what proofs are really like is, in the existing literature, typically only informed by modern and historical case studies, and often by anecdote and personal familiarity. While we do not dispute the value of these, they can only provide part of the picture. In this chapter, we begin to supply another part through the systematic, large-scale, computational analysis of mathematical texts. Specifically, we will analyse texts from the *arXiv*, a pre-print server for mathematics research articles.

We will investigate the relationship between the language used in written, textual mathematics, and how we should understand what proofs are. In particular, we will look at the use of instructions in proofs. One way that instructions are issued in mathematics texts is in sentences in the imperative mood. A sentence in the imperative mood instructs or commands that some action be carried out, such as "Go to your room"; "Eat your greens"; and "Do as you'd be done by". The use of the imperative mood can also be found in mathematics, such as "Suppose that the inductive hypothesis holds"; "Let *I* be a proper ideal of *A*"; and "Notice that *x* is a member of *y*". While the presence of instructions and imperative mood sentences in mathematical writing has been observed and discussed to some extent, there has been no systematic investigation of their relative frequency in modern mathematical proofs, and only to a limited extent the philosophical implications they have.

The central questions we will address through corpus linguistics concern the *frequency* and *diversity* of the instructions given in proofs in modern mathematical writing. In other words, we are asking how common various imperatives are, and how broad the range of permissible imperatives is.

There are several reasons to look specifically at the use of the imperative mood in mathematics. First of all, it contributes to philosophical debates on the nature of proof and rigour, since it tells us about how written proofs are currently presented in research articles. Secondly, it is relevant to the epistemology of mathematics, since the potential ways that knowledge is gained from imperatives, which instruct us to carry out actions, might be importantly different to knowledge from other kinds of sentences. Thirdly, the existence of imperatives in mathematical writing points us to a different class of speech acts being used in proofs, and thereby raises questions about what the function of written proofs might be. Fourthly, it contributes to an overall picture of mathematical style and the language used in professional mathematics, which has impacts in mathematics education on how students are taught to read and write proofs. Finally, it adds to the practical efforts to embed computer assistants into mathematical practice, by showing what computers need to be able to read and produce in order to engage in mathematics alongside mathematicians.

The structure of this chapter is as follows. In section 2 we shall outline the recipe model of proofs, under which a proof provides a recipe of instructions for the reader to follow. In section 3 we will look at the background of existing literature discussing the use of instructions and imperatives in mathematics.

In section 4 we will present our corpus study: 4.1 describing corpus linguistics more generally and the body of texts we are working with; 4.2 offers our hypotheses to test based on the recipe model and the existing literature; 4.3 describes the methods; and 4.4 goes through the various results we obtained. Section 5 will offer an empirically informed discussion of the philosophical questions in light of the results.

2. The Recipe Model of Proofs

In this section, we will describe the *recipe model of proofs*, as defended by Tanswell (forthcoming). This model stands in contrast to the *standard model of proof*, on which proofs are seen as sequences of assertions, which are stepwise inferences from premises to a conclusion, and all of which are declarative statements. From that perspective, all of the activity is reduced to acts of inferring, and the steps in the proof represent the results of those inferences. The standard model of proof has many virtues; however, it also often mischaracterises both what proofs look like in practice and how they are engaged with by practitioners. The recipe model of proof is intended to more faithfully account for these aspects of mathematical proofs. This alternative model develops an analogy that mathematical proofs are akin to cooking recipes. This section will fill out the important aspects of this analogy. Our summary of the recipe model will focus on four key points: (1) the instructions given by the text; (2) the language used to issue them; (3) the relationship between the text and the activity it tells you to carry out; and (4) the epistemic significance of proving. In this section we will discuss these four points.[1]

On point (1), the reason to compare proofs to recipes is that this emphasises the ways in which proofs are used to issue instructions, either explicitly or implicitly, for us to carry out a series of mathematical actions. On the recipe model, we propose that the actions that a proof instructs the audience to perform are far more diverse than inferring alone, and can include instructions for constructing useful mathematical objects, for manipulating objects and equations, for hypothetically considering mathematical assumptions, for fixing notations, for directing the readers' attention, for structuring their reasoning, and for many more possible epistemic actions.

For point (2), the emphasis of the analogy to recipes is on the similarity of the language used to issue instructions. There are, of course, large variations in the presentation of both cooking recipes and

[1] As we are interested in written textual proofs here, we set aside a fifth point by Tanswell (forthcoming) concerning the use of diagrams to issue instructions.

mathematical proofs. However, in cooking recipes the predominant mood is usually the imperative mood, directly listing instructions to follow to carry out the recipe.[2] The recipe model thus predicts that proofs will similarly include imperative sentences to offer instructions to the reader. However, two disanalogies are salient. First of all, recipes might sometimes include elements describing the state of the food, but in mathematics there is reason to find more declarative content because we are likely to see some aspects of the traditional view of proofs, where they do contain declarative statements with inferences from one to the next. Secondly, as a matter of style, cooking recipes tend to use more purely imperatival language, while in maths can also make use of the "We [verb]" constructions. Nonetheless, in both disanalogies the instructions are still there: in the latter case explicitly, but in the case of inferring from statement to statement, the proof is still implicitly instructing the reader to undertake those mathematical acts of inference. This means the recipe model is not dependent on there being many imperatives, but does certainly look more plausible if there are. We expect that there are plenty of proofs without imperatives, but if there are very few proofs that contain any imperatives at all, that would be a bad sign for the recipe model generally.

The next point (3) is about how both proofs and recipes should be understood in terms of the relationship between a written text and the actions they describe. A recipe is intrinsically linked to the activity of making, and by this analogy we should understand proofs to be the same. A proof is a device for recording and communicating how to carry out some piece of mathematics, but it is the correct performance of the relevant actions that constitutes mathematical practice. On this model, therefore, we prioritise the doing of mathematics over the written artefact of a proof itself. Putting it another way: the recipe model makes proofs secondary to proving. Here, the activity of proving is construed broadly, such that the author(s) of the proof have proved the theorem, then use a proof to communicate to their readers how to carry out the mathematical activity and thereby establish the truth of the theorem. Similarly, we expect the author of a recipe to have cooked it themselves, then use the recipe to enable the audience to reproduce it.

The final point (4) concerns epistemology: one of the central functions that proofs play in mathematics is epistemic: a correct and rigorous proof is meant to establish a theorem to be known with certainty.[3] But an important question for the standard model is: how does a presented proof link to

[2] Detailed considerations of the language used in cooking recipes in different contexts can be found in Gerhart et al.'s (2013) edited volume.

[3] This is again a rather strong traditional conception of the epistemic function of proofs, and we agree with De Toffoli (2020) that this needs to be tempered to properly reflect practice.

knowledge at all? What is connecting the proof on the page (or wherever) with the knowledge in your head? A plausible story would already have to be recipe-like, where the proof tells you exactly which inferences to carry out. But as soon as the proof has any inferential gaps (Fallis 2003; Hamami 2014; Andersen 2020), which many proofs do, an additional account is needed. The recipe model's answer to this question is more direct: that knowledge is gained through the mathematical activity that we are instructed in by proofs, with this activity acting as a key intermediary to explain the epistemic functioning of proofs. Furthermore, the recipe model also expands the traditional conception of mathematical knowledge, which is largely about propositional knowledge, to also include mathematical knowledge-how. In order to cook from a recipe, one needs to *know how* to perform the relevant actions; likewise, to carry out a proof one needs to *know how* to perform the mathematical actions it tells you how to do. In fact, the mixture of imperative and declarative forms suggests that the knowledge gained from proving will be a combination of knowledge-how and knowledge-that. (For a longer development of this epistemic picture see Tanswell (2016, ch. 4) and Tanswell (forthcoming, Section 6).)

The recipe model of proofs gives us concrete predictions on our two main areas of investigation, concerning the frequency and diversity of imperatives, namely that they should be common and diverse. They should be common because the issuing of instructions is one of the primary functions of proofs. They should be diverse because the possible space of mathematical activities that one could be instructed to carry out is large. These hypotheses are set out in section 4.2 and tested in section 4.4.

3. Background on Mathematical Language

In this section we will survey previous discussions of the use of instructions and imperatives in mathematics. We will group these into two broad categories, based on the sources of the discussion: 1) mathematics education and philosophy; and 2) formalised mathematics.

3.1. Mathematics Education and Philosophy

Researchers in Mathematics Education and Philosophy have both, at times, noted the use of the imperative mood in mathematical proof texts. However, in many cases this has been just parenthetically or tangentially. We will set out a number of examples from these literatures chronologically.

David Pimm, in *Speaking Mathematically* (1987), analysed the kinds and effects of the language used in the mathematics classroom. As an aside, he remarked on mathematical imperatives:

> "There is also the equally common use of the imperative in mathematical discourse (e.g. *consider*, *suppose*, or *define*) [...] Another form is `Let x be the number of …' Who is giving permission for this to be done? The mathematical imperative […] is a topic worthy of considerable attention." (Pimm 1987, p. 72).

However, Pimm does not go on to give them any such attention, merely suggesting that the impersonal style allows mathematics texts to supress and omit the agent who is carrying out the actions. This is a clear case of a parenthetical mention of the imperative mood in mathematics.

The work of Brian Rotman (1988, 1993) included the unusually careful examination of imperatives in mathematics. His broader project is one of semiotic analysis of mathematical signs, which is the analysis of how signs come to have meaning in mathematical texts. He considers imperatives to be a major feature of mathematical writing, and in establishing the shared meanings used in mathematics:

> "But proof in turn involves the idea of an argument, a narrative structure of sentences, and sentences can be in the imperative rather than the indicative. […] Mathematics is so permeated by instructions for actions to be carried out, orders, commands, injunctions to be obeyed — 'prove theorem T', 'subtract from y', 'drop a perpendicular from point P onto line L', 'count the elements of set S', 'reverse the arrows in diagram D', 'consider an arbitrary polygon with k sides', and similarly for the activities specified by the verbs 'add', 'multiply', 'exhibit', 'find', 'enumerate', 'show', 'compute', 'demonstrate', 'define', 'eliminate', 'list', 'draw', 'complete', 'connect', 'assign', 'evaluate', 'integrate', 'specify', 'differentiate', 'adjoin', 'delete', 'iterate', `order', 'complete', 'calculate', 'construct', etc. that mathematical texts seem at times to be little more than sequences of instructions written in an entirely operational, exhortatory language." (Rotman 1988, p. 8)

He argued that imperatives in mathematics can be separated into two distinct kinds: *inclusive* imperatives, which are for creating shared mathematical discourse (by introducing shared standards, notations, referents etc.), and *exclusive* imperatives, which take a shared mathematical discourse for granted and operate within it. Rotman takes “consider”, “define”, “prove” and synonyms of these to be inclusive, constructing a shared domain of meaning, while other mathematical verbs merely tell us to operate on the shared domain of meanings that is already established. For example, “consider the metric space *(X,d)”* tells us to use *X* as a set with a distance measure *d* obeying the definitions of a metric space, constructing shared meanings for the author and reader for the coming mathematical text, while “add the distances d_1 and d_2” merely instructs the reader to carry out an operation within the space, where meaning for the variables has already been established. This distinction is important for understanding the semiotics of a mathematics text and how some imperatives work to create shared mathematical discourses, but we don’t think the distinction coincides with the epistemic and pragmatic functions of different kinds of imperatives. We also note that while Rotman listed many potential mathematical instructions, there is no study as to how common these actually are. We will carry out such a study below.

Besides the creation of shared meaning in mathematical texts, another feature of the use imperatives in mathematical texts is to mark who is part of the mathematical community, something discussed by Candia Morgan (1996). She says:

> “One common characteristic of academic mathematics texts (and some school texts) is the conventional use of imperatives such as *consider, suppose, define, let* x *be*. Like the use of we, these implicate the reader, who is addressed implicitly by the imperative form, in the responsibility for the construction of the mathematical argument. The use of imperatives and of other conventional and specialist vocabulary and constructions characteristic of academic mathematics marks an author's claim to be a member of the mathematical community which uses such specialist language and hence enables her to speak with an authoritative voice about mathematical subject matter. At the same time it constructs a reader who is also a member of the same community and is thus in some sense a colleague (although the nature of this relationship may vary according to the type of action demanded)” (Morgan 1996, p. 6)

Her analysis was primarily focused on how these issues of community membership work for students of mathematics who are not yet members of the community but are in the process of being inducted and enculturated into it. However, she did not further concern herself with imperatives in particular.

Paul Ernest has also discussed imperatives in mathematics in multiple places. For example, in his book *Social Constructivism as a Philosophy of Mathematics* (1998) he stated:

> “Thus mathematical texts comprise specific assertions and imperatives directed by the writer to the reader.” (Ernest 1998, p. 170)

We are broadly sympathetic to this view of proofs as, at least in part, functioning as communicative devices to transfer knowledge of mathematical reasoning from author to reader. In more recent work, citing Rotman, Ernest went so far as to say:

> ““Imperatives are orders that instruct or direct actions either inclusively, such as: let us …, consider …, or exclusively, such as: add, count, solve, prove, etc. Imperatives occur more frequently in mathematics than in any other academic school subject.” (Ernest 2018, p. 191)

We note that this is a direct empirical claim, offered without evidence, and one that could potentially be refuted by other disciplines which use imperatives in their written work such as chemistry, philosophy, or geology (cf. Swales et. al. 1998). Below we will carry out one study of the frequency of imperatives in proofs in mathematical research articles.

The work of Rotman was also extended and challenged by Roy Wagner (2009; 2010) in his detailed semiotic study of Gödel’s proofs of the first incompleteness theorem. Wagner discussed the imperatives both explicit and implicit in the text, and incorporated these into a nuanced picture of the mathematical subject, agency, and the relationship these have to a mathematical text. One major difference from our findings below is that Wagner finds very few explicit imperatives:

> “First, there are very few imperatives in these texts. Perhaps the imperative mood was considered ill-suited for civilised written communication in the Vienna and Princeton circles. Perhaps it never occurred to writers in these circles to dominate a mathematical text with the imperative mood. Instead of imperatives, we have the frequent use of the indicative mood in both active and passive voices, often attributed to the character we.” (Wagner 2009, p. 48)

It is notable, but maybe not surprising, that the modern writing style is different in some ways to Gödel’s in the 1930s. Wagner’s work also provides a nice contrast for what can be achieved with different methodological approaches: Wagner uses detailed analysis of a single mathematician’s writing, while the corpus linguistics approach below looks for patterns across thousands of papers by thousands of authors.

A notable difference related to action-based vocabulary was also found by Mejía-Ramos & Inglis (2011). On a proof-evaluation task with undergraduate students, they found that changing the question from “Does the argument prove the claim?” to “Is the argument a proof of the claim?” had a significant impact on how the students evaluated a specific visual argument. The students were more willing to agree that the argument proved the claim than that it was a proof for the claim. The researchers built the case that the verb form was more connected to judgments of conviction, while the noun form was more connected to judgments of validity. They related this to issues of semantic contamination: where the everyday meanings of words affect our understanding of them in more formal and technical settings like mathematics (cf. Tanswell 2018, section 4).

As a final example, Lew and Mejía-Ramos (2019) investigated the norms of mathematical writing by testing the reactions of mathematicians and students to cases where those norms were flouted. One of their stimuli was an ill-formed imperative: "Suppose $(S \cdot R)^{-1}$ such that [...]". They found that professional mathematicians especially responded very negatively to the ungrammatical phrasing of this. For the mathematicians, imperatives must be used correctly and grammatically, and they appear to have strong conventions on how that is achieved. In contrast, students often did identify that the construction didn't sound right, but could not clearly express why (Lew & Mejía-Ramos 2019, p. 134).

From a different direction, Paul Halmos (1970) provided a classic style guide for mathematics, and discussed both imperatives in maths and the use of the first-person plural "we". Regarding imperatives he said: "A frequently effective and time-saving device is the use of the imperative. `To find P, multiply q by r.' `Given p, put q equal to r.'" (Halmos 1970, p. 141).[4] He also proceeded to explain using the first-person plural to convey instructions on how a proof goes:

> "There is nothing wrong with the editorial "we", but if you like it, do not misuse it. Let "we" mean "the author and the reader" (or "the lecturer and the audience"). Thus, it is fine to say "Using Lemma 2 we can generalize Theorem 1", or "Lemma 3 gives us a technique for proving Theorem 4"." (Halmos 1970, p. 141)

Indeed, it is discussions of the first-person plural "we" that are found most often in mathematical style guides. For example, Krantz likewise stated that "The custom in modern mathematics is to use the first person plural, or "we". It stresses the participatory nature of the enterprise, and encourages the reader to push on." (Krantz 1997, p. 33). A similar point was made by Knuth et al. (1989, p. 2).

---

[4] Somewhat amusingly, the immediately prior sentence is: ""Most (all?) mathematical writing is (should be?) factual; simple declarative sentences are the best for communicating facts." (Halmos 1970, p. 141).

3.2. Formalised Mathematics

There is a large area of research dedicated to formalised mathematics, where the central goal is to provide formalised counterparts of informal proofs that can be formally verified in one of the proof-checkers, such as Lean, Coq, Mizar, or Isabelle. Standardly, these come with their own language, software, and libraries of completed formal proofs that can be called on in later proofs. There are two obvious goals to this formalisation project: firstly, to check the existing mathematical literature, and secondly, to develop software that can assist mathematicians in new work.

Especially for the second goal, it is important that the formal language used by the computer is accessible to human mathematicians. For this reason, there is some research on the language of everyday mathematics focusing on how to formalise it. Oddly enough, researchers here sometimes seem to downplay the significance of imperatives in proofs. We will look at three examples of work in this area, and what they say with regards to imperatives in mathematical writing.

The oldest we consider is the work of de Bruijn (1994). His goal was to set out the "mathematical vernacular", which is meant to be the native language of mathematics, described as "[...] the very precise mixture of words and formulas used by mathematicians in their better moments" (de Bruijn 1994, p. 865). This is immediately contrasted to the "official", purely formal language of mathematics. De Bruijn then claimed that this mathematical vernacular is in essence the same as the rigorous part of mathematics:

> "Roughly speaking, the [mathematical vernacular] part of a piece of mathematics will be the rigorous part." (ibid., p. 867).

This is then used to exclude various common parts of mathematical texts from the mathematical vernacular as he sees it, such as historical remarks, textual signposting, references to the syntactical form of the presented equations, and indications of how to reproduce omitted material. Included in this list is "Commands, like `show that'." (ibid. p. 867). Hereby de Bruijn excludes imperatives from the domain of rigorous mathematics.

This exclusion is also present in the book *The Language of Mathematics* by Mohan Ganesalingam (2010). The purpose of the book was to analyze the language used in mathematical texts in keeping with a broadly generative tradition, to give a formal, objective, and precise syntactic structure and associated semantics. Ganesalingam said that he will restrict himself to the language used in "more rigorous, careful textbooks" (ibid., p. 7) and only the part in the *formal mode*, where "[a] sentence is in the formal mode if and only if it consists only of assertions about mathematical objects and about mathematical facts, so that its semantic content is purely truth-conditional and can be completely captured in an appropriate logic" (ibid. p. 7). This clearly precludes imperatives, since they are not assertions, nor do they have a purely truth-conditional semantics.[5] Ganesalingam thus stated that:

> "[T]he language of mathematics consists purely of assertions about mathematical objects. As a result, textual mathematics predominantly uses the third person singular and third person plural, to denote individual mathematical objects (or propositions) and collections of mathematical objects (or propositions) respectively. The first person plural ('we') is also used in a more restricted capacity, with a limited and potentially closed class of verbs, typically to refer to the mutual intent of the author and reader." (ibid. p. 21).

This seems to allow constructions similar to imperatives (e.g., "We suppose", "We assume", "We obtain", etc.") but not the imperative forms ("Suppose", "Assume", "Obtain"). Of course, if we restrict ourselves to the formal mode, which by definition is exclusively about mathematical assertions, then the absence of imperatives must follow. Nonetheless, this doesn't justify these bolder claims about *the language of mathematics* being limited to this class of sentences, nor even claims about rigorous textbooks.[6]

Below we will demonstrate that imperatives are found in mathematical writing as found on the *arXiv*. Whether any given article on the *arXiv* is part of rigorous mathematics or not must surely be debatable, but there is no doubt that such a corpus must represent a collection of paradigm examples of the language of mathematics.

---

[5] The formal mode is contrasted with the informal mode: "The informal mode consists of commentary on the ongoing mathematics. Individual sentences in the language of mathematics tend to be either entirely formal or entirely informal, though there are exceptions. It is worth noting that mathematicians have very strong intuitions about the distinction between these two modes." (p. 7). Below we will be showing that imperatives are a part of normal mathematical language, but it bears emphasising that no evidence is provided for the existence or reliability of these intuitions, and even if mathematicians do have robust intuitions about the difference between doing the mathematics and commentary on it, that does not necessarily align with the characterisation he gives of the difference between the formal and informal modes.

[6] A similar criticism can be levelled concerning another exclusion: that of modality in mathematics. Ganesalingam said: "[...] textual mathematics is much more restricted than natural language. [...] Mathematical truths are true in all possible worlds, and therefore there is no modality." (pp. 22-23). However, Hodges (2013) demonstrated that there are many uses of modals in mathematical writing but that these don't refer to metaphysical necessity and possibility.

In contrast to de Bruijn and Ganeslingam, formalised mathematics does not need to exclude imperatives. Here is an example of a simple proof in Lean[7]:

```
theorem even_add : ∀ m n, even m → even n → even (n + m) :=
take m n,
assume ⟨k, (hk : m = 2 * k)⟩,
assume ⟨l, (hl : n = 2 * l)⟩,
have n + m = 2 * (k + l),
by simp [hk, hl, mul_add],
show even (n + m),
from ⟨_, this⟩.
```

The details of how the system works are not important here; all we need is to see the imperatives at work in the proof. What is remarkable is that the imperatives serve the dual functions of both acting as imperatives in a proof in the sense outlined above for the human mathematician, but also acting as imperatives as commands for the computer (in the imperative programming tradition).

The system Naproche-SAD (De Lon et al. 2020) similarly incorporates many imperatives into its "controlled natural language", a fact notable because of the system's explicit aim to emulate normal mathematical language. Here is an example of Naproche-SAD text taken from (Frerix et al. 2018, p. 3):

> ***Theorem 1***. *Assume f is holomorphic and the domain of f is a region. If f has a local maximal point then f is constant.*
>
> ***Proof*** *Let z be a local maximal point of f. Take ε such that Bε(z) is a subset of Dom(f) and |f[w]| ≤ |f[z]| for every element w of Bε(z). Let us show that f is constant on Bε(z). Assume the contrary. Then f[Bε(z)] is open. We can take δ such that* B_δ*(f[z]) is a subset of f[Bε(z)]. Therefore there exists an element w of Bε(z) such that |f[z]| < |f[w]|. Contradiction. end. Hence f is constant.*

Again, observe the natural use of imperatives to construct the proof here. The creators of *Naproche* are amongst a relatively small set of researchers who also use linguistics to study mathematical language, and we hope that our empirical work below will support its further development.

---

[7] This example is taken from https://leanprover.github.io/introduction_to_lean/ written by Jeremy Avigad, Gabriel Ebner and Sebastian Ullrich. (Accessed 11/02/2021).

4. A Corpus Linguistics Study

4.1. Corpus Linguistics

The driving idea underlying the philosophy of mathematical practice is to ensure that our philosophical claims about mathematics are true and accurate of real mathematics, not just some idealised version of it. This means that we need to give empirical grounding to any factual claims about mathematical practice that we rely on (e.g. Aberdein & Inglis 2019). In this chapter we are focused on the language of mathematics, and especially on the language used in proofs. In our review of the literature, however, the main evidence used by previous authors is their own experience as mathematicians (e.g. Ganesalingam 2010) or the detailed analysis of isolated texts (e.g. Wagner 2009, 2010). Here, we take a very different approach: we will examine a large number of mathematics texts using the tools of corpus linguistics.

A corpus in this sense is a substantial body of texts that are machine-readable to allow for computational analysis. In our case, we are using the corpus assembled by Alcock et al. (2017) and used by Mejía-Ramos et al. (2019) in their study of language relating to explanation in mathematics. The corpus is made up of all of the papers uploaded to the arXiv in the first four months of 2009 to the *Mathematics* category. The arXiv is an online repository for texts from mathematics, physics, computer science, and various related disciplines, which acts as the main pre-print server for large sections of research mathematics, especially in English-speaking mathematics communities. Many mathematicians upload copies of their academic papers as pre-prints to make them freely accessible online. There is a level of selection before papers can appear on the arXiv, partially moderated by volunteered and partially by an automated system, and the assigned category for the paper does have social implications, as described by Reyes-Galindo (2016).

Texts on the arXiv are predominantly written in the TeX/LaTeX mark-up language, then compiled into pdf format. In order to process these into a form useable for computational linguistics (plain texts without extraneous code), the texts were processed to strip out the TeX/LaTeX code while preserving the structure of the written language, as described by Mejía-Ramos et al. (2019, p. 249-250).[8] This posed the challenge of what to do with inline mathematics, such as “Let f: X →Y be a bijection.” The problem is that

[8] This process was automated, and Chris Sangwin has made the code available online at https://github.com/sangwinc/arXiv-text-extracter (accessed 23/02/2021).

the same mathematics can be coded in many different ways in LaTeX, which speaks against leaving the code as is because the linguistics software would not be able to identify the different codes as representing the same syntactic unit. However, simply removing it would interfere with some of the linguistics tools (measuring word proximities, for example). As such, they opted to replace all inline mathematics with the tag "inline_math". For the above example, this gives us ""Let inline_math be a bijection." The result of this processing is sufficient for our needs too because we are interested in the natural language components and not the mathematical content.[9]

In total, the corpus then contained 6,988 files and 30,892,695 words of mathematics (with 231,400 distinct words). As a useful subset of this for our purposes, we also have a corpus that is made up of only the proofs found in these papers, containing 3,268 files and 4,973,892 words of mathematics (with 48,025 distinct words).[10] This will allow us to focus specifically on the mathematical writing found in written proofs in these papers. Simple subtraction then also lets us compare the proof-only corpus to the non-proof parts of the mathematical texts. The non-proof part therefore contains 25,918,803 words. Henceforth, we refer to the main corpus as the *All-math* corpus, and the proof only part as the *Proof-only* corpus, and the result of subtracting the *Proof-only* from the *All-math* will be called the *Non-proof* corpus.

Something important to be sensitive to in analysing these texts is how representative they are of textual mathematical proofs more generally. Since the arXiv is the standard repository for pre-print papers for many mathematical researchers, we do take it that the corpus is fairly representative. However, this must come with some caveats. The first is the most obvious: we are conducting this research in English, and so our results apply to mathematical texts written in English. They may still be applicable to nearby languages, where we believe many similar linguistic conventions apply for mathematical writing, but the current study would not reveal that. Secondly, language and other factors will in turn affect who submits their papers to the arXiv, and there are obvious demographic and geographic biases likely to arise in whose

[9] The decision of how to treat the mathematical content of the papers does affect the kinds of results we get. If we were to instead convert mathematical content to verbalised counterparts (i.e., x^2+5 becoming "x squared plus five") this would lead to different frequencies, and "dilute" the natural language parts in the corpora. However, there is certainly no unique or uncontroversial algorithm for this process of verbalisation, so our approach here is the pragmatically obvious one.

[10] Selected automatically as everything between "\begin{proof}" and "\end{proof}". This is the most common way to write proofs, but may have left out a small minority of written proofs coded differently.

work we are seeing.[11,12] Thirdly, since these are pre-print articles primarily from professional mathematicians, the mathematical language we are analysing is theirs and not, say, that of mathematics students or teacher, nor the language used in proofs in maths textbooks or other media (such as spoken proofs, or proofs on blackboards in seminars). Finally, the papers are all from 2009, so represent a snapshot of proof writing from that year.[13] We do not believe there has been any substantial change since then, but writing and presentational styles are always changing to some degree. Obviously, the results of our study will not immediately generalise to historical mathematical writing either.

4.2. Hypotheses

Based on the previous sections, we can form various hypotheses about the prevalence of instructions and imperatives in our corpus.

In the literature from mathematics education and philosophy, we saw that the researchers like Pimm (1987), Rotman (1988, 1993), Morgan (1996) and Ernest (1998, 2018) broadly claimed:

(H1): Imperatives are common in mathematical proof texts.

The recipe model of proof also predicts that (H1) will hold.

In the works of De Bruijn (1994) and Ganesalingam (2010), however, the opposite claim was made regarding imperatives. They saw rigorous proofs as only containing assertions about mathematical objects. So, the opposite hypothesis can be considered:

(H2): There are no imperatives in mathematical proof texts.

The recipe model of proof will predict that (H2) will not hold.

[11] The arXiv themselves do have various statistics about submissions and downloads at https://beta.arxiv.org/help/stats (accessed 23/02/21) but these do not include demographic or geographic data.
[12] For more on how social features of mathematics can involve epistemic exclusion, see Rittberg et al. (2018) and Tanswell & Rittberg (2020). Indeed, Reyes-Galindo (2016) examines how the automated filtering of the arXiv is partially based on linguistic analysis, and can lead to the exclusion of certain researchers.
[13] They are also from only the first four months of 2009. We do not believe that the time of year makes a significant difference to writing styles, though it is possible it affects the demographics of who is submitting their papers to some small degree.

We also saw the mathematical style guides discussing the first person plural in mathematics. This can be used with the same verbs to form "We [verb]..." sentences, which can potentially also be used to give instructions for the reader to follow. We will test how common this is as the following hypothesis:

(H3): The construction "We [verb]..." is common in mathematical proof texts.

The recipe model of proof again predicts that (H3) will hold.

Ganesalingam did note the use of the "We [verb]..." constructions (like "We suppose...", "We assume..." etc.) but states that they are only used for a "limited and closed class of verbs". Generalising to cover imperatives too, and putting this as a hypothesis:

(H4): The verbs used in imperatives and the "We [verb]..." constructions in mathematical proof texts are drawn only from a limited and closed class of verbs.

On the recipe model of proofs, proofs contain many instructions (be they explicit or implicit), and therefore the more diverse the selection of verbs the better for the model. As such, the recipe model of proofs predicts that (H4) does not hold.

In sum, (H1) and (H2) are about the frequency of imperatives in mathematical texts, while (H3) is about "We [verb]..." constructions. The hypothesis (H4) is about the diversity of verbs used to form mathematical imperatives. We believe that both of these can be used in mathematical texts to give instructions so will investigate both below.

4.3. Methods

We used Laurence Anthony's *AntConc* software (Anthony 2020) and Yasu Imao's *CasualConc* software (Imao 2020) to analyse the corpora.[14]

We had two primary targets for investigation from our hypotheses: explicit imperatives and "We [verb]" constructions. Both of these have textual markers that we could use. For explicit imperatives, we could perform a case-sensitive search for capitalised verbs, as this is a reasonable and conservative operationalisation that will capture many imperatives (e.g. "Let...", "Assume...", "Suppose...",

[14] The *AntConc* software is available online at https://www.laurenceanthony.net/software/antconc/ (accessed 24/02/21) and *CasualConc* is available at https://sites.google.com/site/casualconc/ (accessed 06/03/21).

“Construct...” all are formed with a capitalised verb to start the imperative sentence.) A drawback of this is that it will miss any imperatives constructed in different ways, such as “Firstly, suppose...”, but we did not have a systematic way to predict the idiosyncratic ways sentences may be constructed, and we believe that missing these did not alter our findings. For “We [verb]” constructions, the software allowed us to search for these strings directly.

For both explicit imperatives and “We [verb]” constructions, we needed to identify the specific verbs we will investigate in the first place. We required some way to select these systematically, but the verbs most likely to be used in mathematical imperatives are typically mathematical keywords; we shouldn’t expect, for instance, that common mathematical imperatives would use the same verbs as imperatives in general English, so couldn’t use that as a guide. To solve this, we produced a list of the 500 most common words in the *All-math* corpus and selected all of them that can ever be verbs in English. This gave a total of 91 possible verbs, listed in Table 1. One obstacle we faced was that many of the words have multiple meanings: for example, the word “set” is ambiguous between the verb form and the noun

Table 1

The 91 potential verbs among the 500 most common words in the All-Math Corpus.

action	construct	free	lie	order	satisfy	step
apply	curve	function	like	pair	say	structure
assume	define	further	limit	part	second	study
bound	denote	get	map	point	see	sum
bundle	do	give	measure	present	sense	suppose
call	estimate	graph	model	prime	sequence	take
check	exist	group	need	process	set	term
choose	factor	hand	norm	prove	show	time
claim	field	have	note	rank	side	type
complete	figure	hold	number	recall	smooth	use
conclude	find	image	observe	remark	space	will
condition	fix	introduce	obtain	respect	stable	work
consider	form	let	open	ring	state	write

form, and is used in both senses in the corpora. The same held for many of the terms for mathematical objects, such as “group”, “ring”, “field”, and “structure”. A second obstacle is that some of these words were extremely unlikely to be used as verbs in the current context, such as “stable”, “like”, and “hand”. Finally, some of the verbs have existing functions in mathematics papers, such as “claim”, “step”, and “remark”.

To separate out the verbs used as imperatives in the corpus, we generated a file for each of the 91 verbs containing all capitalised instances of that verb in the *All-math* corpus and their contexts (which shows six words either side of the verb). For example, an entry from the file for the verb "choose" is:

[circle of a copy of inline_math. Choose a base point inline_math on inline_math,]

By examining these files, we separated the 91 verbs into three classes: 1) Those where the capitalised form was used exclusively (or almost exclusively) as an imperative; 2) those where there was a mixture of imperatives and other uses; 3) those where there were no imperatives. Separating the verbs into the three lists using the context files gave us the three categories reported in Table 2.

Table 2

The 91 verbs divided into three categories.

Imperatives		Mixed		Non-Imperatives			
apply	note	claim	number	have	hand	respect	structure
assume	observe	complete	order	action	hold	ring	term
call	obtain	do	pair	bound	image	satisfy	time
check	prove	estimate	present	bundle	lie	second	type
choose	recall	factor	process	condition	like	see	will
conclude	say	find	rank	curve	limit	sense	
consider	show	form	set	exist	need	sequence	
construct	suppose	get	study	field	norm	side	
define	take	give	sum	figure	open	smooth	
denote	use	group	work	free	part	space	
fix	write	map		function	point	stable	
introduce		measure		further	prime	state	
let		model		graph	remark	step	

From here we performed a selection of analyses using the corpus linguistics software, reported below.

4.4. Results

4.4.1 Comparing Imperatives in the *Proof-Only* and *Non-Proof* Corpora

To test (H1) and (H2) on explicit imperatives, we compared the frequency (per million words) of the capitalised verbs in our *Proof-Only* and *Non-Proof* corpora. We used only the verbs from the first category of Table 2 so that our operationalisation of using capitalised verbs to track sentences in the imperative mood will be accurate. Recall that the third category contained no imperatives, and the second category

Table 3

Comparing imperative frequencies per million words in the Proof-Only and Non-Proof Corpora.

Verb	Proof-Only	Non-Proof
Let	4523	4035
Suppose	944	512
Note	929	681
Consider	570	314
Assume	556	339
Recall	304	265
Define	272	167
Fix	255	106
Denote	218	145
Observe	213	92
Choose	199	45
Take	178	49
Write	117	39
Apply	53	6
Use	28	9
Call	14	11
Introduce	11	9
Construct	8	4
Say	7	7
Show	3	10
Check	2	2
Prove	1	4
Obtain	1	0
Conclude	0	0
TOTAL	9406	6854

contained a mixture of imperatives and non-imperatives, which would interfere with the goal of the study. The results are displayed in Table 3. What we see is that the total frequency per million words for our selected verbs is 9406 in the *Proof-Only* corpus and 6854 in the *Non-Proof* corpus, a highly significant difference, Fisher's exact test, $p < .001$.

These results demonstrate that imperatives using common mathematical verbs appear more commonly within proofs than outside of proofs. If mathematical texts outside of proofs, as in the *Non-Proof* corpus, are a reasonable baseline for imperatives in mathematical writing, then these results provide evidence for (H1) and against (H2).

The results here also show major variation in the prevalence of imperatives using different verbs. By far the most common is the imperative `Let...', followed by `Suppose...' and `Note...', then by `Consider...' and `Assume'. The imperatives using `Recall...', `Define...', `Fix...', `Denote...', `Observe...', `Choose...', `Take...', and `Write...' were also relatively common. Other verbs we searched for were all present to some degree, but were less common overall. The only three verbs that showed up as a higher frequency per million words in the *Non-Proof* corpus were `Show...', `Prove...', and `Conclude'.

4.4.2 Comparing "We [verb]..." constructions in the *Proof-Only* and *Non-Proof* Corpora

To test (H3) concerning the "We [verb]..." constructions, we took all 91 of the (possible) verbs from the top 500 most common words in the *All-math* corpus (i.e., all of the verbs found in Table 1), and looked for the "We [verb]" string in the *Proof-Only* and *Non-Proof* corpora. We made this case-sensitive to only look at instances where the "We" is capitalised, thus indicating the start of a sentence. The results are reported in Table 4.[15]

The results here are more mixed than in the previous section. The two biggest entries are from `We have' and `We will', used to form the past and future tenses, so they are not in and of themselves giving implicit instructions, but could be used that way. Especially `We have' has a diverse range of potential uses: it could describe carrying out an action in the past tense, as in "We have applied...", or can be used to summarise as in "We have seen...". It can be used in the present tense to emphasise tools, facts, or mathematical objects at our disposal at that point in the proof, as in "We have that...". It can also be used to point forwards to set a shared target, as in "We have to [achieve some goal/carry out some procedure/etc.]". These constructions can also be used in the abstracts of the papers, describing what the authors do in the paper, in which case the "we" is not being used to include the reader, unlike within a proof. Examining the corpus revealed examples of all of these constructions.

While the totals overall again find more "We [verb]" constructions in the *Proof-Only* corpus than the *Non-Proof* corpus (and this difference is again significant, Fisher's exact test, $p < .001$), most of the difference can be accounted for just by the "We have" construction, which we just noted can be used in a variety of ways, so doesn't tell us much about instructions being issued.

Looking more at the specific verbs, the more common imperatives we found above do not appear to be anywhere as common in the "We [verb]" form: none of the verbs used in the thirteen most common imperatives appear in this form more than 100 times per million words in the *Proof-Only* corpus. Nonetheless, several other verbs are more common in this form, especially "We claim", "We conclude", "We prove", "We show", and "We use". What is notable about these constructions is that they appear to

[15] We have omitted all the verbs that returned no results. As such, the zero entries that do appear in the table are the result of rounding.

be more structural than the verbs we found as imperatives, in that they are used to structure where the

Table 4

Comparing "We [verb]" construction frequencies per million words in the Proof-Only and Non-Proof Corpora.

Verb	Proof-Only	Non-Proof	Verb	Proof-Only	Non-Proof
We apply	51	18	We model	0	0
We assume	59	73	We need	92	39
We bound	3	1	We norm	0	0
We bundle	0	0	We note	61	75
We call	12	85	We number	0	0
We check	11	3	We observe	30	20
We choose	55	27	We obtain	65	35
We claim	231	34	We open	0	0
We complete	5	2	We order	1	1
We conclude	137	41	We point	2	11
We condition	0	0	We present	4	17
We consider	95	108	We process	0	0
We construct	21	16	We prove	128	69
We define	96	140	We recall	32	70
We denote	77	158	We remark	14	35
We do	18	35	We satisfy	0	0
We estimate	13	3	We say	9	147
We factor	1	0	We see	40	22
We find	12	16	We set	71	40
We fix	37	24	We show	105	74
We form	1	1	We smooth	0	0
We further	8	10	We state	1	9
We get	53	21	We structure	0	0
We give	20	42	We study	3	23
We group	1	0	We sum	1	1
We have	666	348	We suppose	9	15
We hold	0	0	We take	37	21
We introduce	14	33	We term	0	0
We let	31	27	We use	121	83
We like	0	0	We will	318	387
We limit	0	0	We work	3	7
We map	0	0	We write	47	52
We measure	0	0			
			TOTAL	2923	2522

proof is aiming to get to, or (in the case of “We use”) what method or fact will be employed to get there.

With regards to (H3), we did find plenty of “We verb” constructions in the *Proof-Only* corpus, suggesting that they are not uncommon, but we also didn’t find them to be overall more common than in the more general mathematical texts, as represented by the *Non-Proof* corpus.

4.4.3 Imperative Frequency in *Proof-Only* Files

In Section 4.4.1 we tested (H1) and (H2) by looking at the frequency of imperatives in proofs relative to their frequency in mathematical texts outside of proofs. To provide another angle on how common

Table 5

Number of and percentage of files in the Proof-Only corpus containing the capitalised verb, alongside other keywords appearing at a roughly similar frequency for reference.

Verb	Number of files	% of files	Nearby word	Number of files	% of files
Let	2692	82.4%	then	2706	82.8%
Note	1486	45.5%	function	1477	45.2%
Suppose	1250	38.3%	thus	1274	39.0%
Consider	1186	36.3%	So	1186	36.3%
Assume	1085	33.2%	bounded	1053	32.2%
Recall	844	25.8%	know	843	25.8%
Define	712	21.8%	simple	722	22.1%
Fix	606	18.5%	action	598	18.3%
Choose	541	16.6%	length	553	16.9%
Denote	538	16.5%	precisely	544	16.7%
Take	459	14.1%	always	465	14.2%
Observe	447	13.7%	less	455	13.9%
Write	343	10.5%	derivative	348	10.7%
Apply	173	5.3%	block	173	5.3%
Use	104	3.2%	convenient	122	3.7%
Call	57	1.7%	congruent	54	1.7%
Introduce	44	1.4%	Lagrangian	42	1.3%
Construct	34	1.0%	walk	32	1.0%
Say	27	0.8%	adopt	27	0.8%
Check	7	0.2%	Nash	7	0.2%
Show	5	0.2%	Method	5	0.2%
Prove	5	0.2%	functionally	5	0.2%
Obtain	2	0.1%	unpruned	2	0.1%
Conclude	1	0.0%	Hamiltonianly	1	0.0%

imperatives are, we now investigate the percentage of files they appear in. Recall that our *Proof-Only* corpus consists of 3,268 files. A file in our corpus contains all of the proofs from one file uploaded to the arXiv. Barring a minor complication, we can thus take this to give us a representation of what percentage of papers that contain proofs at all also contain a given imperative somewhere in those proofs.[16]

Once again, to look for imperatives we searched for the capitalised verbs from the first category in Table 2 using a case-sensitive search such that, for example, it will count `Consider' and `consider' separately.

The results are found in Table 5 in the left columns. To put the percentages in some context, we have included an arbitrarily selected word which appears in a similar percentage of the papers, listed on the right.

The results agree with the findings from Section 4.4.1, with the same set of verbs serving most often as imperatives in proofs. The results also show that these imperatives are not just common relative to non-proof mathematical texts, but also common in general, appearing in a substantial proportion of the files of the corpus, and therefore in a substantial percentage of the papers containing proofs.

4.4.4 Imperative Diversity

Our final hypothesis (H4) concerned the diversity of verbs used in imperatives and "We [verb]" constructions in mathematical proofs, alluding to Ganesalingam's claim that they are drawn only from a limited and closed class of verbs. Our more fine-grained results from the previous sections already address this hypothesis to some degree.

Of course, the truth of (H4) depends on what that class is. For example, the English language is finite, so on one reading the claim could be trivially true. We propose that a natural reading of Ganesalingam (2010) in light of our results would be to have the "limited and closed class of verb" be those that we found to be most common in Sections 4.4.1 and 4.4.3, namely: `let', `consider', `assume', `denote', `note', `define', `suppose', `recall', `write', `take', `choose', `fix' and `observe'. From our results in the previous sections, then, it is correct to say that a large number of the instructions in proofs are given with these verbs. This is clearer in the case of imperatives from Sections 4.4.1 and 4.4.3, but also the case in the "We [verb]" constructions.

[16] The minor complication is that if a paper was uploaded as multiple files, then it will remain as multiple files in our corpus. However, only files containing proofs at all will be included in the corpus, which reduces the frequency of this. Only a small minority of the files were uploaded this way, and this is conservative such that the effect on our results is that the percentage of files we report is a slight underestimate of the percentage of papers a given word appears in.

However, the hypothesis of (H4) states that these verbs are the *only* ones used in instructions given by imperatives and “We [verb]” constructions. Above we found many other instances of verbs being used both as imperatives and in “We [verb]” constructions. Here, we offer two more pieces of evidence for there being a diverse number of instructions issued in mathematical proofs.

Table 6

Examples of imperatives from the Proof-Only corpus for the verbs found in the middle column of Table 2.

Verb	Example
Set	Set the total degree equal to the sum of the bidegrees.
Group	Group the non-constant linear generators of inline_math so that […]
Map	Map this product to a neighborhood inline_math of the torus knot as follows […]
Form	Form the commutative cube in which the front and back faces are pullbacks, so that […]
Order	Order the eigenvalues of inline_math so that the inline_math eigenvalue of the spectrum of inline_math is […]
Number	Number the elements of inline_math.
Get	Get rid of the self-intersection as before to obtain as before.
Give	Give the knot inline_math an orientation inline_math.
Find	Find a partition of unity inline_math on inline_math subordinate to the cover inline_math.
Sum	Sum the estimates in the previous corollary.
Do	Do the above coarsening process for every element inline_math with inline_math.
Complete	Complete this elimination process, ending with inline_math.
Pair	Pair these elements and return to inline_math and repeat the process, skipping over any letters already paired.
Estimate	Estimate the difference on the right hand side of by the triangle inequality to find […]
Study	Study the case inline_math.
Factor	Factor inline_math as a product of prime ideals inline_math in inline_math.

First of all, in Section 4.3 we set aside 22 of our 91 verbs because in the *All-Math* corpus they were used in their capitalised form both as imperatives and in some other ways, which would’ve obscured the above studies. These were the verbs listed in the middle column of Table 2. Returning to these verbs, we searched for them in their capitalised forms in the *Proof-Only* corpus. Of these, 16 returned results that were clearly imperatives.[17] For each of these we selected a clear example, displayed in Table 6. Given that these examples sound like normal written mathematics, this demonstrates further diversity of the verbs used as imperatives in the *Proof-Only* corpus.

[17] We avoid giving the numbers of imperatives because not all of the cases were clear, and mostly the numbers were low. That is not important, though, as here we are interested in diversity of imperatives rather than the frequency.

The second way we chose to look at the diversity of verbs used as imperatives in proofs was using the mathematical verbs listed by Rotman in the quote in Section 3.1. We omit all of those that overlapped

Table 7

Comparing the Rotman verb imperative frequencies per million words in the Proof-Only and Non-Proof Corpora.

Verb	Proof-Only	Non-Proof
Add	2.01	2.28
Adjoin	0	0.08
Assign	1.81	1.12
Calculate	0.80	1.50
Compute	2.41	6.29
Connect	2.41	0.58
Count	0	0.15
Delete	0.60	0.81
Demonstrate	0	0.04
Differentiate	0.40	0.58
Draw	0.60	0.77
Drop	0	0.23
Eliminate	0	0.12
Enumerate	2.21	1.00
Evaluate	0.20	0.62
Exhibit	0	0.00
Integrate	3.62	0.50
Iterate	1.01	0.31
List	0.40	2.16
Multiply	7.84	0.93
Reverse	0.20	0.58
Specify	0	0.46
Subtract	1.21	0.12
TOTAL	27.74	21.22

with the previous analysis. Like in Section 4.4.1, we compared the frequency per million words of the capitalised versions of the verbs between the *Proof-Only* and the *Non-Proof* corpora. The results are displayed in Table 7. These verbs appeared significantly more often per million words in proofs than outside of proofs, Fisher's exact test, p = .006.

One striking feature here is that the verbs suggested by Rotman are orders of magnitude less common than those selected by our method above, suggesting that while these may be mathematical words broadly speaking, they are not particularly common ones. More importantly for (H4), we did find that many of them were present as imperatives in our *Proof-Only* corpus. Once again, this result suggests that many more verbs are amenable to being used as imperatives in mathematical proof texts. We thus reject (H4), as the class of verbs does not appear to be closed or limited.

4.5 An Example

It is easy to lose oversight of what these results look like in practice. To address this, and provide some reflection on the methodological choices made above, we will now demonstrate the use of imperatives and instructions in our corpus with an illustrative example. We will look at the preprint of the paper by Mazzeo & Rowlett (2015), selected as the chronologically first paper in our corpus to contain at least one proof.

Here is an excerpt from the introduction section of the paper. This contains three imperatives ("For simplicity, suppose", "Let", and "Note") of which our analysis using capitalisation would have found two. There are also two "We [verb]" constructions ("We now explain" and "We assume"), of which we would have found one. Interestingly, in the context of an introduction, the "We [verb]" phrases are not acting to include the reader in epistemic actions, but instead are simply describing what the authors are doing. What is also noteworthy is that this paragraph is relatively high in instructional language, but is also *proof-like* in the sense that it is describing a piece of mathematical reasoning. Looking further through the

Figure 8 An excerpt from the introduction of Mazzeo & Rowlett's (2015) preprint.

> We now explain the desingularization more precisely. For simplicity, suppose that Ω_0 and Ω_ϵ all lie in some slightly larger ambient open surface $\widetilde{\Omega}$, and that the metrics g_ϵ on Ω_ϵ are all extended to metrics (still denoted g_ϵ) on this larger domain. We assume that this family of metrics converges smoothly on $\widetilde{\Omega}$. Let p be a vertex of Ω_0 and consider the portion of Ω_ϵ in some ball of fixed size around p, $B_c(p) \cap \Omega_\epsilon$. Our main assumption is that the family of pointed spaces $(B_c(p) \cap \Omega_\epsilon, \epsilon^{-2} g_\epsilon, p)$ converges in pointed Gromov-Hausdorff norm, and smoothly, to a noncompact region $Z \subset \mathbb{R}^2$ with smooth boundary, such that at infinity, ∂Z is asymptotic to a cone with vertex at 0 and with opening angle α, the same angle as at the vertex p in (Ω_0, g). Note that this is actually pointed Gromov-Hausdorff convergence for the ambient space $(\widetilde{\Omega}, g_\epsilon, p)$.

paper, this seems to be the case for a large number of the imperatives being used. Direct imperatives appear in theorem statements, lemma statements, numbered remarks, and in the long paragraphs of description of mathematical reasoning. Since our analysis placed all of those in the *non-proof* category, the results we have described are likely to understate the extent to which imperatives are characteristic of proofs in research papers, since this style of proof writing can spill over into the surrounding text.[18]

The first substantial proof is of Theorem 1.6, which spans over several pages. Here are its opening lines:

Figure 9 An excerpt from the proof of Theorem 1.6 of Mazzeo & Rowlett's (2015) preprint.

Proof of Theorem 1.6: We first construct a particular family of parametrices for the heat kernel on Ω_ϵ. For any $0 \leq \epsilon < \epsilon_0$, decompose

$$\Omega_\epsilon = \Omega_{\epsilon,1} \cup \Omega'$$

where $\Omega_{\epsilon,1} = \Omega_\epsilon \cap B_1(p)$, and $\Omega' = \Omega_\epsilon \setminus (\Omega_\epsilon \cap B_1(p))$. Note that Ω' is independent of ϵ. Lemma 2.2 shows that

$$H^{\Omega_\epsilon}(t,z,z) = \chi_1(z) H^{\epsilon Z}(t,z,z) + \chi_2(z) H^{\Omega_0}(t,z,z) + K(t,z), \qquad (3.1)$$

The proof contains eight direct imperatives (using the verbs "decompose", "note", "write", "set", "choose", and "let"), of which our analysis would have found five. In fact, we have not looked for imperatives using the verb "decompose" above, so this reinforces our finding of diversity of imperative phrases. The proof also contains ten instances of using "we" with a verb to indicate activities (using the verbs "construct", "denote", "see", "defer", "examine", "write", "prove", "appeal", and "find"). Surprisingly, only two of these were in the exact "We [verb]" form, and most were embedded in much more complicated sentence structures. This is coherent with our analysis that we did find "We [verb]" phrases in both the *proof-only* and the *non-proof* corpora, but suggests that further analysis might reveal a richer picture. Indeed, this proof also contains several phrases that suggest epistemic and mathematical actions in other ways: "For that, we may as well replace the upper limit of integration by [...]", "To finish the proof, we must analyze the behaviour of [...]", and "The decay of each of term as $|\lambda| \to \infty$ is straightforward, so we can write [...]". These three all make use of modal language, something examined

[18] Our thanks to a referee for pushing us to be explicit about this important point.

by Hodges (2013). Computational corpus linguistics may be a suitable method for capturing these more complex constructions, but would require more advanced tools than we have used here.

We also note that at the time of writing, Keith Weber (submitted) has carried out a "medium-scale" analysis of the use of instructions and imperatives in the entirety of Kunen's classic textbook *Set theory: An introduction to independence proofs* (1980). This will provide a useful piece of methodological triangulation for our results, something which Löwe & Van Kerkhove (2019) have argued is centrally important for empirical approaches to mathematical practice. While our computational analysis allows us to examine trends across a large number of papers, Weber's approach does not rely on automation and thus allows him to find and analyse instructions our methods wouldn't catch. Indeed, Weber is able to carry out an interesting investigation of the different kinds and functions of instructions he finds, providing different insights into the instructional language of proofs. Nonetheless, Weber's results largely confirm our results, finding both a high frequency and broad diversity of instructions and imperatives in Kunen's text. There are some differences in the exact distribution of the imperatives he finds, but this is easily explained by the fact that this is a single author with certain stylistic preferences. In light of Weber's work, we can feel more confident that the various small methodological choices we have made above have delivered reliable results.

5. Discussion

Our analysis of the language used in a large number of mathematics papers allows us to make empirically informed contributions to the philosophy of mathematical practice, and especially questions concerning the nature of real, modern mathematical proofs. We now relate our findings to the philosophical debates discussed earlier in the chapter.

Recall that we were interested in the recipe model of proofs, which tells us that proofs are akin to recipes, commanding that we undertake various kinds of mathematical actions to perform a piece of reasoning. These actions were not limited to inferential actions, but also include constructing and manipulating mathematical objects, fixing notations, entertaining hypotheticals, structuring their reasoning, and more besides. Viewing proofs in this way has important implications for understanding the nature of proofs, the epistemology of proofs, the functions of proofs, how to best educate students to engage with proofs, and how to best integrate formal proof checkers into mathematical practice. As such, while it may appear to be a mere stylistic variation, we hold that the specifics of how proofs can be written is of substantial importance in the philosophy of mathematical practice.

Based on the recipe model, we predicted that imperatives and instructions are both common and diverse in mathematical proof texts. The results of our study certainly lend support to the recipe model because we did indeed find that imperatives and instructions are used often and using a broad range of verbs. However, there were a core set of more common imperatives using the verbs `let’, `consider’, `assume’, `denote’, `note’, `define’, `suppose’, `recall’, `write’, `take’, `choose’, `fix’ and `observe’. This core set of verbs make up the bulk of the imperatives we found in our study, suggesting that these imperatives are used in fairly standardised ways for particular mathematical instructions. Nonetheless, we found many different verbs used as imperatives in our corpus. While many of these were used less frequently, it does indicate that there is also the flexibility in written proofs to deploy a diverse selection of instructions using imperatives.

Finding broad empirical support for the recipe model of proofs is also beneficial for the epistemic picture underlying it. If proofs are giving us instructions to carry out, then mathematical knowledge is more intimately tied to knowledge-how than is standardly recognised.

Our results should also be of interest to the formal mathematics community. If part of the aim of formal proof checkers is to gain widespread use among mathematicians by making the formal tools better at reading and producing natural-sounding proofs, then we will need accurate information on how those proofs do and ought to look. The results in this chapter suggest that such natural-sounding formally-verifiable proof should contain imperatives and “We [verb]” constructions, especially those most common verbs.

We should re-emphasise a note of caution from above. Our results are for papers from the arXiv, and so represent a particular kind of mathematical text. It is thus not clear whether they would generalise to other contexts, such as proofs in textbooks, classrooms, or seminars. We can speculate that journal proofs may be at the more formal end of mathematical language, and therefore we might expect greater flexibility in terms of the verbs used to form instructions in other contexts. Research articles are also a context where one assumes a great deal of competence from the reader, so we might also hope that proofs aimed more at teaching will include more imperatives as a communicative aid.

To conclude, we have made progress in investigating real mathematical proofs in a systematics and large-scale way. If we truly want our philosophy of mathematics to attend to the realities of practice, then these methods should be integral to uncovering what those realities are.

Bibliography

Aberdein, A., & Inglis, M. (eds.) (2019) *Advances in Experimental Philosophy of Logic and Mathematics*, London: Bloomsbury.

Alcock, L.; Inglis, M.; Lew, K.; Mejía-Ramos, J. P.; Rago, P.; & Sangwin, C. (2017) "Comparing expert and learner mathematical language: A corpus linguistics approach", in A. Weinberg; C. Rasmussen; J. Rabin; M. Wawro; & S. Brown (Eds.) *Proceedings of the 21 st annual conference on Research on Undergraduate Mathematics Education*, San Diego, CA: RUME, pp. 478–484.

Andersen, L. E. (2020) "Acceptable gaps in mathematical proofs", *Synthese* 197, pp. 233–247.

Anthony, L. (2020) *AntConc* (Version 3.5.9) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

De Lon, A.; Koepke, P.; and Lorenzen, A. (2020) "Interpreting Mathematical Texts in Naproche-SAD", in C. Benzmüller and B. Miller (Eds.) *CICM 2020, LNAI 12236*, pp. 284–289. https://doi.org/10.1007/978-3-030-53518-6_19

De Toffoli, S. (2020) "Groundwork for a Fallibilist Account of Mathematics", *Philosophical Quarterly,* pp. 1-22. https://doi.org/10.1093/pq/pqaa076

Ernest, P. (1998) *Social Constructivism as a Philosophy of Mathematics*, New York: State University of New York Press.

Ernest, P. (2018) "The Ethics of Mathematics: Is Mathematics Harmful?", in Paul Ernest (ed.) *The Philosophy of Mathematics Education Today*, Cham: Springer Nature, pp. 187-216.

Fallis, D. (2003) "Intentional Gaps in Mathematical Proofs", *Synthese* 134, pp. 45–69.

Frerix, S., & Koepke, P. (2018) "Automatic Proof-Checking of Ordinary Mathematical Texts", in Hasan, O.; Kaliszyk, C.; and Naumowicz, A. (eds.): Proceedings of the Workshop Formal Mathematics for Mathematicians (FMM), Hagenberg, Austria, 13-Aug-2018, published at http://ceur-ws.org

Gerhardt, C.; Frobenius, M.; & Ley, S. (eds.) (2013) *Culinary Linguistics. The chef's special.* Amsterdam: John Benjamins Publishing Company.

Halmos, P. (1970) "How to Write Mathematics", *L'Enseignement Mathématique* 16, pp. 123-152.

Hamami, Y. (2014) "Mathematical Rigor, Proof Gap and the Validity of Mathematical Inference", *Philosophia Scientiæ* 18, pp. 7-26.

Hodges, W. (2013) "Modality in Mathematics", *Logique et Analyse* 56, pp. 5-23.

Imao, Y. (2018) *CasualConc* (Version 2.1.0) [Computer Software]. Osaka, Japan: Osaka University. Available from https://sites.google.com/site/casualconc/

Knuth, D. E.; Larrabee, T.; & Roberts, P. M. (1989) *Mathematical Writing*, MAA Notes Number 14, Washington, DC: Mathematical Association of America.

Krantz, S. G. (1997) *A Primer of Mathematical Writing*, Providence RI: American Mathematical Society.

Kunen, K. (1980) *Set theory: An introduction to independence proofs.* Amsterdam: Elsevier.

Lew, K., & Mejía-Ramos, J. P. (2019) "Linguistic Conventions of Mathematical Proof Writing at the Undergraduate Level: Mathematicians' and Students' Perspectives", *Journal for Research in Mathematics Education* 50, pp. 121-155.

Löwe, B., & Van Kerkhove, B. (2019) "Methodological triangulation in empirical philosophy (of mathematics)", in Aberdein, A., & Inglis, M. (eds*.) Advances in experimental philosophy of logic and mathematics*, London: Bloomsbury, pp. 15-37.

Mazzeo, R., & Rowlett, J. (2015) "A heat trace anomaly on polygons", *Mathematical Proceedings of the Cambridge Philosophical Society* 159, pp. 303-319. https://arxiv.org/pdf/0901.0019.pdf

Mejía-Ramos, J. P.; Alcock, L.; Lew, K.; Rago, P.; Sangwin, C.; & Inglis, M. (2019) "Using corpus linguistics to investigate mathematical explanation", in Fischer, E., & Curtis, M. (eds.) *Methodological advances in experimental philosophy*, London: Bloomsbury, pp. 239-264.

Mejía-Ramos, J. P., & Inglis, M. (2011) "Semantic contamination and mathematical proof: Can a non-proof prove?", *The Journal of Mathematical Behavior* 30, pp. 19–29.

Morgan, C. (1996) "`The Language of Mathematics': Towards a Critical Analysis of Mathematics Texts", *For the Learning of Mathematics* 16, pp. 2-10.

Pimm, D. (1987) *Speaking Mathematically: Communication in Mathematics Classrooms*, London: Routledge & Kegan Paul.

Reyes-Galindo, L. I. (2016) "Automating the Horae: Boundary-work in the age of computers", *Social Studies of Science* 46, pp. 586–606.

Rittberg, C. J.; Tanswell, F. S.; & Van Bendegem, J. P. (2018) "Epistemic injustice in mathematics", *Synthese* 197, pp. 3875–3904. https://doi.org/10.1007/s11229-018-01981-1

Rotman, B. (1988) "Towards a Semiotics of Mathematics", *Semiotica 72*, pp. 1-35.

Rotman, B. (1993) *Ad Infinitum... The Ghost in Turing's Machine*, Stanford: Stanford University Press.

Swales, J. M.; Ahmad, U.; Chang, Y.; Chavez, D.; Dressen-Hammouda, D.; and Seymour, R. (1998) "`Consider This...': The role of imperatives in scholarly writing", *Applied Linguistics* 19, pp. 97-121.

Tanswell, F. S. (2016) *Proof, Rigour and Informality: A Virtue Account of Mathematical Knowledge,* PhD thesis, University of St Andrews. https://research-repository.st-andrews.ac.uk/handle/10023/10249

Tanswell, F. S. (2018) "Conceptual Engineering for Mathematical Concepts", *Inquiry* 61, pp. 881-913.

Tanswell, F. S. (forthcoming) "Go Forth and Multiply: On Actions, Instructions and Imperatives in Mathematical Proofs", in Brown, J., & Bueno, O. (eds.) *Essays on the Philosophy of Jody Azzouni*, Cham: Springer.

Tanswell, F.S., & Rittberg, C.J. (2020) "Epistemic injustice in mathematics education", *ZDM Mathematics Education* 52, pp. 1199–1210. https://doi.org/10.1007/s11858-020-01174-6

Wagner, R. (2009) *S(zp, zp): Post-Structural Readings of Gödel's Proof*, Milan: Polimetrica International Scientific Publisher.

Wagner, R. (2010) "Who speaks mathematics: a semiotic case study", *Philosophical Perspectives on Mathematical Practice* 12, pp. 205–234.

Weber, L. (submitted) "Instructions and constructions in set theory proofs".

Acknowledgements

We are grateful to Lara Alcock, Kristen Lew, Pablo Mejía-Ramos, Paolo Rago, and Chris Sangwin for the assembly and processing of the corpus with which we conducted this work. We would also like to thank Laurence Anthony for the creation and free distribution of the *AntConc* software with which we carried out our main analysis. Likewise, we thank Yasu Imao for the creation and free distribution of *CasualConc*, with which we carried out some subsidiary analyses. We thank audiences in Prague and online via Copenhagen for helpful feedback. We are grateful to Valeria Giardino and Roy Wagner for the careful refereeing of this paper. We also had comments and suggestions from a large number of friends and colleagues, to whom we also extend our thanks.