The Humanity in Mathematicians' Writing

Kathleen Melhuish
Texas State University

Andrew Baas
Texas State University

Mario Gonzalez
McNeese State University

Elizabeth Roan
Texas State University

Blake Dvarishkis
Texas State University

Omkar Devlekar
Texas State University

Mathematical writing is often portrayed as objective and authorless, yet prior work shows that mathematicians' writing contains nonobjective and authority-related features (Burton & Morgan, 2000). We build on this foundation by analyzing 51,325 single-authored mathematics papers from arXiv, focusing on proofs and surrounding narratives. Using curated dictionaries of hedges and evaluative adjectives and mathematics-specific rules, we document their frequency and explore whether gender and publishing output of the author relate to their use. Preliminary results indicate that both features appear to vary with author characteristics. These findings suggest that linguistic choices in mathematics are tied to identity and status, complicating the myth of purely objective discourse.

Keywords: proof genre, mathematician practice, language analysis

The discourse about mathematical proof and writing has often espoused values of author independency and its absolute nature (Dawkins & Weber, 2017). Davis and Hersh (1998) are often quoted for describing the ideal mathematician's writing as "conceal[ing] any sign that the author or the intended reader is a human being" (p. 2). Yet, others point to the fact that mathematical writing and proof are very much the product of humans both in terms of communication styles (Tanswell & Rittberg, 2020) as well as the infusion of authority markers (Burton & Morgan, 2000). That is, a product created by humans cannot be truly dehumanized. Students come to mathematics with diverse linguistic communication styles and rhetorical patterns, some of which align more closely than others with the conventions of proof and mathematical writing. For those whose styles diverge, there may be additional cognitive load in adapting to the ways of writing that have become normative in mathematics. Thus, it is important to understand the features of mathematical writing beyond the idealistic dehumanized products.

This preliminary work builds on Burton and Morgan's (2000) analysis of mathematicians' papers and interviews, which showed that presenting writing in research articles is not dehumanized. They identified markers of authority, such as positioning something as "obvious" (positive authority) or expressing doubt and qualification (negative authority). They suggested that status shapes their use. For example, women may avoid negative authority because their belonging in the field is less secure. In the 25 years since, little work has examined how author identity relates to linguistic choices in mathematical writing. With large corpora now available, we can investigate such questions at scale. In this preliminary report, we focus on two authority markers (hedging and evaluative language) asking: (1) What forms of hedging and evaluative language appear in proofs and surrounding narratives? (2) How do author characteristics such as gender and publication record relate to their use?

Literature Background on Mathematician Writing

Many scholars have focused on the proof product as a particular genre of writing with many distinct linguistic conventions (e.g., Bowers & Küchle, 2020; Selden & Selden, 2013) with

slightly different norms (Lew and Mejía-Ramos, 2019; 2020) depending on context (e.g., student proof or proof in a research article). If we turn to studies specifically on mathematician writing in publications, we find few studies with the majority in linguistic journals. Studies have explored publication patterns in terms of genders of authors and relationship to journal quality (Mihaljević-Brandt, et al., 2016), revision processes (Andersen et al., 2021; McGrath, 2016), and corpus analyses of particular features such as explanatory language (Mejía-Ramos et al., 2019) and use of imperatives (Tanswell & Inglis, 2023). Furthermore, scholars have established that mathematical research articles are quite distinct from standard forms lacking traditional structures such as "Methods" or "Discussion" (Graves et al., 2013) and features such as establishing research gaps (Graves et al., 2014) as well as different types of phrase frames (e.g., "is [also, fairly] easy to see that", Cunningham, 2017). Mathematics articles can be deconstructed into two types of writing, mathematical (e.g., proofs, definitions) and meta-mathematical (the surrounding narrative) (Kuteeva & McGrath, 2015). Andersen et al. (2021) studied the revision process of a graduate student and his advisor writing an article identifying how the mathematical and meta-mathematical have different roles and ways of justifying and engaging with the reader. This suggests any analysis of research articles should attend to these types of writing separately.

If we turn to our focus, we identified Burton and Morgan (2000), McGrath and Kuteeva (2012) and Alasmary (2022) as empirical studies that address authority in mathematical research articles in ways that are relevant to language choices such as hedging and evaluating. They all support the larger claim that mathematical writing is not purely objective in mathematical or meta-mathematical sections. McGrath and Kuteeva used Hyland's (2005) stance and engagement framework to provide counts of words in a set of 25 research articles finding frequent use of imperatives, infrequent use of hedges, but frequent use of boosters ("really") and attitude devices ("obviously"). Alasmary compared graduate level texts, doctoral dissertations, and a set of published research articles (34 articles) via analysis of four-word lexical bundles finding that participant-oriented (affective and engagement with reader) were found the least, but they were found more often in journal articles than in the dissertation set. Burton and Morgan's (2000) qualitative analysis of 59 articles similarly found establishing positive authority (affective judgments such as "obvious") was more common than negative (e.g., hedging). Paired with interviews, they also provided the conjectures we discussed in the introduction.

Theoretical Orientation

We situate this work as a direct extension of Burton and Morgan (2000), who used Halliday's (1973) systematic-functional grammar to analyze 53 mathematical papers. They considered both ideational (presentation of mathematics) and interpersonal functions (relationships between author, subject, and audience). Our interest lies in the interpersonal function, particularly authority. Burton and Morgan showed how authors positioned themselves with positive or negative authority: deferring to others, signaling incompleteness, or hedging (negative), versus asserting expertise through evaluative words such as clearly or obvious (positive). Such markers disrupt the notion that mathematical writing is purely objective.

Burton and Morgan documented variations across their sets of papers in terms of both positive and negative authority markers. In their sample, men more often used markers of negative authority. They conjectured that, women are positioned as not fully belonging in the field and therefore cannot risk using the traditional language of mathematical power. They further suggest that negative authority provides a risk and that, "risk is, of course, relative to the author's preexisting status within the community" (p. 441). They supported this conjecture with

the fact that their outlier in terms of many positive and negative authority markers was a well-established male mathematician in the field. High-status likely means someone's authority is not going to be questioned regardless of using negative authority or even positive authority (which can be perceived as too self-assured for a novice mathematician.)

Building on this foundation, we focus on two authority markers well-studied in other fields: hedging and evaluative language. Hedges (e.g., might, suggests) serve to "express tentativeness" and to qualify "confidence in the truth of a proposition." (Hyland, 1996, p. 433). They serve as a means of negative authority. The author is claiming they do not have complete knowledge of the field. In contrast, evaluative language (e.g., clear, important) positions the author with positive authority. These are adjectives that serve for to assign a value-judgment (Stojanovic, 2015) on a noun and tend to be gradable (scaled between two polarities). Following Burton and Morgan (2000), we make several conjectures. First, hedging language will be more prominent in papers written by men and well-established mathematicians. Second, evaluative language will vary with experience in the field.

Methods

To explore this question, we make use of SIGMathLing's 1.58 million document data set from arXiv.org (Ginev, 2020) years ranging from 1990 to 2020. As we were interested in automatically extracting proofs, we removed all papers that did not use the LaTeX "proof" environment or which had poorly formatted proof environments, yielding a set of 277k papers. For the scope of this paper, we then narrowed this data set to papers whose subject area was research mathematics, excluding papers whose primary subject was not prefixed with "math" and papers whose subjects were in computer science or philosophy. We also filtered to papers that were single authored with identifiable gender in a final dataset of 51,325 papers. To identify gender, we ran author first-names through gender-guesser.py, a python library based on a list of 40k names alongside genders collected by Jörg Michael (2007). We note the limitations of this binary approach and making such strong inferences from names. We then organized authors based on how many papers they had on arXiv into deciles to document output as a proxy for establishment in the field of mathematics. The lowest decile was labeled "Vlow," the next two as "low," deciles four through seven as "mid," deciles eight and nine as "high" and the 10th decile as "Vhigh." For each paper, we then subdivided the portions into "proof" and "nonproof" based on the usage of the proof environment in LaTeX. We also removed any mathematical symbols and replaced them with the generic "formula."

To measure "hedging" and "evaluative" language, we began with existing dictionaries from other projects (Ulinski and Hirschberg, 2019 and Rockladge, et al. 2018, respectively). Initially, we used these sets of words to calculate their frequencies in the entire set of mathematics papers. From this process, it became apparent that words that may have served as hedges or evaluative language in non-mathematical contexts were not always adaptable directly. For example, "estimate" may be a hedge in some contexts, but in mathematics papers was often referring to a precise output of an algorithm. Similarly, a word such as "ideal" would be an evaluative adjective in most settings but is more frequently an algebraic structure in mathematics papers. This necessitated a careful look at each hedge and evaluative adjective. For each word, we selected a sample of papers that used the word to evaluate the contexts. The research team then indicated if the usage seemed to be meeting the definition of hedge or if it had a more precise mathematical meaning. This led to developing and testing several rules. For example, "most" was identified as a hedge when it was part of the phrase "most of" rather than in contexts where

it was describing something precise. In other cases, simple rules did not function well such as was the case for "may." When "may" operates like "might" it is a hedge. Yet, it commonly operated like "can" where something is mathematically permissible. For this hedge, we used semantic embedding to encode the words around each usage in a vector and to compare its distance to the average vectors of "might" and "can". If "may" was much closer to might, it was maintained as a hedge. Presently, we are in the middle of this process reporting a set of 13 hedges and 12 evaluative adjectives (see Results) that we are successfully picking up as intended in our corpora. For each of the hedges and evaluative adjectives, we counted their frequency for each paper in the proof section and the narrative section. We also calculated relative frequency by dividing the frequencies by the total number of words in each respective section. Each of these are treated as outcome variables where gender and publication level were predictor variables. For the scope of these preliminary reports, we provide descriptive statistics and a preliminary logistic regression on the binary of existences of hedging or evaluative adjectives.

Results

In our analysis of 51,325 papers, we found that 66.8% of them had at least one hedge (of our thirteen) in the narrative sections, and 8.6% of the papers had at least one hedge in the proof sections. We found that 88.3% had evaluative adjectives (of our twelve), and 62.7% had at least one evaluative adjective in their proof sections. Figure 1 provides counts of the evaluative adjectives and hedges (meeting our rule requirements) from the analysis.

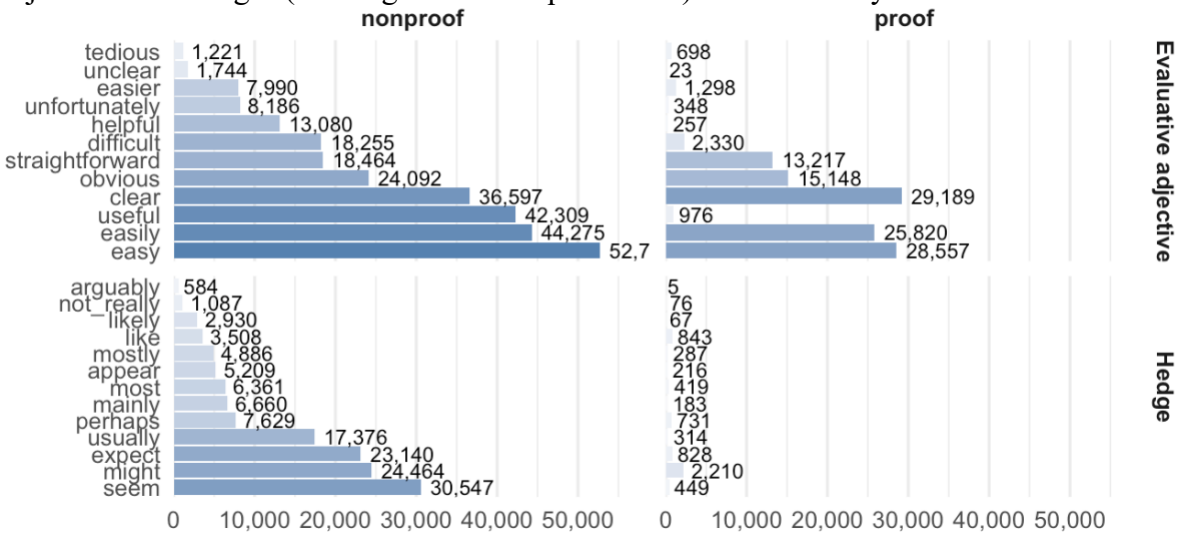

Figure 1. Frequencies of a Preliminary Set of Hedges and Evaluative Adjectives

In Figure 2, we present boxplots representing the relative hedge use (relative to wordcount) broken out by gender and then publishing category for non-proofs. Note that the proof section has so infrequent hedges that these are not boxplots but outliers. In Figure 3, we provide the boxplots for Evaluative Adjectives Relative Frequencies in both proof sections and on proof sections.

Presently, we have run only preliminary models looking at the relationship between gender, publishing output, and the focal language. For each proof and each nonproof, we gave a set of binary variables (has hedge; has evaluative adjective). We found that gender and publishing output were significantly (at the p=0.05 level) related to whether we identified evaluative adjectives in the mathematical (proof) and metamathematical (non-proof) parts of a paper. We

found that both were also predictive of the existence of hedge language in the metamathematical (non-proof) section, but not significantly predictive of hedge language in the mathematical (proof) section. To provide a brief illustration, we share the results for existence of evaluative language in proof whose equation is: $\text{odds}(eval)=1.227\times(1.248)^{GenderMale}\times(1.039)^{publishing\_group}$. According to this model, a woman who is in a low output publishing group would be estimated to have a probability of 56% of using evaluative language in the proofs of her paper. A man in the same publishing group would be estimated at 61%. At the other end, a woman in the high publishing group would have a 59.8% chance of using evaluative language in their proof while their male counterpart would have a probability of 65%.

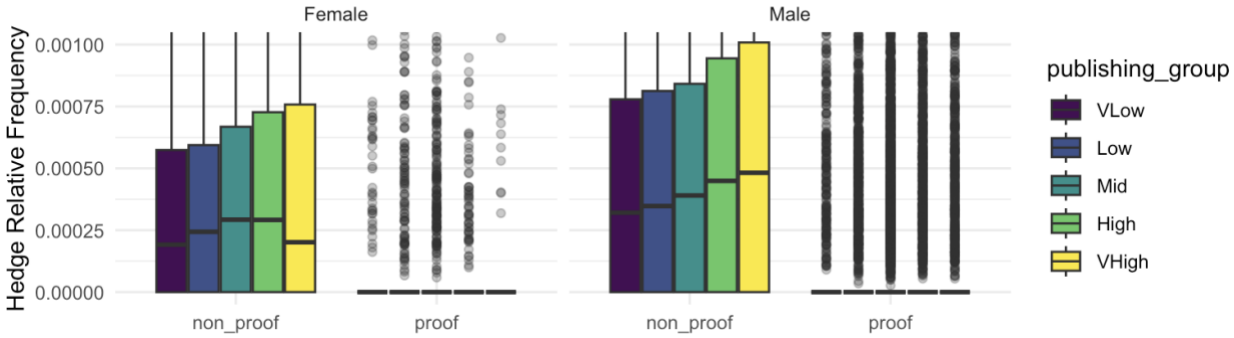

Figure 2. Relative Frequency of Hedges Across Gender and Publication Output Group Narrative Sections.

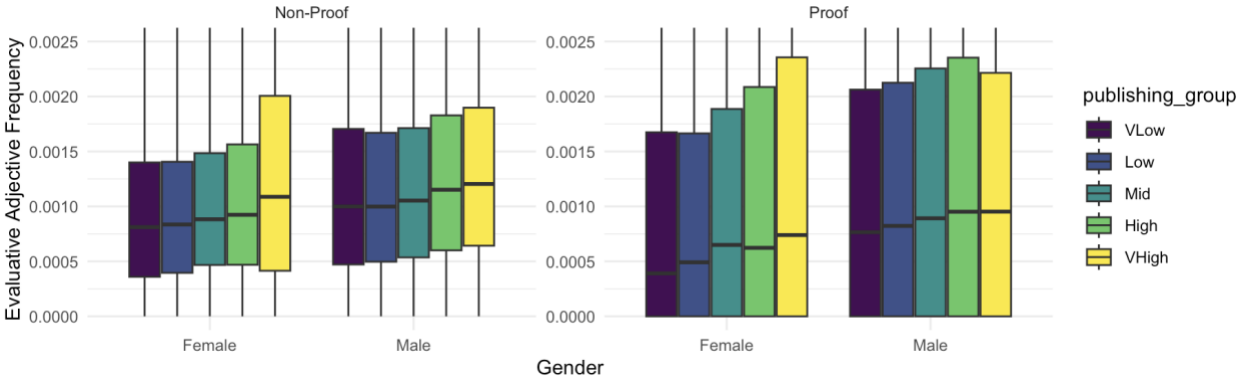

Figure 3. Relative Frequency of Evaluative Adjectives Across Gender and Publication Output Group in Narrative (left) and Proofs (right).

Preliminary Nature of Proposal and Discussion

In this proposal, we are presenting an in-progress analysis of a large corpus of mathematician writing. Our goal was to directly build on Burton and Morgan's (2000) results and conjectures about authority in mathematical writing. We found that the types of markers mentioned by Burton and Morgan did appear in the larger data sets at rates high enough to conduct further analysis. Furthermore, some of their conjectures about status in the field are proving viable. Currently, we have only processed a total of 25 words for this analysis. We intend to continue this process and gather feedback from the RUME audience about how we might consider other factors related to status and address authority markers beyond the set of hedges and evaluative adjectives. We believe the community will find the methods involved interesting. Furthermore, as both scholars and instructors, these results can challenge how we discuss proof and its values based on the reality of how proofs are communicated in published journals.

References

Alasmary, A. A. (2022). Comparing lexical bundles across three advanced mathematical text types: A corpus-based genre-focused investigation. SAGE Open, 12(3), 21582440221113824.

Andersen, L. E., Andersen, H., & Sørensen, H. K. (2021). The role of testimony in mathematics. Synthese, 199(1), 859-870.

Andersen, L. E., Johansen, M. W., & Sørensen, H. K. (2021). Mathematicians writing for mathematicians. Synthese, 198(Suppl 26), 6233-6250.

Bowers, D. M., & Küchle, V. A. (2020). Mathematical proof and genre theory. The Mathematical Intelligencer, 42(2), 48-55.

Burton, L., & Morgan, C. (2000). Mathematicians writing. Journal for Research in Mathematics Education, 31(4), 429-453.

Dawkins, P. C., & Weber, K. (2017). Values and norms of proof for mathematicians and students. Educational Studies in Mathematics, 95(2), 123-142.

Ginev, D. (2020). arXMLiv:2020 dataset, an HTML5 conversion of arXiv.org. SIGMathLing. https://sigmathling.kwarc.info/resources/arxmliv

Graves, H., Moghaddasi, S., & Hashim, A. (2013). Mathematics is the method: Exploring the macro-organizational structure of research articles in mathematics. Discourse Studies, 15(4), 421-438.

Graves, H., Moghaddasi, S., & Hashim, A. (2014). "Let G=(V, E) be a graph": Turning the abstract into the tangible in introductions in mathematics research articles. English for Specific Purposes, 36, 1-11.

Halliday, M. (1973), Explorations in the Functions of Language, Edward Arnold.

Hyland, K. (1996). Writing without conviction? Hedging in science research articles. Applied linguistics, 17(4), 433-454.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. Discourse studies, 7(2), 173-192.

Lew, K., & Mejía Ramos, J. P. (2020). Linguistic conventions of mathematical proof writing across pedagogical contexts. Educational Studies in Mathematics, 103(1), 43-62.

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(2), 121-155.

McGrath, L. (2016). Open-access writing: An investigation into the online drafting and revision of a research article in pure mathematics. English for Specific Purposes, 43, 25-36.

McGrath, L., & Kuteeva, M. (2012). Stance and engagement in pure mathematics research articles: Linking discourse features to disciplinary practices. English for Specific Purposes, 31(3), 161-173.

Mejía-Ramos, J. P., Alcock, L., Lew, K., Rago, P., Sangwin, C., & Inglis, M. (2019). Using corpus linguistics to investigate mathematical explanation. Methodological advances in experimental philosophy, 239-264.

Mejía-Ramos, J. P., Alcock, L., Lew, K., Rago, P., Sangwin, C., & Inglis, M. (2019). Using corpus linguistics to investigate mathematical explanation. Methodological advances in experimental philosophy, 239-264.

Michael, J. (2007, August 6). 40 000 Namen – Anredebestimmung anhand des Vornamens. c't, 2007(17), 182–183. Heise. http://www.heise.de/ct/ftp/07/17/182/

Mihaljević-Brandt, H., Santamaría, L., & Tullney, M. (2016). The effect of gender in the publication patterns in mathematics. PLoS One, 11(10), e0165367.
Rocklage, M. D., Rucker, D. D., & Nordgren, L. F. (2018). The Evaluative Lexicon 2.0: The measurement of emotionality, extremity, and valence in language. Behavior research methods, 50(4), 1327-1344.
Selden, A., & Selden, J. (2013). The genre of proof. Mathematics and mathematics education: Searching for common ground, 248-251.
Stojanovic, I. (2015). Evaluative adjectives and evaluative uses of ordinary adjectives. Proceedings of LENLS12: language engineering and natural language semantics. The Japan Society for Artificial Intelligence.
Tanswell, F. S., & Inglis, M. (2023). The language of proofs: A philosophical corpus linguistics study of instructions and imperatives in mathematical texts. In Handbook of the history and philosophy of mathematical practice (pp. 1-28). Cham: Springer International Publishing.
Tanswell, F. S., & Rittberg, C. J. (2020). Epistemic injustice in mathematics education. Zdm, 52(6), 1199-1210.
Tanswell, F. S., & Rittberg, C. J. (2020). Epistemic injustice in mathematics education. Zdm, 52(6), 1199-1210.
Ulinski, M., & Hirschberg, J. (2019, August). Crowdsourced hedge term disambiguation. In Proceedings of the 13th Linguistic Annotation Workshop (pp. 1-5).