The use(s) of 'is' in mathematics

Paul Dawkins, Northern Illinois University
Matthew Inglis, Loughborough University
Nicholas Wasserman, Teachers College, Columbia University

This paper analyzes some of the ambiguities that arise among statements with the copular verb is *in the mathematical language of textbooks as compared to day-to-day English language. We identify patterns in the construction and meaning of* is *statements using randomly selected sample statements from corpora representing the two linguistic registers. In particular, for the grammatical form "[subject] is [noun]," we compare the relative frequencies of the subcategories of semantic relations conveyed by that construction. Specifically, we find that this construction – in different situations – conveys a symmetric relation, an asymmetric relation, or an existential relation. The intended logical relation can only sometimes be inferred from the grammar of the statement itself. We discuss the pedagogical significance of these patterns in mathematical language and consider some strategies for helping students interpret the intended meaning of the mathematical text they hear or read.*

Keywords: mathematical language, corpus analysis, copular verbs

What does 'is' mean in mathematics? This is an important question because 'is' is used much more often in mathematical English than it is in day-to-day English. In both British and American English "is" represents around 1.01% of words (Davies, 2017), however in mathematics research papers the figure is 2.66% (Alcock, Inglis, Lew, Mejia-Ramos, Rago & Sangwin, 2017). The relative frequency of 'is' in mathematics can perhaps be explained by its ability to encode logical relationships. Linguists categorize 'is' as a copular verb, meaning it is used to join an adjective or noun to a subject. While copular verbs are known to be confusing in all languages – they can mean both predication (an asymmetric relation) and identity (a symmetric relation) (e.g., Geist, 2008; Russell, 1919) – they can be especially problematic in mathematics teaching and learning because of potential logical misinterpretations (e.g., Moschkovich, 1999; Schleppegrell, 2007).

Inspection of some ready examples suggests that 'is' can have at least three distinct logical meanings, as biconditional ($\leftrightarrow$), conditional ($\rightarrow$), and existence ($\exists$):

i. In "a square *is* a regular quadrilateral," *is* is intended to represent a biconditional ($\leftrightarrow$) relationship: an object is a square if and only if it is a regular quadrilateral;
ii. In "a square *is* a rectangle," *is* is intended to only represent a conditional ($\rightarrow$) relationship: if an object is a square then it is a rectangle;
iii. In "there *is* a rectangle that's a square" *is* is intended to assert existence ($\exists$): there exists a rectangle that is also a square.

The potential confusion between the biconditional (i) and conditional (ii) interpretations is especially challenging. From our experience, high school geometry students often object to the statement "a square is a rectangle"; however, it is unclear if they do so because they do not recognize the entire set of objects that fulfill the definition of a rectangle (i.e., their concept image of rectangle is at odds with the given definition), or because they interpret this statement as a biconditional, rendering it false. In other words, correctly interpreting a mathematical statement, at times, requires knowing the conveyed relationship prior to reading the statement.

Consider another example. The statement "an isosceles trapezoid *is* a quadrilateral with congruent diagonals" intends to assert a conditional relation and not a biconditional relation. We know this despite the fact that this sentence structure looks nearly identical to the biconditional in (i) above. Even though in many cases one communicates a biconditional by providing a narrowing clarification (e.g., a square is not just a quadrilateral but a *regular* quadrilateral), in this example, the narrowing of quadrilaterals to only those with congruent diagonals is still not narrow enough to be defining.

The major point is that *is* can be logically ambiguous, which means there may be important issues that arise in the teaching and learning of mathematics around use of this word that are worth further consideration. In this paper, we investigate the various uses of and grammatical constructions with the word 'is' in mathematics (in comparison with common English), as a means to reflect on communication in the teaching and learning of mathematics.

A corpus approach

Corpus linguists study language by analyzing large collections of texts – corpora – intended to be representative samples of particular types of language. Our goal here was to compare the usage of *is* in day-to-day English and in mathematical English in pedagogical contexts. To this end we randomly sampled occurrences of *is* from two corpora. We used the Brown and LOB corpora (Kucera & Francis, 1967; Johansson, Leech, & Goodluck, 1978) to represent day-to-day written English and a corpus of mathematics textbooks compiled by Alcock et al. (2017).

Kucera & Francis (1967) compiled the Brown corpus in the 1960s. It contains 500 samples of American English text, totaling around 1 million words, from a balance of sources (e.g., newspaper articles, biographies, government documents and so on). Johansson, Leech and Goodluck (1978) compiled a British English version of the Brown corpus using texts taken from a similar range of sources, and in similar proportions. It too contains around 1 million words. We combined these two corpora to form a supercorpus of day-to-day English.

To study pedagogical language in mathematics, we used the textbook corpus constructed by Alcock et al. (2017). This consists of processed versions of language taken from undergraduate-level textbooks (Alcock et al. describe the process required to convert LaTeX source files into analyzable plain text). All the textbooks were taken from the Open Textbook Library, the College Open Textbooks site, or the American Institute of Mathematics Approved Textbook list. Topics included abstract algebra, analysis, linear algebra, complex analysis, and transition to proof. In total, 21 complete undergraduate textbooks are included in the pedagogical corpus, comprising of 1.5 million words. In order to conduct the analysis reported below, we randomly selected 250 instances of the word *is* from each corpus, together with the surrounding context.

Analytic strategy

The rationale for our analysis strategy was the belief that the comparison of *is* in day-to-day language and pedagogical mathematical language would lead to insights about the kinds of mathematical statements likely to be difficult for students to interpret appropriately. Motivated by our examples of ambiguity described in the introduction, we initially began by coding the randomly selected sample of *is* statements as expressing *symmetric relations* (if and only if), *asymmetric relations* (if, then), or *existential relations* (there is). It became clear that we needed to distinguish an additional fourth category of *verb phrases* such as "is graphed" or "is rolling" since *is* operated as part of the conjugation of another verb rather than as a simple linking verb. Doing so, however, led to the realization that there was great variation among such structures.

One of the most problematic issues with this coding related to the role of verbs in past participle form. For example, in mathematics we use phrases like "is graphed" or "is connected" that consist of *is* followed by a past participle verb. However, the former is a verb phrase expressing the result of past action and the latter is a property attribution where *connected* acts as an adjective. Because mathematicians are careful to define terms like *connected*, this distinction can be made with some certainty. In the Brown and LOB corpora (representing American and British English respectively), we found more challenging statements like "Mrs. Lavaughn Huntley is accused of driving the getaway car used in a robbery of the Woodyard Bros. Grocery." In this case, *accused* could be an adjective describing Mrs. Huntley or a verb phrase describing ongoing action. This distinction appeared much more challenging to apprehend.

The fact that we had to rely on our understanding of mathematical content to recognize that *connected* acted as an adjective led us to develop a two-stage coding process that distinguished words' grammatical form from their operative role in the statements. Doing so helped us differentiate what information the grammatical form of a statement makes available from what information the reader's knowledge of semantic relations must provide. Using the TagAnt software package (Anthony, 2015) – which identifies parts of speech in a corpus – the first stage in our coding process involved determining the subject and object of each of the 500 *is* statements. While both the subject and object often constituted phrases, we identified one representative word as the object of *is* and then categorized each statement by the object word's part of speech (which we shall call the *grammatical category*). The object words were coded as: 1) nouns; 2) adjectives; 3) verb phrase, in gerund or infinitive form; 4) verb phrase, in past participle form; and 5) prepositions. Then, the second stage in our coding process involved analyzing the sentences within each grammatical category by determining the semantic role that the object word played in each *is* sentence (which we shall call the *semantic subcategory*). The semantic subcategory thus identifies the type of relation *is* is intended to convey.

In what follows, we elaborate on the first grammatical category, [Subject] *is* [noun], and its semantic subcategories. We deemed this category to be of particular interest because it involved both a broad range of *semantic variation*, as well as apparent differences between its uses in day-to-day and mathematical language – what we refer to as *register variation*. In other words, we were interested in *is* constructions in which students would have to use semantic cues to infer the logical relations conveyed in the statement because the grammatical cues are ambiguous. This seems more likely to be difficult if there are a variety of possible semantic subcategories and the frequencies of these subcategories differ between day-to-day and mathematical usage.

[Subject] *is* [noun]

In this and the following sections we shall present our analysis of the statements coded in each grammatical category along with the frequency of each category in our sample. We shall begin our discussion with example *is* statements taken from the corpora.

- Example 1 (Ped): "Inlinemath *is* the standard basis for inlinemath."[1]
- Example 2 (Ped): "The definite integral of a constant times inlinemath *is* the constant times the definite integral of inlinemath."
- Example 3 (Ped): "A rational number *is* a fraction built out of integers."
- Example 4 (Ped): "This map *is* an isomorphism because it has an inverse."
- Example 5 (B/LOB): "a distinction must, however, be drawn between that which is traditional and enduring and that which *is* the result of current political necessity."
- Example 6 (Ped): "Show that there *is* one dimensionless product."

[1] The mathematical corpora replaced all mathematical symbols and expressions with "inlinemath" to facilitate search functions and word counts without having to account for the complexity of LaTeX code for mathematical notation (Alcock et al., 2017).

- Example 7 (B/LOB): "And there *is* enough truth in that to set you thinking."

We identified three semantic subcategories of statements of the form "[subject] *is* [noun]" that correspond closely to our original categories: symmetric relation (1-3), asymmetric relation (4-5), and existential statements (6-7).

Symmetric relation

When *is* conveys a symmetric relation, it indicates "is the same as." We present three cases of the symmetric relations because we observe there are subtle variations among them. In Example 1, the subject and object noun phrase both refer to the same mathematical object, so the two are being identified as the same. Here both are understood as singular, though if either involved variables the entire claim may be understood as implicitly quantified. Example 2 similarly conveys that both the subject and object phrases refer to the same object, though in this case that object is a number. In both of these cases, the article *the* before the object noun provides an explicit cue that *is* conveys a symmetric relation. This was common among our sample of statements in the symmetric relation subcategory, as displayed in Table 1. Example 3 portrays how statements conveying symmetric relations can nevertheless use *a* or *an* before the object word. Because the object phrase "a fraction built out of integers" can be taken to define the subject "rational number," the relation is symmetric.

Table 1. Article choice within the symmetric and asymmetric relation subcategories.

Corpus	Ped	B/LOB		Ped	B/LOB
Total symmetric statements with articles (SYM)	31	19	Total asymmetric statements with articles (ASM)	59	32
- SYM with *a/an* before object	2 (7%)	2 (11%)	- ASM with *a/an* before object	53 (90%)	25 (78%)
- SYM with *the* before object	27 (87%)	17 (89%)	- ASM with *the* before object	0 (0%)	3 (3%)

Asymmetric relation

When *is* conveys an asymmetric relation, it signifies "is one of" or "is an element of the set of." Example 4 is a prototypical example of this form because the object noun is preceded by *a* or *an* (see Table 1), which cues that the subject noun is an example of the class specified by the object noun (and not *the* class itself). Example 5 portrays how statements in this subcategory can still use the article *the* before the object word. It uses *is* to say "that" is an example "result of political necessity," meaning *is* conveys an asymmetric relation. Thus, the article on the object noun usually provides a grammatical cue for whether *is* conveys a symmetric or asymmetric relation, but there are both symmetric and asymmetric constructions that use the alternative articles.

Existential relation

Though questions of existence may differ between day-to-day and mathematical contexts, we did not observe semantic ambiguity in statements of this form in either corpus. The phrase "there is" seems to clearly distinguish statements in this subcategory. However, we observed an interesting trend in the frequency of this semantic subcategory, as presented in the next subsection.

Frequencies of this grammatical category and semantic subcategories

Figure 1 presents the frequencies of "[subject] *is* [noun]" statements in our two samples and the relative frequency of each subcategory. This grammatical category was much more common in our sample of mathematical statements, which may reflect mathematicians' tendency to use nominalizations for concepts or processes (Morgan, 1996). There was a significant difference in the balance of subcategories found in each corpus (Fisher's exact test, $p = .001$), symmetric relations occurred with about equal frequencies while mathematics text conveyed asymmetric relations more often and day-to-day text conveyed existence relations more often. The latter fact seems surprising, though we expect this is because mathematicians more often use the more formal "there exists" (instead of "there *is*"), or the symbol ∃, since existential claims are by no means scarce in mathematics text.

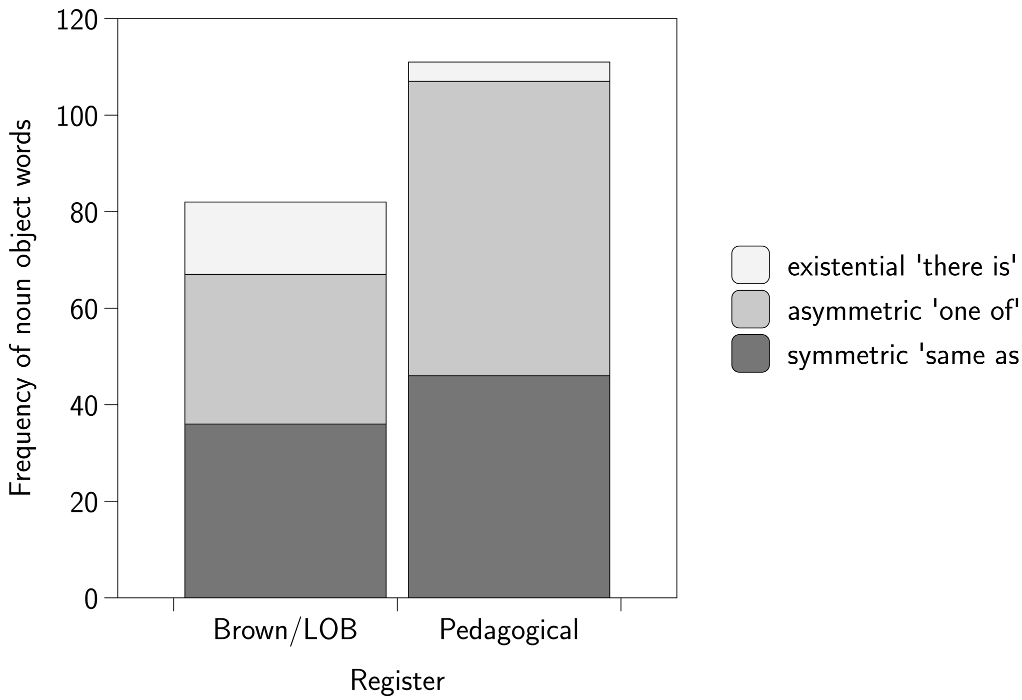

Figure 1: Frequencies of noun object words and subcategories thereof.

Quantification in "A [Subject] is a [noun]" constructions

The construction that began our investigation of *is* statements occurs when *is* links two nouns each with articles *a* or *an*. In this section, we explore further ambiguities that arise in this construction, particularly as they pertain to quantification and generalization, including a few more examples from the pedagogical corpus for discussion:

- Example 8 (Ped): "If inlinemath *is* a complete binary tree of height inlinemath, then…"
- Example 9 (Ped): "If inlinemath *is* a family of sets which covers inlinemath and inlinemath *is* a subfamily of inlinemath which also…"
- Example 10 (Ped): "The Cartesian product of two sets inlinemath and inlinemath, written inlinemath, *is* the set of all ordered pairs inlinemath, where inlinemath and inlinemath."
- Example 11 (Ped): "It can be shown that the best strategy is to pass over the first inlinemath candidates where inlinemath *is* the smallest integer for which inlinemath."
- Example 12 (Ped): "If inlinemath *is* a type 1 integer and inlinemath is a type 2 integer, then inlinemath is a type 2 integer."
- Example 13 (Ped): "If inlinemath, we say that inlinemath *is* a compact subset of inlinemath if, regarded as a subspace of inlinemath, it is a compact metric space."

As noted above, these *is* statements generally convey either a symmetric relation ("same as") or an asymmetric relation ("one of"). In most all cases the nouns on either side of *is* are singular with singular articles (*the, a, an*). However, given the value placed upon generalization in mathematics, these singulars are understood to represent entire classes through arbitrary selection (Durand-Guerrier, 2008). It is this implicit generalization that introduces so much of the ambiguity into statements of this grammatical form.

For instance, Example 1 seems to identify two singular objects. The subject of the sentence is the same as "the standard basis" for some other object. However, if this sentence is introducing a general notation for standard bases, it means to convey a universal relationship. Without recognizing whether *inlinemath* in that sentence represents a generic placeholder or some representation of a singular mathematical object (or a placeholder for some more specialized class), one cannot discern what relation *is* conveys. Example 2 conveys a general law of integrals, not merely a naming convention (despite being structurally similar to the definition in Example 10). However, one cannot tell from the grammatical form of the statement whether this sentence is stating the law in general or applying it to a particular case (Example 12 is similar in this regard). In Example 2, the article *the* is misleading. *The* marks the singularity of indefinite integrals, but the function being integrated should likely be understood as a placeholder representing any function. In other words there is one indefinite integral per function, but the statement almost certainly applies to a range of functions. Example 3 quite clearly means to convey a universal (defining) relationship, despite the singular article on both sides of *is*. The key point is that one cannot discern this merely grammatically – familiarity with the mathematical concepts is essential. In contrast, the grammatical cues in Example 4 convey more accurately that *is* relates a particular object ("this map") to a general class ("an isomorphism").

Our examples reveal other common grammatical cues that mathematicians use to convey the implicit generality behind nouns and noun phrases with singular articles. For instance, the *if*s at the beginning of Examples 8 and 9 are there to convey universal quantification of the subject of the *is* claim[2]. Example 13 presents an odd case where *if* is used in two slightly different ways in the same definition. The first *if* calls out an arbitrary metric space (a context assumption) while the second presents the defining condition for being a compact subset. In cases where *is* could relate an entire class represented by an arbitrary placeholder or a particular case, deciding whether the variable or name given to an object has appeared before or not (i.e. is already *bound*, Epp, 2009) provides a subtle cue. For instance, this would resolve some ambiguity in Examples 2 and 12. If the variable is not bound then the claim is likely universal; otherwise it may be an application of a warrant to a particular case or an introduction of cases within an argument. The mere grammar of the construction "If [subject] is a [noun]" does not distinguish between these uses. Furthermore, Example 11 shows how mathematicians sometimes compress the process of binding and using a variable by referring to a quantity before defining it in an appended clause.

What we gather from these examples is that the "[subject] is [noun]" grammatical structure entails semantic ambiguity that is only partially resolved by other grammatical cues (articles and conjunctions). In other words, one cannot infer the relationship between the subject and object nouns merely by the statement's construction. Mathematicians tend to state the general using arbitrary particulars, usually using placeholder variables or names with singular articles. This construction is not unique to mathematics (e.g. "The redeemed soul is a debtor to mercy alone"),

[2] Indeed one of our philosopher colleagues argues that such claims are not really conditional at all, but rather universal (L. Clapp, personal communication, December, 2016; c.f. Durand-Guerrier, 1996).

but it appears from our samples to be much more common. This means students will likely need to be trained to properly interpret such common constructions in the mathematical register.

Reflections

The goal of our grammatical analysis was to 1) identify differences between *is* usage in day-to-day and mathematical language and 2) to identify the semantically ambiguous *is* constructions in mathematical language. Due to space limitations, we have only presented our analysis of "[subject] is [noun]" constructions.

We proffer two tentative points from preceding analysis regarding the nature of the issue and what can be done to address it. First, we do not mean to belittle or demonize semantic ambiguity in mathematical discourse. We view it as inevitable, despite mathematicians' pursuit of precision and explicitness. However, we observe there is a tradeoff between simple statements that entail semantic ambiguity and complex statements that are grammatically hard to parse (c.f. Schleppegrell, 2004). Pedagogically speaking, we must create ways for students to be apprenticed into mathematical knowledge and language, requiring that we make it easier to parse and interpret. Simplifying language often incurs a cost in precision. In many cases, we judge that this price must be paid. However, problems arise when mathematics instructors treat dense constructions like "A square is a rectangle" or "A rational number is a fraction built out of integers" as completely unambiguous, without recognizing the role their expertise plays in rendering these claims interpretable.

Second, we comment on what might be done to maintain efficiency in pedagogical language while increasing the fidelity of communication. We recognize that empirical study must ultimately determine this, but we offer two ideas for consideration. One, it may help to alternate the grammatical cues we use to convey similar relationships. For instance, one could state and restate one of our first examples – "A square is a rectangle" – in multiple ways:

- "Each square has all the properties of a rectangle."
- "All squares are also rectangles."
- "Each square is also in the class of rectangles."

Similarly, statements conveying symmetric relations – "A square is a regular quadrilateral." – can be restated:

- "A regular quadrilateral is known as a square."
- "A square is the only kind of regular quadrilateral."

Alternating *a* and *an* with *any* and *each* or clearly designating defining actions with phrases such as *is called* and *is known as* can help cue students to the relations that *is* statements convey. We do not think any one of these is uniquely best. A longer statement is more explicit while "A square is a rectangle" is easy to recall. We recommend that instructors practice parallel articulations conveying the same relations to scaffold mathematical parlance. Some of our other work in mathematical logic demonstrates the importance of students associating mathematical properties with the sets of objects exhibiting the properties (what Dawkins, 2017, calls *reasoning with predicates*). Helping students to manage some ambiguities tied to implicit quantification aligns closely with developing a set-oriented way of thinking about mathematical claims. Two, we perceive that interpreting these mathematical statements is directly tied to understanding mathematical practices such as defining, representing, equating, stating general claims, and applying general warrants to particular cases. Future research on linguistic interpretation may benefit from integrating analysis of students' emergent interpretations of mathematical practices.

References

Alcock, L., Inglis, M., Lew, K., Mejia-Ramos, J. P., Rago, P., & Sangwin, C. (2017). Comparing expert and learner mathematical language: A corpus linguistics approach. *Proceedings of the 21st Annual Conference on Research on Undergraduate Mathematics Education*. San Diego, CA.

Anthony, L. (2015). *TagAnt (Version 1.2.0) [Computer Software].* Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Davies, M. (2017). *Word frequency data from the Corpus of Contemporary American English (COCA) and British National Corpus (BNC).* Downloaded from http://www.wordfrequency.info on 7th April 2017.

Dawkins, P.C. (2017). On the Importance of Set-based Meanings for Categories and Connectives in Mathematical Logic. *International Journal for Research in Undergraduate Mathematics Education,* 1-27. DOI 10.1007/s40753-017-0055-4.

Durand-Guerrier, V. (1996). Conditionals, necessity, and contingence in mathematics class. *DIMACS Symposium Teaching Logic and Reasoning in and Illogical World*. Rutgers, The State University of New Jersey, July 25-26.

Durand-Guerrier, V. (2008). Truth versus validity in mathematical proof. *ZDM, 40*, 373–384.

Geist, L. (2008). Predication and equation in copular sentences: Russian vs. English. In I. Comorovski and K. von Heusinger's (Ed.), *Existence: Semantics and Syntax*, Volume 84 of the *Studies in Linguistic Philosophy* series (pp. 79-105), Dordrecht, Netherlands: Springer.

Johansson, S., Leech, G. N., & Goodluck, H. (1978). *Manual of Information to Accompany the Lancaster-Oslo/Bergen Corpus of British English, for Use with Digital Computers*. Department of English, University of Oslo.

Kucera, H. & Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Providence, Rhode Island: Brown University Press.

Morgan, C. (1996). "The language of mathematics": towards a critical analysis of mathematics texts. *For the Learning of Mathematics, 16*(3), 2-10.

Moschkovich, J. N. (1999). Supporting the participation of English language learners in mathematical discussions. *For the Learning of Mathematics, 19*(1), 11-19.

Russell, B. (1919). Introduction to mathematical philosophy. London: Allen and Unwin.

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah, NJ; Lawrence Erlbaum Associates.

Schleppegrell, M. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading and Writing Quarterly, 23*, 139-159.