Journal of Mathematical Behavior 30 (2011) 19–29

Contents lists available at ScienceDirect

The Journal of Mathematical Behavior

journal homepage: www.elsevier.com/locate/jmathb

Semantic contamination and mathematical proof: Can a non-proof prove?

Juan Pablo Mejía-Ramos [a], Matthew Inglis [b,*]

[a] Graduate School of Education, Rutgers University, United States
[b] Mathematics Education Centre, Loughborough University, Schofield Building, Loughborough, United Kingdom

ARTICLE INFO

Available online 26 January 2011

Keywords:
Language
Mathematics
Proof
Reasoning
Semantic contamination

ABSTRACT

The way words are used in natural language can influence how the same words are understood by students in formal educational contexts. Here we argue that this so-called semantic contamination effect plays a role in determining how students engage with mathematical proof, a fundamental aspect of learning mathematics. Analyses of responses to argument evaluation tasks suggest that students may hold two different and contradictory conceptions of proof: one related to conviction, and one to validity. We demonstrate that these two conceptions can be preferentially elicited by making apparently irrelevant linguistic changes to task instructions. After analyzing the occurrence of "proof" and "prove" in natural language, we report two experiments that suggest that the noun form privileges evaluations related to validity, and that the verb form privileges evaluations related to conviction. In short, we show that (what is judged to be) a non-proof can sometimes (be judged to) prove.

Argumentation and proof are widely accepted as being central to mathematics (Heinze & Reiss, 2007; Hilbert, Renkl, Kessler, & Reiss, 2008). Not surprisingly then, educators generally agree that argumentation and proof should be incorporated into mathematics learning and instruction at all levels (Hanna, 2007; Schoenfeld, 1994). However, many studies have shown that students find engaging with proof difficult, regardless of whether such engagement takes the form of evaluating given proofs or constructing novel proofs (Coe & Ruthven, 1994; Fischbein, 1982; Harel & Sowder, 1998; Knuth, 2002; Selden & Selden, 2003). There is a long tradition in the literature of trying to account for these difficulties by looking at individual students' conceptions of proof, and finding mismatches between these and the agreed standards in the discipline (Bell, 1976; Healy & Hoyles, 2000; Segal, 1999). Our goal in this paper is to build on this long research tradition, but also to attempt to connect it to a rather different and under-researched area: the linguistics of mathematical proof. We first discuss some theoretical background issues associated with both these areas.

1. Semantic contamination

The influence of language on learning has been widely recognized in the education literature. It is now accepted that the way linguistic structures are used in natural language (by which we mean language from day-to-day life) can differ from how the same linguistic structures are used in formal contexts and, in particular, in mathematics. When analyzing this phenomena, Halliday (1975) used the term "register" to refer to "a set of meanings that is appropriate to a particular function of language, together with the words and structures which express those meanings" (p. 65). Halliday suggested

* Corresponding author.
E-mail address: m.j.inglis@lboro.ac.uk (M. Inglis).

0732-3123/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmathb.2010.11.005

Table 1
Frequency of noun and verb forms of proof in the spoken component of the BNC, by the language-type of the source.

	Noun	Verb
Specialist language	121 (54%)	105 (46%)
Informal language	51 (35%)	95 (65%)

that the mathematical register is relatively unusual for a technical register, in that when naming new phenomena, it often redefines simple words from natural language rather than coining novel technical terminology. This can give rise to what Pimm (1987) called *semantic contamination*: where the meaning or usage of a term from natural language influences how the term is understood by a learner in the mathematical register.

Several examples of semantic contamination have been discussed in the mathematics education literature. In the context of advanced mathematics for example, Monaghan (1991) found that the natural language meaning of words and phrases associated with the limit concept ("tends to," "approaches," "converges," etc.) can impact upon students' concept images of the formal limit concept in calculus and analysis classes. Similarly, Tall and Vinner (1981) suggested that colloquial meanings of the term "continuity" influence how students engage with the formal mathematical concept. Difficulties which arise from such issues will need to be overcome if the learner is to successfully engage with the mathematical register (Pimm, 1987; Schleppegrell, 2007).

In this paper we explore whether students' understanding of mathematical proof is influenced by semantic contamination. In other words, we ask does the way that words associated with proof are used in natural language influence how they are understood in the mathematical register?

To establish semantic contamination three steps are required. The first is to investigate the way in which the to-be-analyzed concept is referred to in natural language. In the case of limit discussed by Monaghan (1991), for example, the natural language use and meanings of the terms "limit," "approaches," "tends to" and so on were investigated. The second step is to empirically examine how these concepts are understood by students in mathematical contexts: Monaghan conducted a survey which interrogated students' interpretations of these terms. The third and final step is to argue that the results from the analyses conducted in the first two stages correspond sufficiently to suggest that natural language referents to the concept influences students' understanding of the concept in mathematical contexts.

We begin this three step process by analyzing the contexts in which the concept of proof appears in natural language.

2. Proof in natural language

In the English language there are two main linguistic methods of referring to the concept of proof: it can appear as a noun ("proof") or as a verb ("prove"). To investigate the way that these noun and verb forms are used in natural language, we searched for instances of the noun form ("proof" or "proofs") and the verb form ("prove" and "proves") in the British National Corpus (BNC) World Edition (Burnard, 2000). The BNC is a comprehensive collection of 100 million words of spoken and written English, designed to represent a cross-section of current English usage. Analyses of large scale corpora, such as the BNC, are widely used by researchers interested in the usage patterns of various linguistic features. In particular, such techniques have been highly productive at comparing the use of words in different registers (Stubbs, 2004).

Our aim was to determine the frequency of the verb and noun referents to proof in specialist language (i.e. language normally associated with a specific formal context or topic, such as education, business, legal, and medical) and informal day-to-day language (i.e. language which could be found spoken on popular radio or in informal conversations, and so on). In Halliday's (1975) sense, we aimed to compare the occurrences of the verb and noun referents to proof in typical formal and informal registers.

The spoken component of the BNC consists of approximately 10 million words split into two sections: impromptu conversational spoken English recorded from the day-to-day life of 124 representative volunteers, and spoken English recorded from timetabled events in various different contexts: educational (e.g. university lectures), business (e.g. trade union talks or business meetings), public/institutional (e.g. parliamentary proceedings) and leisure (e.g. sports commentaries). We formed a specialist language category by grouping the educational, business and public/institutional context domains, and an informal language category by grouping the conversational component and the leisure context domain.

Table 1 shows the frequency of the noun and verb referents in the spoken component of the BNC, separated by the language-type of the source document (specialist or informal, defined as described above).[1] The noun form was found significantly more often in specialist language (54% of occurrences of proof in specialist language were of the noun form) compared to informal language (where 35% of occurrences were of the noun form), $\chi^2(1) = 12.355$, $p < .001$.[2] Importantly, this disproportionate occurrence of the noun form of proof in specialist language does not merely reflect an overall overabundance

[1] This word frequency analysis included several unusual meanings of "proof" and "prove" (for example "proof" can refer to the aeration of dough by a raising agent before baking). Because semantic contamination refers to how the ways in which a word is used in natural language influences how it is understood in the mathematical register, we did not attempt to remove any such atypical uses (which could have an impact upon the nature of any semantic contamination into the mathematical register) from the sample.

[2] An equivalent analysis on the written component of the BNC revealed a similar pattern of results.