Relating Vocabulary in Mathematical Tasks to Aspects of Reading and Solving

Ewa Bergqvist[1], Anneli Dyrvold[1], and Magnus Österholm[1,2]
[1]Umeå University, Sweden; [2]Monash University, Australia

This paper focuses on relationships between vocabulary in mathematical tasks and aspects of reading and solving these tasks. The paper contains a framework that highlights a number of different aspects of word difficulty as well as many issues to consider when planning and implementing empirical studies concerning vocabulary in tasks, where the aspect of common/uncommon words is one important part. The paper also presents an empirical method where corpora are used to investigate issues of vocabulary in mathematical tasks. The results from the empirical study show that there are connections between different types of vocabulary and task difficulty, but that they seem to be mainly an effect of the total number of words in a task.

Introduction

When using written tests to assess students' mathematical ability, one aspect of validity is to measure mathematical competence and nothing else. However, there is always a possibility that the test also measures the students' reading ability. Language can be seen as a necessary tool for doing mathematics and as such it is not supposed to add unnecessary difficulty to mathematical tasks. Some researchers also describe language as separated from mathematics (e.g. Greenlees, 2010). Still, language is an essential component of doing mathematics, since there are words, symbols, phrases, and grammatical structures that mathematics cannot exist without. Therefore, it might be more reasonable to include a language component in the concept of mathematical knowledge. This is also supported by the aspect of communication presented as a part of mathematical proficiency in several frameworks (e.g. NCTM, 2000). The complex relationship between language and mathematics described here is the starting point for several studies that we are currently carrying out or planning (see also Österholm & Bergqvist, in press-a, in press-b). The overarching goal of our research is to better understand and be able to describe the connection between reading and solving mathematical tasks. In this particular paper we focus specifically on connections between vocabulary in mathematical tasks and aspects of reading and solving. In particular, we discuss and examine different types of words, regarding the interplay between everyday words and mathematical words.

Background

Many studies highlight the importance of vocabulary, both in relation to aspects of reading comprehension in general and in relation to the solving of mathematical tasks, and several different methods are used to study aspects of vocabulary. *Counting letters and/or syllables* to measure word length is one method. It has long been known that using syllables to measure sentence and/or word length is an important linguistic measure of readability, and the number of letters as a measure of word length is also part of some readability formulas (DuBay, 2004). When *judging word difficulty and/or familiarity,* experts, often in the form of teachers or assessment experts, are sometimes used to judge the words (e.g. Shaftel, Belton-Kocher, Glasnapp, & Poggio, 2006). A third type of method is using different *lists of words* to compare with. Helwig, Rozek-Tedesco, Tindal, and Heath (1999) define *word familiarity* using a list of words that contains information about grade level and percentage of students that correctly could identify the meaning of the word. Another version is to use a list of words that children in some particular age understand, or are supposed to understand (e.g. Dempster & Reddy, 2007; Stahl, 2003).

An alternative to using word lists is to *calculate frequencies in particular corpora* in order to characterize words as being familiar or not. This method is based on the assumption that more common words are also more familiar. Empirical results support this assumption (e.g. see Breland, 1996). In modern linguistics, a corpus can be defined as "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" (Sinclair, 2005, p. 23). Different corpora are composed to represent different types of language and therefore they consist of words from different sources of text, for example newspapers, scientific articles, or oral communication. Something that makes corpora useful in research is the annotations that are usually added to the text. Such meta-information can make it possible to search not only for particular words, but for example also for grammatical forms, types of clauses or even particular meanings of words.

Many studies, including several of the ones mentioned above, use statistical methods to investigate the relation between different aspects of the language in the tasks and the students' performance on the tasks, but these methods have serious limitations. The information gained from statistical computations usually concerns to what extent different linguistic aspects of a task (e.g. the amount of long words) correlate with the difficulty of the task, but it does not inform us on *why*. For example, longer words can be more difficult to decode phonologically, an issue that relates specifically to aspects of reading ability, but perhaps mathematically complex concepts are usually represented by longer words, an issue that relates specifically to aspects of mathematical ability. Therefore, if we only rely on statistical correlations between variables describing a task and students' performance on the task, we cannot conclude that a language aspect is simply related to reading ability. Other methods have been utilized in order to overcome this issue, often by using data of students' results not only on a mathematics test but also on a reading test of some kind. However, when we examine these methods, which use correlations and regressions in different ways, all have problems with aspects of validity or reliability (Österholm & Bergqvist, in press-a). Based on our previous methodological analysis, we suggest an approach using principal component analysis to measure a task's demand of reading ability. This method is described in detail in a previous publication (Österholm & Bergqvist, in press-a) and is summarised in the section *Empirical study* below.

Purpose

This paper presents the first part of a study where the purpose is to increase the understanding of the connection between different aspects of the vocabulary of a mathematical task and other aspects of the task; difficulty and demand of reading ability. First, we propose a framework for the study of vocabulary in mathematical tasks, which is intended to function as a basis for planning and implementing empirical studies. The framework gathers different perspectives on vocabulary from prior research and also includes discussions of empirical methodology, issues we see as missing in prior research. Second, we present a small pilot study where we introduce an empirical method where corpora are used to investigate issues of vocabulary in mathematical tasks. The empirical analysis in this study is a first step towards exploring benefits of combining information from different corpora, in order to analyse words that are common in some discourse/situation but uncommon in another, and not only use information about whether words are "universally" common or uncommon, as is usually done in previous research. Our research questions are:

- What crucial factors are there to consider when issues of vocabulary in mathematical tasks are to be studied?
- Is there a connection between the difficulty or the demand of reading ability for a mathematical task and whether the words in the task are common or uncommon in mathematical language and/or everyday language?

A framework for the study of difficult vocabulary

What can be considered a "difficult" word? Here we describe a first version of a conceptual framework regarding the notion of *difficult vocabulary*. In this framework we include what can be seen as different aspects of difficulty regarding the words in mathematical tasks, together with perspectives on how to analyse these aspects in empirical studies. However, for the present paper we focus our attention on the one aspect of common/uncommon words, in particular of words being (un)common in mathematical language and/or everyday language.

The framework has been created based on issues highlighted in previous research (in particular, see the background in this paper) together with our suggestions of alternatives to the perspectives described in existing research literature. Such alternatives have been noticed as relevant and important while planning the empirical study described later in the present paper. Described here is a first version of the framework and we plan to develop it further in future publications. Note that some parts of the framework are here described more briefly, due to space restrictions.

Analysing the difficulty of a word

Here we focus on the analysis of singular words, possibly in relation to the text in which it is included. We include in our framework the following five aspects of word difficulty (see also the background), of which we only elaborate on the fourth aspect in the present paper: (1) word length, (2) word form (e.g. verbs in a passive voice or nominalised verbs/adjectives), (3) word type (e.g. pronouns or modal verbs), (4) common/uncommon words (word familiarity), and (5) word meaning (e.g. the complexity of a concept or a word's potential ambiguity).

We here include four issues to take into consideration regarding the notion of common/uncommon words, and thereafter we discuss the process of analysing these issues in empirical research.

1. When, where, how and who? When a word is labelled as common or uncommon, this needs to be in relation to a certain population or discourse community. For example, oral everyday language can be seen as specifying the language used in a certain type of situation (where) and in a certain form/modality (how). This issue also includes considerations of whose vocabulary is referred to (e.g. regarding age or ethnicity) and the question of when, since language changes continuously.

2. Discourse-specific vocabulary. This issue refers to a relationship between different discourses (or populations), regarding information about whether a word is specific to a certain discourse. For example, in her research review, Schleppegrell (2007) discusses the technical vocabulary of mathematics as an important issue.

3. Derivations and lexemes. This issue highlights the question of whether to focus on a specific word as being common/uncommon or to focus on its components or *lexeme*. Lexeme refers to the set of different forms a specific word can have. For example, "stand", "stood", and "standing" are elements of the same lexeme, since they are different forms of the same word. An argument for focusing on lexemes instead of words is that even if a specific word is relatively uncommon, a reader can perhaps directly see the word as a form of a more common word from the same lexeme. Thus, the word is not as difficult as could be believed from how uncommon it is (e.g. see Dempster & Reddy, 2007). In a similar manner, derivations of words, using for example prefixes and suffixes, might also be misrepresented as highly difficult when relying on how (un)common they are (Stahl, 2003).

4. The context. Since the same word can have different meanings in different contexts, it could be necessary to distinguish between different meanings of words when analysing how common/uncommon they are.

With these four general issues as a basis, we now turn our attention to more practical issues, regarding the planning and implementation of empirical studies about common/uncommon words in mathematical tasks. We focus on the use of some type of explicit reference material in the analysis, in particular word lists and corpora. We describe three steps in the analysis of how common/uncommon a word is using a reference material.

A. Choosing/creating reference material. There are a few questions that can be asked regarding any type of reference material. First, we have the question of *what material to include*, which refers primarily to the first general issue; for the material to be representative of some specified population, situation, and time. For example, to have a corpus for "school language" we need texts from all subjects. Second, we have the question of *what type of meta-information to include*, if any. For corpora, different types of linguistic meta-information are usually included (Leech, 2005). Word lists usually include only the most common words, possibly together with some type of meta-information, for example about frequency of words or about the fraction of students at different school levels who know the meaning of each word (Helwig et al., 1999).

B. Searching for words in reference material. When searching for a word in a reference material, you need to decide what to count as a word and what to count as the same word. Primarily this decision can be about issues number 3 and 4 above. For example, when a corpus includes lexical meta-information, it is possible to search for lexemes and not only specific words. Also, you might need to take into consideration some issues at a more practical level, such as hyphens and spaced words (e.g. if "lifestyle", "life-style", and "life style" are seen as the same word).

C. Characterizing words. The frequencies of words can be used in different ways to measure how common they are. In corpora, a relative frequency can be used as a direct measure, comparable between corpora, of how common a word is. It is also possible to use frequencies, absolute or relative, as a basis for ranking words, and the ranking would then be possible to compare between corpora. The words can also be labelled as common or not (or uncommon or not) in different ways, not directly based on relative frequencies, but on whether a word is included in a given list of (the most) common words (e.g. Dempster & Reddy, 2007) or similarly, whether a word is included among the 1000 (or any chosen number) most common words in a certain corpus (e.g. Österholm & Bergqvist, in press-b).

Analysing mathematical tasks regarding difficult vocabulary

After analysing each word in a mathematical task, the next question is how to use the information about each word to characterize the task. Here we discuss two issues that were necessary to handle when planning the empirical part of this paper, and that need to be handled when planning this type of empirical analysis in general.

Characterizing the "amount of difficulty" in a task. This issue can be measured in several different ways, for example by focusing on: the mere existence of difficult words; the number of difficult words; the proportion of difficult words; the mean of some quantified word property (e.g. to calculate the mean of the frequency of all words); and the spread of some quantified word property (e.g. to calculate the standard deviation of the frequency of all words). The question of what method to use can be seen as an empirical question, regarding which method most truly captures a potential difficulty in a task or set of tasks. For example, a comparison between the effect of the mere *existence* of difficult words and of the *proportion of* difficult words could answer the question of whether it is, in general, of critical importance for students to know *every* word in mathematical tasks.

Different parts of the task text. Included here are decisions about whether and how to include certain parts of the task text in the analysis, for example regarding: sub tasks and the leading text (i.e. the part of the task text that is common for all sub tasks); background information and the prompt (i.e. the question or description of what to do); tables, diagrams, figures, and symbols; and the repeating of a word. The questions concerning this issue can be seen as empirical questions. For example, the creation of separate difficulty variables for different parts of task texts could make it possible to examine whether a potential difficulty in the leading text of the actual question is of most importance for a certain set of tasks.

Empirical study

This empirical study is a small pilot study with the purpose to examine whether and how the presence of common or uncommon mathematics or everyday words is connected to the demand of reading ability and/or to the difficulty of a mathematical task. The study explores how corpora could be used in an analysis of the vocabulary in mathematical tasks. Details of the analysis and choices made should therefore be seen as preliminary and the results as tentative. The method will be developed and the choices more thoroughly argued for in coming publications.

Method in the empirical study

Our analysis consists of three steps. First, values for the variables *demand of reading ability* and *difficulty* are calculated for each mathematical task included in the PISA tests 2003 and 2006. Second, two different corpora are used to determine how common the words in the mathematical tasks are in two contexts (mathematics and everyday contexts). Third, the correlations between the information from the first and second step are analysed, and based on the results we discuss what the presence of words common/uncommon in different contexts means for the difficulty or demand of reading ability of the tasks. Each step is presented in more detail below.

In order to measure a mathematical task's *demand of reading ability*, a principal component analysis (PCA) is used. This method is presented and discussed in more detail in previous papers (Österholm & Bergqvist, in press-a, in press-b) and only briefly described here. In this study, all Swedish students' scores on all PISA mathematics and reading tasks from 2003 and 2006 are entered into the PCA, from which the first two components are extracted, which are expected to correspond to the two abilities of mathematics and reading. The loading value on the reading component for each mathematical task is taken as a measure of the demand of reading ability.

As a measure of *task difficulty* we use the percentage of credited responses for the task (the p-value), which means that if the sum of all credits that the students get are 75 % of all possible credits, the p-value for that task is 0.75.

In order to determine how *common* or *uncommon* particular words in the mathematical tasks are in different contexts, we use two different corpora, at this point chosen partly based on easy access. The corpus we use to represent everyday language (of society in general) is composed of 58 novels (about 4.7 million words) and newspapers from the years when the PISA tests were distributed (i.e. 2003 and 2006; about 42 million words) [1]. To represent mathematical language, we use a corpus consisting of two mathematics textbooks intended for year 8 students (the same age group as the students that take the PISA tests; about 70,000 words), which are part of the OrdiL project (Lindberg & Johansson Kokkinakis, 2007).

We analyse the words in the PISA mathematical tasks by searching for them in both our corpora, and retrieving the *frequency of each word in each corpus*. The search is made on the specific form of each word, and not on lexemes or derivations of any kind (see the framework), mainly due to non-existing meta-information in the corpora. Some words contain a hyphen, in which case the hyphen is included as part of the word during the search. Due to technical limitations in the search procedure, words with "strange" mixtures of upper and lower case letters (e.g. "woRd") are treated as separate words, but "Word", "word", and "WORD" are treated as the same word. All words in the mathematical tasks, also from tables, diagrams and figures, are included in our analyses, but "words" that consist of or include symbols, numbers, or punctuation marks (except hyphens) are excluded. Words consisting of only one letter are also excluded since they sometimes denote variables.

Based on the information on word frequency in each corpus, we sort the words into *four categories*, by labelling each word as common or uncommon in each of the two corpora. The separation of common and uncommon words is done by, for each corpus, dividing the group of unique words into two groups of equal size. The words with zero frequency are excluded in the creation of the two groups since these words are seen as representing a flooring effect in the data, but they are thereafter included in the group of uncommon words for the continued analysis.

For the mathematical tasks, we define *four different variables*, as the number of words in the task in each of the four categories of words. Sometimes a group of tasks are preceded by a common introductory text (leading text). In those cases, we include the words in the introductory text in each of the tasks, since at this point no analysis of the need of the leading text for the understanding of each task is done.

To determine whether the types of words used in the tasks are related to the demand of reading ability and/or the difficulty, two-tailed non-parametric correlations are used (Spearman R coefficient), with a significance level of 0.05. Besides the four different types of words described above, the total number of words in tasks is also added as a variable in the correlation analysis. This variable is included in order to examine whether any significant correlations to the number of certain types of words in tasks could be an effect of the total number of words in tasks.

Empirical results

Vocabulary property	**Demand of reading ability**	**Difficulty**
Uncommon both	r = -0.071 (p = 0.578)	**r = 0.366 (p = 0.003)**
Uncommon math, common everyday	r = -0.230 (p = 0.070)	**r = 0.451 (p = 0.000)**
Common math, uncommon everyday	r = -0.142 (p = 0.267)	r = 0.059 (p = 0.648)
Common both	**r = -0.275 (p = 0.029)**	**r = 0.497 (p = 0.000)**
Total number of words	r = -0.232 (p = 0.068)	**r = 0.442 (p = 0.000)**

Table 1: Correlations between the number of words of different types in tasks and the tasks' demand of reading ability and difficulty (N=63).

In Table 1 it can be seen that the total number of words in the tasks correlates in a significant way with difficulty and almost significantly with demand of reading ability. These results make it difficult to draw conclusions about any effects of the number of *different types* of words in the tasks on difficulty or demand of reading ability. All correlations to difficulty are positive, while the correlations to demand of reading ability are all negative. That is, tasks with more words tend to be more difficult but also tend to have lower demand of reading ability, and the opposite is true for tasks with fewer words. However, there is also a significant negative correlation between the difficulty of a task and its demand of reading ability ($r = -0.589$, $p = 0.000$), making it even more difficult to interpret the correlations in Table 1. In addition, causal interpretations are of course not possible to make from these analyses.

Conclusions and discussion

Our empirical analyses show that there are clear connections between vocabulary and task difficulty, but these connections seem to be mainly an effect of the total number of words in a task. Connections between vocabulary and demand of reading ability are generally weak, but existing tendencies may also be an effect of the number of words in general, rather than of any specific types of words. Other types of studies/analyses are needed in order to handle the uncertainties in these conclusions.

Our results, showing that the effect of the total number of words might be primary, question the results from other studies where the number of “difficult” words has been included as an important factor without considering the effect of the total number of words. However, Shaftel et al. (2006) use regression analyses to examine both the total number of words and several different aspects of difficult vocabulary, with task difficulty as independent variable. Their results show a non-significant regression coefficient for the total number of words, but significant coefficients for several other vocabulary properties, including the feature they label as “Math vocabulary”. However, in their study they use expert judgments for which the description of “Math vocabulary” is somewhat unclear: “unusual or difficult but specific mathematics vocabulary words” (p. 126). More in-depth analysis is needed in order to explain the discrepancies between their study and our results.

Through the framework described above we have created a coherent description of factors to consider when analysing effects of vocabulary on the reading and solving of mathematical tasks. The framework will be used as a basis for a more structured planning and implementation of future empirical studies, and as a basis for summarizing, analysing and criticizing existing research. One important aspect in the future development of our framework is to include relationships to theories of reading comprehension, in order not to limit the framework to describing only practical aspects of empirical research but also to include an explanatory dimension.

Notes

1. From the Swedish Language Bank (http://spraakbanken.gu.se/korp/): the part with novels is labelled *SUC-romaner* and the parts with newspapers are labelled *GP 2003* and *GP 2006*.

References

Breland, H. M. (1996). Word frequency and word difficulty: A comparison of counts on four corpora. *Psychological Science, 7*, 96-99.

Dempster, E. R., & Reddy, V. (2007). Item readability and science achievement in TIMSS 2003 in South Africa. *Science Education, 91*, 906-925.

DuBay, W. H. (2004). *The principles of readability*. Costa Mesa, CA, USA: Impact Information. Retrieved October 26, 2011, from http://www.impact-information.com/impactinfo/readability02.pdf

Greenlees, J. (2010). The terminology of mathematics assessment. In L. Sparrow, B. Kissane & C. Hurst (Eds.), *Shaping the future of mathematics education: Proceedings of the 33rd annual conference of the Mathematics Education Research Group of Australasia* (pp. 218-224). Fremantle, Australia: MERGA.

Helwig, R., Rozek-Tedesco, M. A., Tindal, G., Heath, B., & Almond, P. J. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students. *The Journal of Educational Research, 93*(2), 113-125.

Lindberg, I., & Johansson Kokkinakis, S. (Eds.). (2007). *OrdiL: En korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år [OrdiL: A corpus-based study of the vocabulary in lower secondary school text books]*. Gothenburg, Sweden: The Institute of Swedish as a Second Language, University of Gothenburg.

NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA, USA: National Council of Teachers of Mathematics.

Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading & Writing Quarterly, 23*(2), 139-159.

Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment, 11*(2), 105-126.

Sinclair, J. (2005). Corpus and text - Basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 5-24). Oxford: Oxbow Books.

Stahl, S. A. (2003). Vocabulary and readability: How knowing word meanings affects comprehension. *Topics in Language Disorder, 23*(3), 241-247.

Österholm, M., & Bergqvist, E. (in press-a). Methodological issues when studying the relationship between reading and solving mathematical tasks. *Nordic Studies in Mathematics Education*.

Österholm, M., & Bergqvist, E. (in press-b). What mathematical task properties can cause an unnecessary demand of reading ability? In *Proceedings of the Sixth Nordic Conference on Mathematics Education, NORMA 11*.