Linguistic features and their function in different mathematical content areas in TIMSS 2011

IDA BERGVALL, JENNY WIKSTEN FOLKERYD AND CAROLINE LIBERG

This study investigates how written language is used as a resource to express meaning in different mathematical content areas; algebra, geometry, statistics and arithmetic, in the Swedish version of TIMSS 2011. Based on previous research we identify linguistic features that fulfill the function of expressing four central meaning dimensions of written academic language in general and in language used in school mathematics in particular; Packing, Precision, Personification and Presentation of information. These four meaning dimensions constitute the foundation for the analysis. The results show differences in how the language is used within the different mathematical content areas in TIMSS 2011. These differences consist primarily of to what extent the language is subject specific and used to express the specific mathematics in each of the four content areas. In this way the notion of a single mathematical language is also challenged.

According to Shanahan and Shanahan (2012), language is a subject-specific tool to create meaning and provides unique opportunities to express diverse goals and approaches to the epistemology of different disciplines. Like most school subjects, mathematics has been described as having its particular way of expressing meaning with its special grammatical features and special ways of using different semiotic resources like images and mathematical symbols (O'Halloran, 2005; Schleppegrell, 2007). A successful teaching should emphasize the specific challenges and specific meanings that language entails in the different subjects rather than stress general academic language (Shanahan & Shanahan, 2012). To implement a teaching that draws attention to the subject specific language and its function, knowledge of the specific subject language is required.

Ida Bergvall, Jenny Wiksten Folkeryd & Caroline Liberg
Uppsala University

Bergvall, I., Wiksten Folkeryd, J. & Liberg, F. (2016). Linguistic features and their function in different mathematical content areas in TIMSS 2011. *Nordic Studies in Mathematics Education*, 21 (2), 45–68.

In a research survey, Österholm and Bergqvist (2013) show that there are many claims regarding language used in mathematical texts. However, there are very few studies presenting empirical support for these claims. Morgan, Craig, Schuette and Wagner (2014) also point out that it is likely that the mathematical language is different at different levels of schooling and in different genres such as in scientific reports compared with school books. The subject of mathematics consists of various content areas that have grown out of different traditions and needs. It is possible that the language used is not just specific to mathematics as a whole, but fulfills different functions in the different content areas of the subject. How language is used in different content areas within mathematics is seldom described. However, there are some studies investigating some aspects of language use in single content areas (see e.g. Barwell, 2005; Shreyar, Zolkower & Pérez, 2010; Wagner, Dicks & Kristmanson, 2015). The present study examines empirically how the language of school mathematics is used in four mathematical content areas; algebra, statistics, geometry and arithmetic. The mathematical texts which are examined are the Swedish versions of the tasks in the *Trends in international mathematics and science study 2011* (TIMSS). The aims of this study are to

- Investigate linguistic features that are characteristic of different mathematical content areas in the Swedish version of TIMSS 2011, and what functions these linguistic features fill.
- Highlight differences and similarities between language use in different mathematical content areas, and discuss them in terms of mathematical registers.

Previous research

Compared to spoken informal language, written academic language in general is characterized by high density of information. This information density is mainly provided by a high frequency of nouns (Halliday & Matthiessen, 2004; Martin, 1991). A noun can fulfill the function of summarizing information already written or expected to be known by the reader (Fang, Schleppegrell & Cox, 2006). A lot of information is then packed in one word. Examples are subject specific words or concepts which require knowledge of the definition delimiting the concept. The word *rhombus* is an example of a noun which assumes that the reader knows what a rhombus is, and what distinguishes it from other objects such as squares.

Nouns can also form incongruent grammatical expressions which describe a process or an attribute that in everyday language normally would be expressed by a verb or an adjective (Martin, 1991; Veel, 1997).

These incongruent grammatical forms make the text more dense with information since the information is packed in one clause and expressed with fewer words compared to if the process was expressed in a more everyday manner by using the congruent verb form (Fang et al., 2006; Veel, 1997). In mathematics, this incongruent usage of nouns can reify an activity into a thing, and the activity becomes a concept (Veel, 1999). Examples of this reification are for example when *multiply* is nominalized to *multiplication* or *transform* to *transformation*. As Veel points out, there is a qualitative difference between being able to multiply numbers and to understand the generalized and more abstract concept of multiplication. Reification has also been highlighted by Sfard (1991) who stresses that the concept of number or the concept of a function can be understood in two different ways; as objects or as processes. If the mathematical entity is understood as a process it is conceived for example as a computational process or as a transformation and on the other hand, if the mathematical entity is conceived as an object, the mathematical entity is conceived as a static structure. These two approaches are complementary to each other and the development of a structural understanding of mathematics is important for the development of a mathematical understanding. When a complex problem has to be solved, the solver often needs to repeatedly change between the two approaches. The ability to understand mathematics both as an object and as a process, and to shift between these approaches, is an important part of doing mathematics and the advance from a process understanding to a structural understanding is often a difficult step and involves an ontological change (Sfard, 1991). In language, the transformation from process to object is often represented by nominalizations (Morgan, 1998; Veel, 1999).

Word length is another linguistic feature that fulfills the function of decreasing or increasing the information density of a text (Mühlenbock, 2013). The average word length tends to be higher in written academic language. Word length counted by characters has, in combination with sentence length, been used as a measure of readability in the Swedish readability index *LIX* (Läsbarhetsindex) (Magnusson & Kokkinakis, 2008). The threshold for what counts as a long word in Swedish has been set at more than six letters. The reason is that 90 percent of all polysyllabic words contain more than six letters. 90 percent of all disyllabic words, on the other hand, contain six or fewer letters (Mühlenbock, 2013). But there are also long disyllabic words with more than six letters. In Swedish it is also common with compound words which often consist of more than six letters. Compound words express the same information that would require two words in many other languages. The use of compound words means that a complex concept can be expressed in a compact form and these words are considered to be more difficult than shorter disyllabic words (Björnsson, 1986). Long words are often, but not always, subject specific. These words frequently consist of several morphemes, that is, the smallest grammatical units in a language (Fang, 2006). Examples or words consisting of several morphemes are *subtrac-tion*, *centi-meter*, *tri-angle* and *circum-ference*. Subject specific words put large demands on students' knowledge of the concept and its definition. These words provide a precise meaning and are not replaceable with everyday expressions without some part of the generality and precision being lost (Fang).

Another linguistic feature common in written academic language is extended noun phrases (Fang et al., 2006; Mühlenbock, 2013). In a noun phrase adjectives, adverbs and certain participles are modifiers to the noun. The modifiers fulfill the function of making the main word in the phrase more specified and precise. In mathematics, these modifiers are often classifiers or qualifiers to a thing (Veel, 1999). Classifiers express taxonomic relations, for example that a triangle is of the subtype isosceles or making precise which type of numbers the expression concerns. Examples of classifiers from TIMSS 2011 are *whole* in *whole numbers* or *equilateral* in *equilateral triangle*. Qualifiers specify and limit the meaning of the thing in the noun group often by providing numerical precision as in *a number between six and nine* (Veel, 1999).

Furthermore, mathematics is a subject that has many practices outside the discipline. Mathematics can be used in other school subjects or in everyday situations. In mathematical tasks, concrete and real situations are often described, and these situations have to be solved mathematically. Personal references in mathematical texts are features that fulfill the function of making a text more concrete, and are used to describe and create real situations (Palm, 2002). There are various definitions of what constitutes a realistic task, ranging from a requirement that students need to *image themselves* in the task to the definition that a realistic task includes persons or non-mathematical objects (Palm, 2002). Generally, personal references to present or known participants expressed by pronouns and proper names are more common in a colloquial and spoken communication compared to in written academic language (Schleppegrell, 2001). For personal pronouns and proper names the reference is most often unique and refers to a particular person, place or institution (Halliday & Matthiessen, 2004). In written academic language, on the other hand, more generic descriptions expressing generalizations to groups are used (Edling, 2006; Schleppegrell, 2001). In school texts there is often a progression from the more concrete and specific to the general in more advanced texts for the later school years (Edling, 2006).

Concreteness and an emphasis on human experience can be expressed by persons acting in the tasks (Herbel-Eisenmann, 2007). These persons can be named by proper names, professional title or group membership. In the study of Herbel-Eisenmann (2007) first-person pronouns, indicating the author's personal involvement, were absent in the texts. Second-person pronouns address the reader directly, but also second-person pronouns were rare. More often, the mathematics itself was described as acting on its own, independent of human involvement. The lack of such features obscured the humanist subtext by imaging mathematics as a system acting independently of humans.

Another way to look at pronouns and proper names is as expressions of human interest, which has been done in readability studies. Such features are considered to provide a personal stance in the text and could thereby make the text more vivid and interesting for the reader by providing a description of an everyday situation (Mühlenbock, 2013).

Passive voice is a linguistic feature, which in contrast to pronouns and proper names, fulfills the function of making the text more abstract, agentless and distanced. Passive voice expresses general relationships that are not dependent on a connection to a particular situation or individual, and that focuses an objective production of knowledge where people and relationships are less prominent to the reader. Passive voice is frequently used in written academic language found in science and mathematics (Fang, 2006).

A frequent way to organize texts in written academic language is to use subordinate clauses (Schleppegrell, 2004), that is, they are subordinate to a main clause and cannot stand by themselves. There are different kinds of subordinate clauses. For instance subordinate clauses can begin with adverbs such as *when* or a pronoun like *which* (Hellspong & Ledin, 1997) or with conjunctions like *that* or *because* (Fang, 2006; Veel, 1997). Although subordinate clauses fill several functions and vary extensively in terms of complexity, sentences with subordinate clauses are more complex on the whole. The frequency of subordinate clauses is therefore a good indicator of sentence complexity (Mühlenbock, 2013).

As seen from the overview of earlier research, different linguistic features characteristic of written academic language in general and in school mathematics in particular fill important functions such as to make the information in the text denser, to make the information more precise, to upgrade or downgrade personal relationships, and to give a more complex presentation of the information.

Theoretical framework

A theoretical framework that has been used in studying functions in academic language from different angles like the ones presented in this article, is the *Systemic functional linguistics* (SFL) (Christie, 2012; Coffin & Donohue, 2014; Halliday & Matthiessen, 2004). In SFL, the fundamental property of language is considered to be its function to carry meaning. A text and its linguistic features are seen to fulfill different functions in order to express meanings. Three functional components of the semantic or meaning system, that is, three meta-functions of a text, are considered. The first meta-function carries the so called *ideational* meaning through which the content of a text is expressed. The ideational meaning concerns primarily everything from everyday aspects of the content to more technical and specialized aspects. The second meta-function carries *interpersonal* meaning and expresses personal relationships in the text and with the reader/listener. It concerns attitudinal aspects and dichotomies such as subjective/objective and informal/formal. The third meta-function carries a *textual* meaning in order to express how the text is structured, concerning ways of organizing and presenting information by, for example different types of cohesive relations.

Furthermore, SFL theory proposes that language differs from one context of situation to another. A linguistic choice is most often not the only linguistic expression possible, but if the language is changed this also entails a change of the meaning that is expressed. For example, a mathematical formula could be described in natural written language, but this implies that the expressed meaning would be different. In some situations it is most suitable to use an informal and subject specific expression and in another situation an everyday and personified linguistic form is a more appropriate choice. The language in these two situations expresses different meaning and realizes two different social situations.

The constellation of lexical and grammatical features characterizing a certain language used in a certain social situation is termed a register within SFL (Halliday & Hasan, 1989; Martin, 1991). Language is used in different ways for different purposes in different contexts of situations, and therefore the registers are different. This study examines empirically how the language of school mathematics is used in four mathematical content areas; algebra, statistics, geometry and arithmetic and thereby possibly constitutes different registers.

The SFL-framework provides a solid framework to motivate a selection of relevant linguistic features and to investigate what function they fulfill in order to express different types of meaning. Moreover, this framework may also support future comparisons between languages, since these functions are not language specific. It is the linguistic features that may be language specific.

Central meaning dimensions

As seen from the overview of earlier research, different linguistic features that are characteristic of written academic language fulfill the functions of making the information in a text packed and precise, downgrade personal relationships, and have a more complex presentation of the information. Written academic language and more specifically school language in mathematics can thereby be studied through the lenses of the three meta-functions and what meaning they carry: the meta-function carrying ideational meaning (*packing* the information and making it more or less *precise*), the meta-function carrying interpersonal meaning (*personification* of the information) and the meta-function carrying textual meaning (*presenting* the information). In this way Packing, Precision, Personification, and Presentation of information will be considered as central meaning dimensions of school language of mathematics. The contribution of this study compared to previous descriptions of school language of mathematics is that these meaning dimensions highlight the functions different linguistic features fulfill.

Method

The material

The material in this study consists of mathematical tasks from the Swedish version of the large-scale international test TIMSS 2011, grade eight. It contained 217 mathematical tasks (algebra: $n = 70$, statistics: $n = 43$, geometry: $n = 43$, and arithmetic: $n = 61$).

TIMSS evaluates knowledge based on the TIMSS framework which is not identical to the policy documents for education in Sweden or any other country. The TIMSS framework can be described as a hybrid of the participating countries' policy documents (Skolverket, 2012). An analysis made by the *Swedish national agency for education* shows that the questions posed in the TIMSS have relatively large bearing on the Swedish policy documents in both mathematics and science, especially in grade eight (Skolverket, 2012). This large bearing on the Swedish policy documents provides validity of the results also when it comes to other practices than TIMSS. In combination with the possibilities to make automatic linguistic analysis of this large-scale material, the large bearing on the Swedish policy documents has been decisive for the choice of the empirical material.

The analyses

The framework of SFL provides a base for making several different types of analyses of language use. Examples of some such analyses are to investigate the type of nouns (participants) and verbs (processes) used in order to express ideational meaning, the appraisal system used in order to express interpersonal meaning, and theme-rheme-patterns and lexical chains in order to express textual meaning. However, these features require a manual analysis. To allow an analysis of such a large material as the tasks in TIMSS 2011 an automatic computational linguistic tool is needed and has been used in this study as a first step in the analysis. This analysis has then been combined with a manually performed analysis of some key features concerning the interpersonal and textual meaning since the program has not been trained to do this kind of analysis. This applies to subordinate clauses, passive form and the identification of personal pronouns.

The four meaning dimensions Packing, Precision, Personification and Presentation have been used as an operationalization of the meaning the three meta-functions serve to carry, and as a base for the analyses of the TIMSS tasks. Based on the overview of previous research and on the empirical material that has been available, the linguistic features in the Swedish language that have been chosen as relevant to study are presented in table 1.

Table 1. *Meaning dimensions and the linguistic features they are expressed by*

Metafunction that carries	Meaning dimensions	Linguistic features
Ideational meaning	Packing	proportion of nouns and long words *
	Precision	proportion of adjectives, adverbs, participles and numerals
Interpersonal meaning	Personification	proportion of proper names and personal pronouns
Textual meaning	Presentation	proportion of subordinate clauses and passive voice

Note. * Long words = words with more than six letters in Swedish as in the readability formula LIX described by Mühlenbock, (2013) and Björnsson (1968).

To enable an automatic linguistic analysis of this large amount of material, a computer pre-processed database with the TIMSS tasks was used (see Liberg & Forsbom, 2009). In this computer based automatic parsing, each task was classified according to the following linguistic features: the word class that each word belongs to, number of words per task and word length. In order to capture the linguistic features subordinate clauses, personal pronouns and passive voice, a manual analysis was added. All words belonging to a task were included, also words in tables, diagrams and within the multiple-choice alternatives. For analytical reasons all mathematical symbols and digits were excluded from the analysis, even though symbols and digits have grammatical functions in the sentences in which they occur. The aim of this study was to investigate the function of linguistic features in the language. Symbols and digits are analyzed in another, forthcoming study.

In order to compare the meaning dimensions in each content area in mathematics, meaning dimension profiles were calculated in two steps. First a *z*-score was computed for each separate linguistic feature in each task. A *z*-score allows comparisons of variables on different scales, showing how many standard deviations each task differ from the average (Borg & Westerlund, 2006). The formula for calculating the *z*-score is

$$z = \frac{x - \bar{x}}{s_x}$$

where x is the value for the actual task, $\bar{x}$ is the total average for all tasks in all content areas and s_x is the standard deviation for all tasks in all content areas.

As a second step, the *z*-scores for the linguistic features included in one meaning dimension were added in an index. In this way each task got an index value for each meaning dimension. As an example the index for information packing was calculated by adding the *z*-score for nouns and the *z*-score for long words. This means that long nouns are counted both as nouns and as long words, which is considered as relevant because these words contributes both to function as a noun and as a long word. Descriptive data on these four meaning dimensions-indices are described in table 2.

As mentioned above, a more detailed analysis of the ideational and textual meaning was also conducted. In this analysis the question is whether linguistic features that express ideational meaning belong to a subject specific or a colloquial language (cf. af Geijerstam, 2006). Examples of this is expressions such as *konstant* (constant) and *parallell* (parallel) which have been analyzed as subject specific, while the expression *större* (bigger) and *många* (many) have been analyzed as belonging to colloquial language. Personal pronouns, proper names and passive form have not been included in this analysis, since these features are colloquial respectively academic by themselves. The age of the students were also taken into account in this analysis. Things that can be regarded as outside everyday language for the age group was analyzed as subject specific. The

Table 2. *Descriptive data of the meaning dimensions in four content areas*

		Packing	Precision	Personification	Presentation
Arithmetic	SD	1.67	1.97	1.94	1.83
	Mean	-0.36	0.28	0.37	0.00
	Max	4.46	8.15	6.52	6.46
	Min	-3.82	-2.43	-0.81	-1.27
	n	61	61	61	61
Geometry	SD	1.26	2.13	0.76	1.61
	Mean	-0.01	-0.07	-0.56	0.04
	Max	2.20	5.30	2.71	5.50
	Min	-2.33	-2.43	-0.81	-1.27
	n	43	43	43	43
Statistics	SD	1.27	1.98	1.45	1.18
	Mean	0.59	0.33	0.39	0.40
	Max	2.83	4.99	4.17	4.09
	Min	-2.07	-2.43	-0.81	-1.27
	n	43	43	43	43
Algebra	SD	1.87	2.18	1.27	1.53
	Mean	-0.04	-0.41	-0.22	-0.28
	Max	4.41	8.62	5.07	4.74
	Min	-3.82	-2.43	-0.81	-1.27
	n	70	70	70	70

proportion of subject specific versus colloquial features where calculated by dividing the number of occurrences of subject specific versus colloquial features by the total number of occurrences of the feature. The result is presented in table 3 and shows the proportion of subject specific versus colloquial in each linguistic feature within each content area. 1 is the maximum and 0 is minimum.

To give an overall picture of the length of the tasks within the various content areas in TIMSS 2011, the average number of words per task was also calculated for each content area. There is some variation between content areas. Most noteworthy is that tasks in statistics, on average, are nearly twice as long as tasks in the other content areas with 58 words per task on average compared to 29 words per task in algebra, 28 in geometry and 31 in arithmetic.

Results

In this section the results are first presented for each content area concerning the profiles of meaning dimensions in the mathematical tasks in TIMSS 2011 (see Figure 1). Secondly, the result of the more detailed analysis concerning subject specific versus colloquial linguistic features is shown

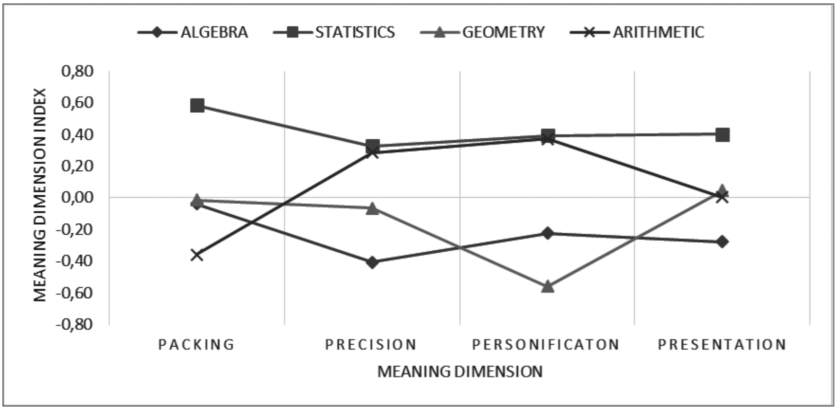


Figure 1. Profiles of the meaning dimensions in each content area.

in table 3. This is followed by a description of each of the four content areas and a more detailed examination of the different linguistic features that fulfill the function of expressing the four meaning dimensions.

Algebra

The profile for algebra shows an average or low level for each of the four meaning dimensions. Information packing in algebra converges with the total average level for this meaning dimension. The nouns and long words

Table 3. *Proportion of subject specific versus colloquial linguistic features within the meaning dimensions in each content area*

			Algebra	Statistics	Geometry	Arithmetic
Packing	Nouns	Colloquial	0.29	0.80	0.11	0.70
		Subject specific	0.71	0.20	0.89	0.29
	Long words	Colloquial	0.40	0.80	0.17	0.74
		Subject specific	0.60	0.20	0.83	0.25
Precision	Participles	Colloquial	0.80	0.74	0.60	0.75
		Subject specific	0.20	0.26	0.40	0.25
	Adverbs	Interrogative	0.16	0.26	0.23	0.22
		Colloquial	0.51	0.51	0.55	0.62
		Subject specific	0.32	0.24	0.23	0.16
	Adjectives	Colloquial	0.83	0.90	0.56	0.90
		Subject specific	0.17	0.10	0.44	0.10
Presentation	Subordinate	Colloquial	0.81	0.96	0.93	0.95
	Clauses	Subject specific	0.19	0.04	0.07	0.05

consist mostly of subject specific words and descriptions of mathematical concepts, for example the noun *värdet* (value) in figure 2.

FRÅGA M042226	Egenkonstruerade svar
$k = 7$ och $l = 10$. Vad är värdet av P om $P = \frac{3kl}{5}$? Svar: ______________	

[English version: k=7 and l=10. What is the value of P when P=3kl/5? Answer: _]

Figure 2. *Example of Algebra task (Skolverket, 2014, p. 49)*

Furthermore, in the algebra content area a low degree of precision is used. One form of precision that occurs is expressed by adjectives in the few tasks where a real world situation is created. One example is the adjective *long* in the following example: *A piece of wood was 40 cm long.* A specific function of adverbs in the algebra area is to specify conditional information as in; *om (if) x = 4* and *om (when) P = 3kl/5* (see figure 2). These adverbs also introduce subordinate clauses describing mathematical conditions. These subordinate clauses thus express mathematical precision and, to a large extent, distinguish algebra from other content areas because this form of subordinate clauses is very rare in the other content areas.

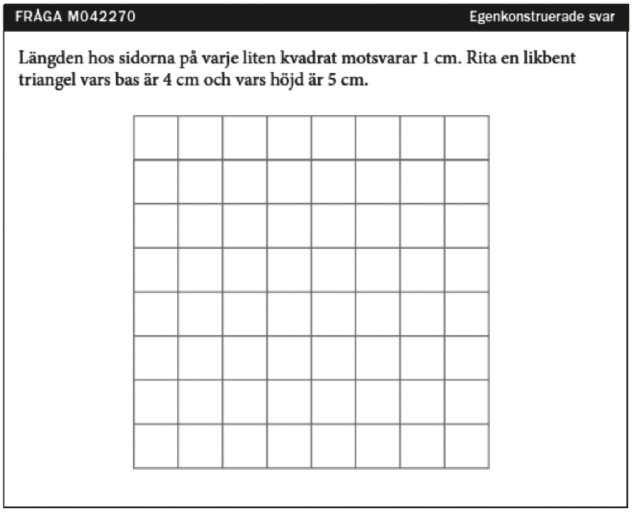
Personification is very seldom used in the language in algebra tasks in TIMSS 2011. Instead of proper names the language contains generalized concepts such as *students* or *boys*.

Geometry

Geometry has a profile which is very average except for low personification. In the meaning dimension of information packing the long words and nouns to a large extent are mathematical subject words such as *triangle*, *circumference* or *rotation*. Everyday words are rare among the long words and nouns in geometry.

Geometry is the content area with the lowest degree of personification in the tasks. The texts are predominantly descriptions of non-representational geometric figures (see figure 3). Tasks containing persons or descriptions of everyday geometric figures are rare in TIMSS 2011.

The frequency of subordinate clauses in geometry tasks is close to the average. Subordinate clauses sometimes fulfill the function of describing mathematical conditions which are also common in algebra. But more



[English version: The length of the sides of each of the small squares represents 1 cm. Draw an isosceles triangle with a base of 4 cm and a height of 5 cm]

Figure 3. *Example of Geometry task (Skolverket, 2014, p. 75)*

commonly the subordinate clauses in geometry serve a function more similar to how subordinate clauses work in narrative texts, that is, adding information which is not always mathematical.

Precision, expressed by adjectives, adverbs and participles, often provides explanations of mathematical concepts. Examples are modifiers to the main word in a noun phrase, as in *parallel sides*, *isosceles triangle* and *right angled triangle*. As Veel (1999) describes, these modifiers fulfill the function of being classifiers to the main word, the mathematical object. In this way, the classifiers express both precision and information packing when they refer to mathematical concepts like *isosceles*.

In TIMSS 2011, geometry is thus a content area with a high level of generality, expressed by subject specific words and mathematical classifiers. Moreover, it has a low degree of personification and everyday aspects expressed by the language.

Statistics

In TIMSS 2011, statistics is, as already stated, the content area with the most words per task, on average almost twice as long as in the other content areas. In contrast to algebra and geometry, statistics exhibits a profile with quite high values in all four meaning dimensions. The high values might partly be explained by the answer alternatives and the figures which in this content area often include written language. In other content areas, it is more common with numbers or formulas that are repeated in the answer alternatives and in the figures.

The meaning dimension profile shows high values of information packing. The tasks contain a high proportion of long words. These long words are sometimes subject specific words, but more often they have an everyday character and may not be as dense of mathematical information as the long words in for example geometry. Examples of long words in statistics in TIMSS are *candies*, *machine* and *together.* These everyday words usually have the function of providing a real world background that forms the basis of the statistical calculations required in the tasks. In statistics there are also long words that do not have an everyday character but can neither be considered as mathematical subject words. Examples of this type of words are *population*, *citizens* and *business*. These words also fill the function of providing a real world background for the statistical calculations but are rather retrieved from other school subjects than from an everyday situation. Although these words are information dense, they are not dense in a mathematical sense.

The texts in the statistics content area also have a high degree of personification which indicates texts with some form of everyday situation. Names of persons, sport teams, cities or companies also contribute to the real world background, which forms a basis for the statistical calculations, as with the long words mentioned above. Personal pronouns are common and are used to refer to the persons present. Tasks in statistics that solely rely on general mathematical relationships, with no personification or any description of a real world situation, are very rare in TIMSS 2011. An example of a task in the statistics area expressing a personified situation is shown figure 4. It contains the proper names *Sara* and *Carl* and also a personal pronoun *hans* (*his*).

When it comes to the meaning dimension of presentation, subordinate clauses are often used in the statistics subject area to express this meaning dimension. These subordinate clauses have a more literary form, connecting clauses to each other. An example is the subordinate clauses in the alternatives in figure 4, for example "... *att hans karamell är rosa*" (... *that his candy will be pink*).

Passive voice is another common feature in statistics with relevance for presentation of the mathematics, which fulfills the function to make the text more complex. An example is the formulation "*480 elever ombads uppge sin favoritsport*" (*480 students were asked to name their favorite sport*).

Within statistics there is also a high degree of precision, mainly expressed by adjectives. These adjectives fill the function of describing

FRÅGA M032132 **Flervalsfråga**

En automat innehåller 100 karameller. När man vrider om ett handtag kommer det ut en karamell. Automaten innehåller en blandning av samma antal blå, rosa, gula och gröna karameller. Sara vred om handtaget och fick en rosa karamell. Därefter vrider Karl om handtaget.

Hur sannolikt är det att Karl får en rosa karamell?

(A) Det är säkert att hans karamell är rosa.

(B) Det är mer sannolikt än det var för Sara.

(C) Det är exakt lika sannolikt som det var för Sara.

(D) Det är mindre sannolikt än det var för Sara.

[English version: A machine has 100 candies and dispenses a candy when a lever is turned. The machine has the same number of blue, pink, yellow, and green candies all mixed together. Megan turned the lever and obtained a pink candy. Peter turned the lever next. How likely is it that Peter will get a pink candy? A. It is certain that his candy will be pink. B. It is more likely than it was for Megan. C. It is exactly as likely as it was for Megan. D. It is less likely than it was for Megan.]

Figure 4. *Example of Statistics task (Skolverket, 2014, p. 107)*

and distinguishing different objects in a real world situation, so that they can form the basis of, for example, a probability calculation. In figure 4, the adjectives *blue*, *pink*, *yellow* and *green* separate the candies and enable probability calculations on the chance to get a candy within a particular color. Veel (1999) states that written academic language in mathematics often contains adjectives describing general mathematical properties, but in statistics, it is found in this study that adjectives providing this type of mathematical classifications are rare.

Arithmetic

Compared to the other content areas in TIMSS 2011, the meaning dimension profile of arithmetic is most similar to the profile of statistics, when it comes to precision and personification. Moreover, in arithmetic the meaning dimension profile shows a very low degree of information packing. The long words and nouns present in these tasks are usually not mathematical subject specific words. And subject words from other areas, which are common in statistics, are also rare in arithmetic. Instead, the long words are more often of an everyday character such as *workman*, *cartons* and , indicating everyday situations rather than general mathematical conditions.

Another indicator of an everyday situation is the high degree of personification in the tasks as shown in figure 5. This is a task with a personified content expressed by proper names and personal pronouns.

FRÅGA M052061 Egenkonstruerade svar

Kim packar ägg i kartonger.

Varje kartong har plats för 6 ägg.

Hon har 94 ägg.

Vilket är det minsta antalet kartonger hon behöver för att packa alla äggen?

Svar: ______________ kartonger

[English version: Kim is packing eggs into boxes. Each box holds 6 eggs. She has 94 eggs. What is the smallest number of boxes she needs to pack all the eggs? Answer: ____boxes]

Figure 5. *Example of personified content (Skolverket, 2014, p. 19)*

There is a large variation between different tasks concerning personification. In this respect arithmetic is a quite heterogeneous content area. In some tasks there is a connection to an everyday situation, even if no names or personal pronouns are used. Instead, a person might be referred to as a *workman* or a *pupil*, or everyday objects such as *caps* or a *school class* might be described. However, it is also common with tasks without any connection to an everyday situation at all. These tasks usually involve pure calculations as shown in figure 6.

FRÅGA M032725 Egenkonstruerade svar

Skriv $3\frac{5}{6}$ i decimalform avrundat till 2 decimaler.

Svar: ______________

[English version: Write 3 in decimal form, rounded to 2 decimal places. Answer:_]

Figure 6. *Example of task without connection to everyday life (Skolverket, 2014, p. 33)*

When it comes to presentation, the subordinate clauses in arithmetic are similar to subordinate clauses common in statistics and serve to organize the text, see for example the last sentence in figure 5 "... *she needs to pack all the eggs*". The subordinate clauses in arithmetic are not primarily used to express mathematical conditions, an opposite observation compared to algebra (see figure 2).

The precision which is mainly expressed by adjectives in the arithmetic content area, is often extensively used and increases the level of concreteness. The adjectives do not, to a large extent, fill the function of providing classifications of mathematical concepts as Veel (1999) describes. In this study adjectives describing mathematical concepts have been found in geometry, for example in the formulation an *isosceles triangle*. The adjectives in arithmetic on the other hand fulfill the function to specify an everyday content.

Discussion

This study investigates linguistic features that are characteristic of mathematical content in the areas of algebra, statistics, geometry, and arithmetic in the Swedish version of TIMSS 2011, and what functions these linguistic features fill. Based on an analysis of meaning dimensions, it has been possible to illustrate how the language in the tasks differs between the various content areas and to reveal important nuances in how language is used to express meaning in these mathematical areas. The differences and similarities identified between language use in the different mathematical content areas will here be discussed in terms of mathematical register. In accordance with the theoretical framework of SFL these contexts of situations can be noted as different registers used in the different content areas.

In the results of this study, the most salient trait of the language used in algebra in TIMSS is an average degree of information packing which is mostly expressed by subject specific nouns and long words. On average there are low degrees of precision and personification and the subordinate clauses are of a subject specific type in the studied tasks. The language use in geometry in TIMSS shows some similarities to the language use in algebra, the difference being that the proportion of subject specific words expressing the information packing and precision are even higher. On the other hand, the degree of personification is very low. Furthermore the content area of statistics in TIMSS is characterized by high average values on all four meaning dimensions with a high proportion of colloquial information packing and precision and a highly personified content. The studied tasks in arithmetic draw on a similar language as used in the content area of statistics when it comes to precision and personification. The degree of information packing is very low and the degree of complexity in presentation is on an average level.

A conclusion that can be drawn is that there are differences between how language is used in various content areas in TIMSS 2011. Due to these differences in language in the four content areas which have been made visible by the results from this study a conclusion is that the four content areas in TIMSS, algebra, statistics, geometry and arithmetic constitute four different registers. As shown in the meaning dimension profile the meaning dimensions are expressed in a similar way in some content areas for example when it comes to precision and personification in statistics and arithmetic. In other meaning dimensions the results show great differences between the content areas, and therefore the content areas can not be regarded as the same register.

These differences in the meaning dimension profiles, and thereby in the registers, are mainly a result of the different ways to do mathematics in the various content areas in TIMSS. For example, algebra does not refer to everyday mathematical problems as much as statistics does. In algebra, on the other hand, a more general perspective on mathematical structures is taken. Thus, the different profiles of algebra and statistics are a product of their different ways of doing mathematics and language and mathematics can not be separated. The language used is an expression of the mathematical content. Furthermore, different ways of doing mathematics is not solely expressed by digits, mathematical symbols or diagrams. The written language in a task is also part of the different ways of doing mathematics.

Generality in various content areas

This study shows that the written language used in the various content areas in TIMSS 2011 expresses various degrees of personification, information packing and precision Altogether, this shows different degrees of generality which varies between the various content areas. In statistics and arithmetic the degree of personification is high, which, according to Edling (2006) indicates a low level of generality and that the text rather concerns specific objects and situations. The linguistic features expressing information packing and precision in these content areas also often have an everyday character. The difference from the content areas of algebra and geometry is apparent. In algebra and geometry the degree of personification is low, linguistic features expressing information packing and precision often describes subject specific concepts independent of time and place, or classifications and relationships, rather than specific situations. Edling (2006) points out that the degree of generality often increases in later school years. To move from the specific to the general is an important, but often difficult, step for a student. A challenge is to understand the meaning of the generalized concept of the subject without a connection to a specific situation. This study shows that both specific and general content is expressed, occurring simultaneously in various mathematical tasks in TIMSS 2011. Students are expected to be capable to handle both specific and general texts and to be able to alternate between them.

Mathematics as an object or a process expressed in language

Sfard (1991) describes two approaches to mathematical concepts. These approaches imply that a notion such as a number or a mathematical function can be conceived structurally, as an object, or operationally, as a process. In language, this corresponds to processes expressed by congruent forms using verb constructs, or by an incongruent expression using a noun, and thus turning the process into an object. The incongruent noun-form is a common feature in written academic language (Halliday & Matthiessen, 2004). This study of tasks in TIMSS 2011 shows that information packing, in the form of subject specific nouns, expressing mathematics as an object is more common in algebra and geometry compared to statistics and arithmetic. Examples of subject specific nouns describing mathematics as objects are *estimation* or *value*. Another example from TIMSS 2011 is the question, *What is the value of* $2a + 2b + 4$? A congruent linguistic form, expressing the task as a process, could have been a verb construction such as, *Calculate* $2a + 2b + 4$. The different use of incongruent versus congruent expressions in the different content areas show that language can be regarded as an integral part of the subject areas and is also an important factor for how students' are offered, or required, to understand the mathematical concepts.

Mathematical registers and teaching

The results of this study show that school mathematics does not have one subject specific language, rather the subject areas have different registers. This variance in language is an important issue for teaching since the choice of linguistic features are of significance for how mathematical concepts are perceived and for how the specific and the general is expressed. Previous research stresses that learning in a subject also involves learning the subject language (Abel & Exley, 2008; Lemke, 1990; Schlepperell, 2007; Unsworth, 1997). As Schleppegrell (2007) puts it:

> Of course the challenges of mathematics go beyond the language issues, but the linguistic challenges need to be addressed for students to be able to construct knowledge about mathematics in the ways that can ensure their success. (p. 156)

Concluding remarks

In this study linguistic features that fulfill the function of expressing central meaning dimensions of written language in mathematical tasks in TIMSS 2011 have been explored, and the language registers that these meaning dimensions express in the different mathematical content areas have been discussed.

The aim of this study was to investigate the function of linguistic features. Mathematical symbols and visual images were excluded, even though it is recognized that they just like linguistic features fill important functions. But the four meaning dimensions developed in this study of linguistic features are also useful for the study of mathematical symbols and visual images. Some indications of this are found in e.g. Fredlund's study (2015) where he has described how formulas are used for packing information in physics. An information packed formula becomes a useful tool for those who understand the meaning and can extract the information properly. Furthermore, regarding personification O'Halloran (2005) states that it is not possible to express appraisals, attitudes or feelings by symbols. Instead symbols appear to be value-free and expressing a high level of non-negotiable truth. When there are more symbols and fewer words in a task, the possibilities to express a concrete and personified content decreases. Symbols and digits in mathematical subject language will therefore be analyzed in light of the four meaning dimensions in a forthcoming study.

This study is based on an analysis of the TIMSS material, which was designed to correspond to the participating countries' policy documents. Therefore the content and how it is expressed by the four meaning dimensions can be assumed to correspond also to other teaching materials in the participating countries. However, there are certain aspects that make TIMSS a particular practice. Especially distinctive is the structure with multiple choice questions which might not be common in textbooks and other teaching materials. It is likely that this multiple choice structure of the tasks reduce the generalizability because some words and phrases are repeated in the answer alternatives. To what degree the findings of this study might apply to other mathematical practices such as textbooks, national or local tests, and classroom discussions, remains to be investigated.

Using the meaning dimensions to investigate the language in other practices could provide an enlarged understanding of the subject language. In this way it is not just the linguistic features as such that is scrutinized, it is also what function they fulfill in order to express different types of meaning. The theoretical framework and the results of this study can thereby form a launching pad for further research. The meaning dimensions form a meta-language in order to study the function of the language used in school mathematics, and registers typical of different content areas and in different levels of schooling may be identified.

References

Abel, K. & Exley, B. (2008). Using Halliday's functional grammar to examine early years worded mathematics texts. *Australian Journal of Language and Literacy*, 31 (3), 227–241.

Barwell, R. (2005). Working on arithmetic word problems when English is an additional language. *British Educational Research Journal*, 31 (3), 329–348. doi:10.1080/01411920500082177

Björnsson, C. H. (1968). *Läsbarhet* [Readability]. Stockholm: Bokförlaget Liber.

Borg, E. & Westerlund, J. (2006). *Statistik för beteendevetare* [Statistics for social science]. Stockholm: Liber.

Christie, F. (2012). Language education throughout the school years: a functional perspective. *Language Learning*, 62 (Supplement 1), 1–247.

Coffin, C. & Donohue, J. (2014). A Language as social semiotic-based approach to teaching and learning in higher education. *Language Learning*, 64 (1), 1–308.

Edling, A. (2006). *Abstraction and authority in textbooks: the textual paths towards specialized language.* Uppsala: Acta Universitatis Upsaliensis.

Fang, Z., Schleppegrell, M. & Cox, B. E. (2006). Understanding the language demands of schooling: nouns in academic registers. *Journal of Literacy Research*, 38 (3), 247–273. doi: 10.1207/s15548430jlr3803_1

Fang, Z. (2006). The language demands of science reading in middle school, *International Journal of Science Education*, 28 (5), 491–520. doi: 10.1080/09500690500339092

Fredlund, T. (2015). *Using a social semiotic perspective to inform the teaching and learning of physics*. Uppsala: Acta Universitatis Upsaliensis.

Geijerstam, Å. af (2006). *Att skriva i naturorienterande ämnen i skolan* [Writing in the science subjects at school] (Ph. D. dissertation). Institutionen för lingvistik och filologi, Uppsala universitet.

Halliday, M. A. K. & Hasan, R. (1989). *Language, context, and text: aspects of language in a social-semiotic perspective* (2nd ed.). Oxford University Press.

Halliday, M. A. K. & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (3rd ed.). London: Arnold.

Hellspong, L. & Ledin, P. (1997). *Vägar genom texten: handbok i brukstextanalys.* [Roads through the text: a manual to text analysis]. Lund: Studentlitteratur.

Herbel-Eisenmann, B. A. (2007). From intended curriculum to written curriculum: examining the "voice" of a mathematics textbook. *Journal for Research in Mathematics Education*, 38 (4), 344–369.

Lemke, J. L. (1990). *Talking science: language, learning, and values.* Norwood: Ablex.

Liberg, C. & Forsbom, E. (2009). Text and language in assessment of mathematics and science. In A. Saxena & Å. Viberg (Eds.), *Multilingualism. Proceedings of the 23rd Scandinavian Conference of Linguistics* (pp. 328–332). Acta Universitatis Upsaliensis. Studia Linguistica Upsaliensia 8.

Magnusson, U. & Kokkinakis, S. J. (2009). Quantitative measures on student texts. *Papers from the ASLA symposium in Stockholm, 7–8 November, 2008. ASLA:s skriftserie*, 22, 43–56.
Martin, J. R. (1991). Nominalization in science and humanities: distilling knowledge and scaffolding text. In E. Ventola (Ed.), *Functional and systemic linguistics* (pp. 307–337). Berlin: Mouton de Gruyter.
Morgan, C. (1998). *Writing mathematically: the discourse of investigation.* London: Falmer Press.
Morgan, C., Craig, T., Schuette, M. & Wagner, D. (2014). Language and communication in mathematics education: an overview of research in the field. *ZDM Mathematics Education*, 46 (6), 843–853. doi:10.1007/s11858-014-0624-9
Mühlenbock, K. H. (2013). I *see what you mean: assessing readability for specific target groups.* Institutionen för svenska språket, Göteborgs universitet.
O'Halloran, K. L. (2005). *Mathematical discourse: language, symbolism and visual images.* London: Bloomsbury UK.
Palm, T. (2002). *The realism of mathematical school tasks: features and consequences* (Ph.D. dissertation). Department of Mathematics, Umeå University.
Schleppegrell, M. J. (2001). Linguistic features of the language of schooling, *Linguistics and Education*, 12 (4), 431–459. doi:10.1016/S0898-5898(01)00073-0
Schleppegrell, M. J. (2004). *The language of schooling: a functional linguistics perspective.* London: Lawrence Erlbaum Associates.
Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: a research review. *Reading & Writing* Quarterly, 23 (2), 139–159. doi:10.1080/10573560601158461
Shanahan, T. & Shanahan, C. (2012). What is disciplinary literacy and why does it matter? *Topics in Language Disorders*, 32 (1), 7–18. doi: 10.1097/TLD.0b013e318244557a
Shreyar, S., Zolkower, B. & Pérez, S. (2010). Thinking aloud together: a teacher's semiotic mediation of a whole-class conversation about percents. *Educational Studies in Mathematics*, 73, 21–53. doi:10.1007/s10649-009-9203-3
Sfard, A. (1991). On the dual nature of mathematical conceptions: reflections on processes and objects as different sides of the same coin, *Educational Studies in Mathematics*, 22, 1–36. doi: 10.1007/BF00302715
Skolverket. (2012). *TIMSS 2011. Svenska grundskoleelevers kunskaper i matematik och naturvetenskap i ett internationellt perspektiv.* [Swedish primary school students' knowledge of mathematics and science in an international perspective]. Stockholm: Skolverket.
Skolverket. (2014). *TIMSS 2011. Uppgifter i matematik, årskurs 8.* [TIMSS 2011. Tasks in mathematics, grade 8] (Rapport 401). Stockholm: Fritzes.

Unsworth, L. (1997). Scaffolding reading of science explanations: accessing the grammatical and visual forms of specialized knowledge. *Literacy*, 31(3), 30–42. doi: 10.1111/1467-9345.00061

Wagner, D., Dicks, J. & Kristmanson, P. (2015). Students' language repertoires for prediction. *Mathematics Enthusiast*, 12(1–3), 246–261.

Veel, R. (1997). Learning how to mean-scientifically speaking: apprenticeship into scientific discourse in the secondary school. In F. Christie & J. R. Martin (Eds.), *Genre and institutions: social processes in the workplace and school* (pp. 161–195). London: Cassell.

Veel, R. (1999). Language, knowledge and authority in school mathematics. In F. Christie (Ed.), *Pedagogy and the shaping of consciousness: linguistic and social processes* (pp. 185–216). New York: Cassell.

Österholm, M. & Bergqvist, E. (2013). What is so special about mathematical texts? Analyses of common claims in research literature and of properties of textbooks. *ZDM Mathematics Education*, 45(5), 751–763. doi: 10.1007/s11858-013-0522-

Ida Bergvall

Ida Bergvall has a PhD from the Department of Education, Uppsala University, Sweden. Ida has a background as mathematics, science and special education teacher. Her research and thesis, *Bokstavligt, bildligt och symboliskt i skolans matematik – en studie om ämnesspråk i TIMSS*, focus on language in school mathematics.

ida.bergvall@edu.uu.se

Jenny Wiksten Folkeryd

Jenny Wiksten Folkeryd is assistant professor at the Department of Education, Uppsala University, Sweden. Her research focuses educational linguistics including reading and writing development, language across the curriculum, text analysis and literacy assessment. She participates in the construction of national tests in Swedish and is a member of the Reading Development Group of PIRLS (Progress in International Reading Literacy Study).

jenny.folkeryd@edu.uu.se

Caroline Liberg

Caroline Liberg is a Professor of educational sciences focusing on reading and learning processes at the Department of Education, Uppsala University. Her research has focused on literacy practices in early childhood and in different subject areas in school. Together with her doctoral students and colleagues, Caroline Liberg is currently investigating linguistic and content dimensions of test items in international studies of students' competences in math and science and furthermore the function, content and form of students' text-making in different subject areas in early school years.

caroline.liberg@edu.uu.se