International Electronic Journal of Mathematics Education
2023, 18(1), em0719
e-ISSN: 1306-3030
https://www.iejme.com

MODESTUM

Research Article

OPEN ACCESS

Unsupervised machine learning to classify language dimensions to constitute the linguistic complexity of mathematical word problems

David Bednorz [1]*, Michael Kleine [2]

[1] Department of Mathematics Education, IPN-Leibniz Institute for Sciene and Mathematics Education, Kiel, GERMANY
[2] Department for Mathematics Education, Bielefeld University, Bielefeld, GERMANY
***Corresponding Author:** bednorz@leibniz-ipn.de

Citation: Bednorz, D., & Kleine, M. (2023). Unsupervised machine learning to classify language dimensions to constitute the linguistic complexity of mathematical word problems. *International Electronic Journal of Mathematics Education, 18*(1), em0719. https://doi.org/10.29333/iejme/12588

ARTICLE INFO

Received: 03 Jun. 2022

Accepted: 26 Sep. 2022

ABSTRACT

The study examines language dimensions of mathematical word problems and the classification of mathematical word problems according to these dimensions with unsupervised machine learning (ML) techniques. Previous research suggests that the language dimensions are important for mathematical word problems because it has an influence on the linguistic complexity of word problems. Depending on the linguistic complexity students can have language obstacles to solve mathematical word problems. A lot of research in mathematics education research focus on the analysis on the linguistic complexity based on theoretical build language dimensions. To date, however it has been unclear what empirical relationship between the linguistic features exist for mathematical word problems. To address this issue, we used unsupervised ML techniques to reveal latent linguistic structures of 17 linguistic features for 342 mathematical word problems and classify them. The models showed that three- and five-dimensional linguistic structures have the highest explanatory power. Additionally, the authors consider a four-dimensional solution. Mathematical word problem from the three-dimensional solution can be classify in two groups, three- and five-dimensional solutions in three groups. The findings revealed latent linguistic structures and groups that could have an implication of the linguistic complexity of mathematical word problems and differ from language dimensions, which are considered theoretically. Therefore, the results indicate for new design principles for interventions and materials for language education in mathematics learning and teaching.

Keywords: language dimensions, mathematical word problems, linguistic complexity, machine learning, unsupervised machine learning

INTRODUCTION

Mathematical word problems are a typically exercises for mathematics instruction. Specific for word problems are, that beside mathematical requirements also language requirements are very important (Abedi et al., 2020). One aspect of this language requirements is the linguistic complexity of word problems (Martiniello, 2008, 2009). The linguistic complexity is particular high for word problems with academic (and mathematics) language (Biber & Gray, 2016; Snow & Uccelli, 2009). On the perspective of different mathematical task types, it can be assumed that the influence of linguistic complexity can be high for word problems, especially with real world context (Verschaffel et al., 2020). This is supported by studies comparing monolingual and bilingual students (or learners with a migration background). Regarding to this, Heinze et al. (2007) point out that the language proficiency, for certain mathematical problems (different subscales), has an influence on the construction of mental representations (e.g., mental models (vom Hofe et al., 2005)). This applies to word problems with real world context, whereas simple addition and subtraction tasks are less dependent on the language proficiency. The individual language proficiency of students has an influence how good word problem are solved which are embedded in a real-world context, because of the higher linguistic complexity of the word problems (Pongsakdi et al., 2020). Furthermore, Daroczy et al. (2020) assume that first, text features and content features exist independently and second, text features and content features exist intrinsically related to each other (cognitive function of academic language). This in turn can have an impact on how linguistic complexity affects solving the word problem (Daroczy et al., 2015). Therefore, dimensional language approaches in mathematics education research could help to structure and classify language in word problems and to make implication on the linguistic complexity (Prediger et al., 2019) and helping educational researchers and educators to know which language structures can be a possible obstacle in solving word problems and to subsequently develop suitable action plans for classroom instructions (Schleppegrell & O'Halloran, 2011). Most of language dimensional approach build dimensions according to theoretical considerations. Machine learning (ML) is a newly high statically approach in educational research, e.g., for natural language processing (NLP) application (Alexopoulou et al., 2017). Such approaches could help to make empirical submissions about the linguistic dimensional structures e.g., of mathematical word problems (Biber & Gray, 2013a). These linguistic structures could help to explore the linguistic complexity of word problems and to classify mathematical word problems along these structures to make implication on the linguistic complexity of specific groups of mathematical word problems. Therefore, the study focuses on the exploration and the classification of the dimensional linguistic structures of mathematical word problems with ML approaches.

LITERATURE REVIEW

Linguistic Complexity of Mathematical Word Problems

Mathematical problems can be numerical or can be phrased in words. Word problems are typical problems for mathematics instruction and are investigated in various facets in the literature (for an overview see Verschaffel et al., 2020). Word problems can be distinguished between mathematical word problems without contextual references to the real words e.g., "what is the sum of 1/2 and 2/3", and word problem with contextual references to the real world such as "Karin take a half pizza funghi and a quarter pizza margarita. How much pizza does Karin eat?" Word problems with real world contextual references are also known as *modeling problems or modeling task* (Verschaffel et al., 2020).

One mayor challenge for solving mathematical words problems is besides the required domain specific requirements, e.g., mathematical abilities, cognitive demands, the personal or tasks-based requirements, especially language abilities of the student and linguistic feature of the word problem (Boonen et al., 2013, 2016; Vilenius-Tuohimaa et al., 2008). Therefore, language can have an influence on the difficulties of mathematical word problems. To explain the difficulty of linguistic features one perspective is the use of language in different contextual situation such academic situations (academic language or more specifically mathematics and mathematics language), everyday situations (everyday language) (Halliday, 1993, 2004, 2007, 2014; Lemke, 2012; Schleppegrell, 2006).

Textual features of academic language (Halliday, 2014; Lukin et al., 2011; Morek & Heller, 2012) are associated with high linguistic complexity. The concept of academic language can be determined from different perspectives (Morek & Heller, 2012). In the context of academic language, the cognitive or epistemic function is highlighted (Maier & Schweiger, 1999; Morek & Heller, 2012). This means that academic language is used on a cognitive level as a tool for thinking, and thereby assumes the function as a descriptive and interpretive tool to extend, simplify, and abbreviate thinking (Klix, 1995; Tomasello, 2008). For the focus on text features, academic language can be described under the umbrella term register. The term register is a concept of language variation and defines the correspondence between language variation and situation variation (Biber, 2006; Reid, 1956). Thus, a register is meant to represent how the use of language corresponds with certain situations (Halliday, 2005). For academic language as a register, this means that (typical) text features can be identified that (typically) occur in the context of academic situations. A selection of academic language text features is offered by Morek and Heller (2012). In addition, there are further divergent lists of academic language text features in the literature, which can be linguistically complex (Gogolin & Lange, 2011; Heine et al., 2018; Heppt et al., 2020; Schleppegrell, 2005). Central textual features of academic language are the vocabulary used, nominalization, connectors and the frequent use of long nominal phrases or subordinate clause and passive constructions, composites, impersonal language. These text features are assumed to lead to high linguistic complexity in a text.

For other text features there is also evidence that they can reduce the linguistic complexity of a text. For example, text features of the everyday register are related to reduce linguistic complexity. This can happen, for example, through contextualization, personalization (pronouns), and formation of active sentences of the text (Abedi & Lord, 2001; Cummins et al., 1988; Gogolin & Lange, 2011). This can lead, as in the case of contextualization, to an increase in the number of certain text features. Likewise, main clauses, known vocabulary, and short sentences are supposed to reduce linguistic complexity, which are often directly related to the academic text feature counterpart that increase linguistic complexity, such as pronounced subordinate clause constructions, unknown vocabulary, long sentences. Other text features that can reduce linguistic complexity are reference text features for coherence building. Text coherence is achieved through strategies to create meaningful connection between sentences through certain text features, for example, conjunctions, polysemy, or anaphora, and through other features of the text, such as representations (Dittmar et al., 2017; Leiss et al., 2017; McNamara et al., 1996, 2011; McNamara & Kintsch, 1996). Heine et al. (2018) list the relevance of coherence structures for linguistic complexity, in the category of uniqueness of form-meaning relations, in their analyses. For this reason, textual reference structures and text-image references (as textual features of a discontinuous text) also seem to be relevant to reduce linguistic complexity. In view of this, Martiniello (2009) was able to show for representations that they can contribute to reduction of linguistic complexity, in case of his study, of mathematical test items.

To localize where language obstacles occur due to the mentioned text features, in mathematics education it is a frequent practice to build dimensions of linguistic features (Heine et al., 2018; Prediger et al., 2019). In the most cases these dimensions are construe after theoretical reasons and only few using empirical approaches to evaluate language dimensions (Bednorz, 2021).

For mathematics education different studies researched about the influence of the linguistic complexity and specific linguistic features on the difficulty of mathematical word problems, among other things, for mathematical assessments (Abedi & Herman, 2010; Abedi & Lord, 2001; Ufer & Bochnik, 2020). Abedi and Lord (2001) have determined differences between numerical formats and word problems. Linguistic features have a high relevance for mathematical assessments, especially for learners with low language abilities, whose mathematical abilities are likely to be underestimated in assessment due to language obstacles (Abedi & Gándara, 2006; Abedi & Herman, 2010; Hofstetter, 2003). Additional studies shows that linguistic feature of mathematical word problems have an influence of the difficulties for this kind of tasks. Studies point out that language difficulties can be the basis for test constructs irrelevant difficulties for learners who differ in their language abilities (Abedi et al., 2008; Haag et al., 2015; Martiniello, 2008, 2009). Martiniello (2008) indicate that the linguistic difficulties of mathematical word problems increase for learners with low language abilities. As a reason for these increased difficulties, relevant linguistic features that are relevant for the solution process are worked out in a qualitative analysis. In addition, Haag et al. (2013) consider in their study the effect of academic language features on differential functioning of test items with regard to the group of learners with lower language abilities. The results also indicates that the linguistic features have an effect to differences in the difficulties of word problems and that the linguistic features are cause a high proportion of differential item functioning (Haag et al., 2013). Martiniello (2009) can show that the influence of linguistic features on linguistic complexity results in a positive effect on the difficulties of mathematical word problems. Martiniello (2009) explicitly refers to the connection between linguistic complexity and linguistic features, especially of the academic language, which contribute significantly to the difficulties of mathematical word problems.

Application of Machine Learning in Education for Text Analysis

ML is a relatively new high potential approach for the analysis of data (especially for big data). Jordan and Mitchell (2015) have defined ML as learning process of computers with the goal of automatically improvement. This definition focused on ML supervised learning techniques of a computer, especially neural networks. In addition, unsupervised learning techniques as dimension reduction, clustering, and regressions are also potential ML approaches to identify underlying patterns of data (Gentleman & Carey, 2008; Nelson, 2020; Rosenberg & Krist, 2021). Different applications in the fields of Technology and Science, Marketing, Sociology or Politics and also in Education and educational research are possible in using ML.

Educational researchers are using ML, e.g., for construing or evaluating intelligent tutoring systems, interactive learning environments, adaptive learning and for NLP, e.g., for text analysis (Abidi et al., 2018; Aleven & Koedinger, 2002; Amir et al., 2014; Balyan et al., 2020; Chen et al., 2020; Conati & Merten, 2007; Gobert & Sao Pedro, 2016; Holstein et al., 2019; Mousavinasab et al., 2021; Ortmann & Dipper, 2019; Roll & Wylie, 2016).

NLP approaches and the potential of text analysis are used in different ways in educational research (Almatrafi et al., 2018; Lin et al., 2009; Wu et al., 2020). One major application is the automation of coding procedures for educational research (Nelson et al., 2018). A second applications is the analysis of linguistic features and the evaluation of linguistic complexity with NLP methods (Alexopoulou et al., 2017; Weiss & Meurers, 2019). Another perspective are unsupervised ML techniques like dimension reduction methods, to analysis linguistic features to reveal latent linguistic structures (Biber, 2006). The dimension reduction procedure systematizes the linguistic feature and can be used to explore the dimensional structure of linguistic feature (Biber & Egbert, 2018; Conrad, 2015). Biber (2006) interpreted the dimension and the linguistic patterns as variation of linguistic feature and stated that the constitution of the patterns is because of a mutual language function of the linguistic features (Biber & Conrad, 2019; Biber & Gray, 2013a; Biber et al., 1998, 2016; Conrad, 2015).

The analysis process for language dimension can be differentiated in three steps with two steps which involves ML techniques. The first step is an evaluation of the frequencies of linguistic features, mostly with NLP techniques to build up the text corpus as data source. The second step is the dimensions reduction process of the linguistic features with a factor analysis to explore hidden linguistic pattern. The third step is a functional interpretation of the evaluated linguistic patterns (Biber & Egbert, 2018). Biber (2006) used the dimensional approach for macro- and microscopic analysis for the evaluation of linguistic pattern structures, which are called language dimensions, for different type of text, such as texts with academic or everyday language, (Biber, 1985, 2006; Biber & Gray, 2013a, 2013b, 2016). The awareness of this linguistic pattern could be used for further analysis of the linguistic complexity of mathematical word problems.

In order to make conclusions about the language dimensions, it is necessary to collect a higher number of text and additionally linguistic features (Biber & Reppen, 2002). Therefore, McEnery and Hardie (2012) called this approaches highly statically and time-consuming analysis. Regarding to this, Conrad (2015) stated that this problems, especially in an effectiveness perspective, could be a reason why this approach has a modest spreading in education research so far. This is why ML techniques are needed to analyse language dimensions to make this a rational approach for educational research, with existing digital infrastructures such as NLP analysis tools, existing references corpora and software for the analysis.

Research Questions

To solve a mathematical word problems domain specific abilities are necessary. Different studies pointed out that also language specific requirements are important and can have an influence for the solution of mathematical word problems (Abedi, 2006; Abedi & Lord, 2001; Boonen et al., 2013; Kiplinger et al., 2000; Martiniello, 2009; Pongsakdi et al., 2020). Linguistic features, such as features of the academic language, can lead to language obstacles for the word problems (Kieffer et al., 2009; Snow & Uccelli, 2009).

To analyzing language obstacles a typically process in mathematics education is to sort the linguistic features in dimensions. Mostly researcher build these dimensions after a theoretical perspective such as a sorting of linguistic features after word, sentence, and text basis (Prediger et al., 2019). However, in the context of mathematical words problems and the linguistic complexity of these tasks, it is unclear what latent linguistic structure could be determined empirically and which implication on the linguistic complexity of mathematical word problems can be made. Dimensional reduction of linguistic features such as the approach of Biber (2006) can be a possible way to identify hidden linguistic feature patterns in language dimensions which are possible causal reasons for the linguistic complexity and for feasible obstacles in solving mathematical word problems. Therefore, the objective of this study is to identify different language dimensions of mathematical word problems, classify this patterns and discuss potential implication for the linguistic complexity. The authors of this study address following research questions:

1. RQ1. Which linguistic dimensional structures existing for mathematical word problems?
2. RQ2. Which classification of mathematical word problems with the dimensional structures can be made?
3. RQ3. What implication on linguistic complexity could be made for the different classified mathematical word problems?

METHODOLOGY

Sample

We have selected n=342 mathematical word problems for the study from German secondary mathematical schoolbooks, whereby 47 mathematical word problems were chosen from a mathematical assessment instrument (PALMA) (Murayama et al., 2013). The mathematical word problems are distrusted as a percentage to algebra with functions and calculus (58.57%), geometry (25.58%), and stochastics (15.58%). The distribution of the word problems to the mathematical content fields is in accordance with the curricular occurrence in the secondary school in Germany. Furthermore, we have raised 17 linguistic features for all of the selected 295 mathematical word problems. We have selected the linguistic features according to the relevance in literature, with a focus on linguistic features of the academic language (Biber & Gray, 2013b; Heppt et al., 2020; Schleppegrell, 2012). Firstly, we have selected linguistic features with a direct mathematical relationship (mathematical terms, numbers, discontinuous text, or mathematical symbols). Secondly, we selected linguistic features that are important for text relations (conjunctions, prepositions). Thirdly, we have collected features that have a connection to academic language (impersonal language, lexical diversity, proportional density, nominalization, compounds). And finally, we have chosen feature with a high occurrence in the everyday language (filler words, common words, present perfect, modal verbs, and synonyms).

Data Analysis

We distinguished the data analysis of this study in three processes.

Text to data

The first data analysis process is the analysis of the linguistic features and using text as data, with NLP. Therefore, we have analyzed the 295 mathematical word problems using POS-tagging to determine the 17 linguistic features automatically and calculated the frequencies of the features. We have used R for the text analyses with the package *koRpus* (Michalke, 2018).

Factor analysis

The second process of data analysis are using the collected data and to perform a factor analysis to reveal the dimensional linguistic structure. For the analysis we have used a principal component analysis with an oblique rotation technique. We have chosen the oblique rotation, because the dimensional structures of linguistic features can correlate with each other (Biber, 2006). We have used R for the factor analysis with the package psych (Revelle, 2021). The test criteria for a factor analysis were satisfied. For the correlation matrix of the quantified text features, Bartlett's test becomes significant with χ^2=1,672.36 (p<.001). Thus, the existing correlation matrix is significantly different from an identity matrix. The KMO or MSA values are the most appropriate method for testing the correlation matrix and should be necessarily tested before conducting a factor analysis (Kaiser, 1974). The overall KMO value for the present data for factorization is 0.67, which is in the upper range of medium suitability (≥.6 medium and ≥.7 fairly good). The MSA values for all variables are ≥0.50, which is in the adequate range. Removing variables with low MSA values could increase the global KMO value and improve the fit of the correlation matrix. Since variables that are particularly relevant for mathematics, such as mathematical terms, have low MSA values, we will continue the factor analysis with the existing linguistic features variables.

Latent class analysis

For the third process of the data analysis, we have performed a latent class analysis to classify the dimensional linguistic structures of the factor analysis in cluster. Therefore, we have used the z-scored factor scores from the language dimensions. Because of the z-scoring of the dimensions, positive numbers mean it is in the upper part of the distribution of the factor scores, and a negative number mean it in the lower part of the distribution of factor scores. z-scores have a distribution like, that a plus one means a positive standard deviation and minus one means a negative standard deviation. Which we have selected as the bases for a weighting of the results for the interpretation. After performing the cluster analysis, we have computed the distribution of the z-scores according to the clusters, for the following classification. We have used R for the cluster analysis with the package *mclust* (Scrucca et al., 2016).

RESULTS

Dimensional Reduction with a Factor Analysis

The first decision for a factor analysis to make, is to choose the numbers of dimensions. There are different methods to determine the number of dimensions. The two common methods are the very-simple-structure (VSS) criterion and the parallel analysis. We have carried out the two procedures below and then we have discussed, which number of factors we have extracted. Revelle and Rocklin (1979) designed the VSS criterion as a procedure for determining the optimal number of dimensions. By using the VSS criterion, the fit of a given number of dimensions loadings to the loading matrix is determined by deleting all but the c largest loadings per item, where c is a measure of factor complexity. This allows to compare a simplified model to the original correlations, with the VSS criterion reaching the highest value between 0 and 1 for an optimal number of dimensions.

Figure 1 shows the results of the VSS criterion and that for a VSS complexity of c=1, the highest VSS fit value is obtained for two factors. Also, **Figure 1** shows that the highest VSS fit values have the three-dimensional solution with a complexity of c=3 and for a five-dimensional solution with a complexity of c=4. The highest VSS fit for three-dimensional VSS (3)=0.70 is as high as that

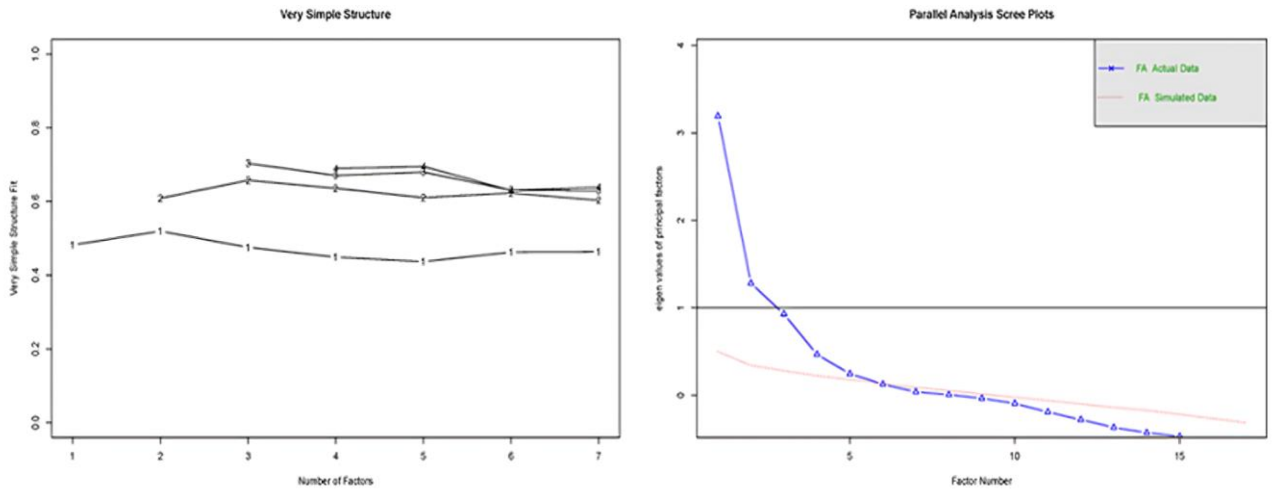

Figure 1. Selection of the number of dimensions for factor analysis by two possible methods. On the left the VSS procedure to determine the optimal number of dimensions. On the right, the parallel analysis, which uses a bootstrap procedure to repurpose a data structure to determine the optimal number of dimensions. The VSS criterion points to an optimal number of three dimensions for a dimensional complexity of two and five dimensions for a dimensional complexity of three. The parallel analysis represents that a factor solution of five factors is probably optimal (Source: Authors' own illustration)

for five-dimensional with VSS (4)=0.70. The VSS criterion results tend to indicate a three- or five-dimensional solution. To compare the results of the VSS fit, we have performed a parallel analysis as another criterion. This provides another indicator to check which number of dimensions we should extracted for further analysis, thus allowing a more explicit determination. Parallel analysis compares the existing data set with solutions of random data that have the same properties as the existing data set. The bootstrap procedure draws 1,000 bootstrap samples from the existing data set to reproduce the dimensional structure through a secondary sample to obtain an empirical sampling characteristic distribution. In **Figure 1**, the lower non-straight horizontal line indicates the result of parallel analysis.

According to the parallel analysis criterion, a solution with five dimensions is optimal. This result coincides with the VSS fit with five dimensions at a dimensional complexity of c=4. Considering and comparing the two methods, the extraction of five dimensions seems to be the optimal choice. When choosing the number of dimensions, two aspects are important. On the one hand, if the number of dimensions is too high, there is a risk that not all factors can be interpreted in a meaningful way. On the other hand, if the number of dimensions is too low, there is a risk that information regarding relevant structures will be lost. For this reason, we have extracted the three- and five-dimensional solutions to examine which of the dimensional solutions show relevant latent linguistic structures. To contrast the results and to be sure that the three- and five-dimensional solution are optimal, we have extracted the four-dimensional solution with an also high VSS fit too.

In **Figure 2**, the different solutions of the factor analysis are shown. Recognisable are the latent linguistic structures in the correlation diagrams for all three feasible solutions. For the three-dimensional solution the first dimension structures especially conjunctions, prepositions, impersonal language, modal verbs, filler words, synonyms, lexical diversity, nominalization, and present perfect. The second dimension structures particular compounds, discontinuous text, and passive. The third dimension have a high correlation with numbers, mathematical symbols, and mathematical terms. And with a negative correlation common words and propositional density.

For the four-dimensional solution the first dimension combined particular mathematical terms, preposition, numbers, conjunctions, nominalization, mathematical symbols, and filler words. The second dimension have high correlations with compounds, passive, discontinuous text, and lexical diversity. The third dimension have a high relationship with linguistic features like lexical diversity, synonyms, modal verbs, and present perfect. The last dimensional structure sums up common words, impersonal language, and propositional density. The five-dimensional solution have high correlation for the first dimension with conjunctions, impersonal language, prepositions, nominalization, and filler words. The second dimension have high correlation with compounds, passive, discontinuous text, and lexical diversity. The third dimension have a strong positive relationship with lexical diversity, synonyms, modal verbs, present perfect and negative relationship with discontinuous text, and mathematical terms. The fourth dimension structures common words and propositional density and have a negative correlation with numbers. The fifth dimension is strongly characterized by mathematical symbols and in smaller proportion by mathematical terms.

Stable linguistic structures above all dimensional solutions

Single structures are stable above all dimensional solutions. One strong relationship between linguistic features are conjunctions and prepositions as text relation structures with filler words and nominalization. These linguistic features occur in all dimensional solutions for the first dimension. A second stable linguistic structure is a combination of compounds, passive and discontinuous text which occur for all dimensional solution in the second dimension. Another linguistic structures which occur is the relationship between lexical diversity, synonyms, modal verbs, and present perfect. This linguistic structure occurs for the three-dimensional solution in the first dimension, constituting mostly the third dimension in the four and five-dimensional solutions. Furthermore, common words and propositional density have a strong relationship. In the three-dimensional solution

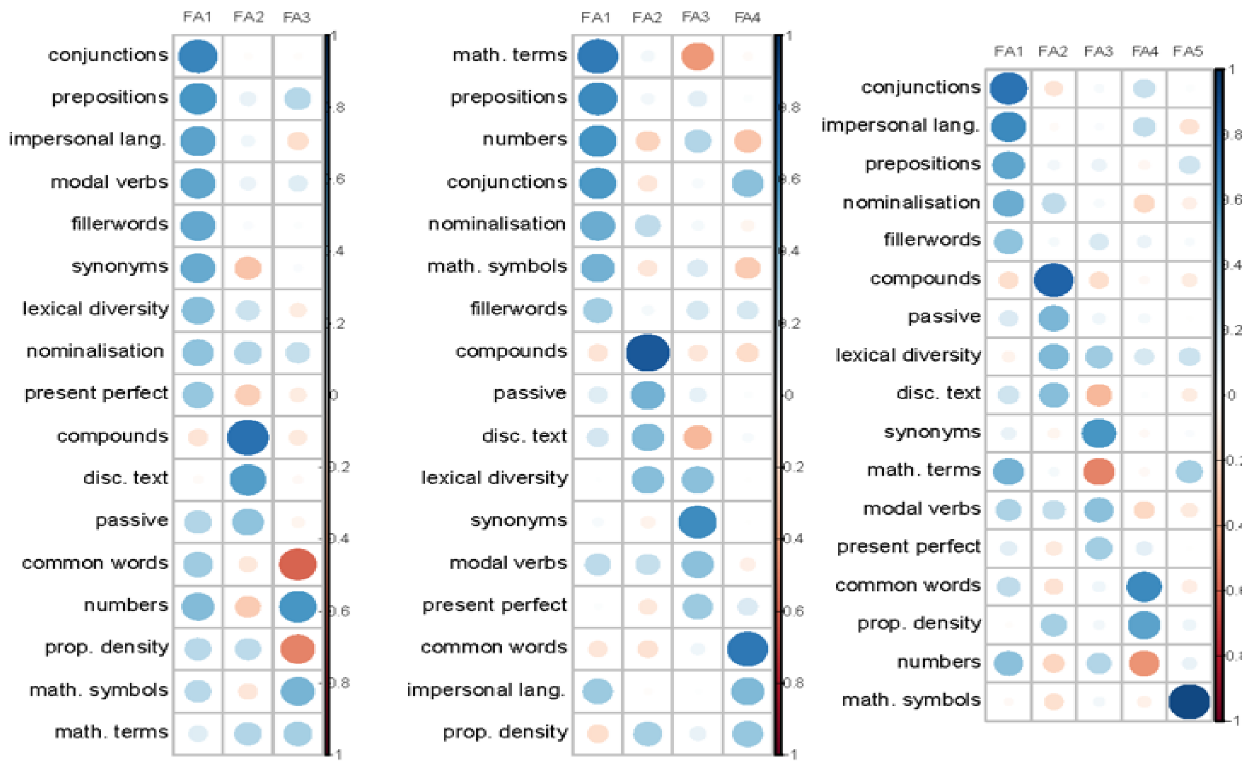

Figure 2. Three rotated dimensional solution with different dimension extracted. On the left the solution with three dimension. On the middle the four dimensional solution. On the right the solution with five dimensions considered. To see in the figure are a correlation diagram of the dimensional structures of the different solutions. The size and the colouring of the circles indicates the correlative relationship to the dimension. Blue indicates a positive correlation between linguistic feature and dimension. Red indicates a negative correlation between both (Source: Authors' own illustration)

Table 1. Variable correlation with the dimensional structure

		Three dimensions			Four dimensions				Five dimensions				
		1	2	3	1	2	3	4	1	2	3	4	5
Linguistic features variables													
Mathematical language	Mathematical terms			0.33	0.70		-0.43		0.48		-0.50		0.33
	Numbers	0.44		0.58	0.59				0.42			-0.45	
	Discontinuous text		0.56			0.43	-0.32			0.42	-0.36		
	Mathematical symbols			0.47	0.47								0.91
Text relations	Conjunctions	0.65			0.57			0.42	0.73				
	Prepositions	0.58			0.63				0.53				
Academic language	Nominalisation	0.41			0.50				0.50				
	Passive		0.41			0.47				0.46			
	Impersonal language	0.53			0.37			0.45	0.63				
	Lexical diversity	0.43				0.43	0.42			0.45	0.45		
	Propositional density			-0.49		0.33		0.38		0.33		0.54	
	Compounds		0.74			0.84				0.81	0.81		
Everyday language	Common words	0.36		-0.59				0.71				0.63	
	Present perfect	0.39					0.36				0.35		
	Synonyms	0.51					0.63				0.59		
	Modal verbs	0.52					0.42		0.31		0.42		
	Filler words	0.52			0.37				0.41				

Note. Linguistic features correlation with r>.3 with different dimensional solutions. Linguistic features are ordered after typical allocation that are made in the literature. From mathematical linguistic features, features for text relations, features, which are often connected to academic and last one to everyday language. This is a simplification of possible allocation, but makes interpretation of linguistic structures feasible

both linguistic features have a negative correlation with the third dimension and distinguished from numbers, mathematical symbols, and mathematical terms. The negative relationship between these two linguistic features and numbers also occurs in the five-dimensional solution in the fourth dimension. Not so strong are the relationship between mathematical terms, mathematical symbols, and numbers. They constituted in positive correlation the last dimension in the three-dimensional solution, have a strong correlative impact for the first dimension for the four-dimensional solution, but in the five-dimensional solution mathematical symbols are the major linguistic feature for the fifth dimension although it has only a low relationship with mathematical terms left and no considerable correlation with numbers.

The correlative relationship between the linguistic features and the dimensional structures are shown in **Table 1**.

Table 2. Summary of the results of the dimensional reduction of the linguistic features and labelling of characteristic structures

Stability of the linguistic feature structures		Allocation to different types of linguistic features structures		Label
Strong relationships	**Occurs**	**Types**	**Examples**	
Conjunctions, prepositions, filler words, & nominalisation	All dimensional solutions (first dimension)	1. Mixture over all types of linguistic features	Dimension one all dimensional solutions	ST mix
Compounds, passive, & discontinuous text	All dimensional solutions (second dimension)	2. Linguistic features mostly for mathematics purposes and academic language	Dimension two all dimensional solutions	MAL mix
Lexical diversity, synonyms, modal verbs, & present perfect	Three-dimensional solution (first dimension) Four- & five-dimensional solutions (mainly constituting the third respectively five dimension)	3. Linguistic features of academic or everyday language and features for text relations	Dimension four for the four-dimensional solution	AELR mix
Common words & propositional density	Three-dimensional solution (third dimension with negative correlation to numbers, mathematical symbols, & terms) Four-dimensional solution (second dimension with correlation to lexical diversity)	4. Linguistic features of academic or everyday language and negative correlated to features for mathematics purposes	Dimension three & four for the five-dimensional soliton. Dimension three for the four-dimensional solution	AEL-NM mix
Mathematical terms, mathematical symbols, & numbers	Three-dimensional solution (third dimension) Four-dimensional solution (strong impact for the first dimension) Five-dimensional solution (only a small correlation between mathematical symbols & terms)	5. Only (positive) linguistic features for mathematics purposes	Dimension three for the three-dimensional solution Dimension five for the five-dimensional solution	MA mix

Additionally, we have made a connection of the linguistic features with typically associated contextual situations. It should be notice that the made connections are only helping structures for the interpretation and labelling of the results. The allocation between linguistic features and situation is not discrete but continuous as Biber (2006) mentioned. Therefore, we have made these connections between features and typically situations, because of the most reasonable frequency of linguistic features in the different contextual situation.

Labelling the linguistic structures

To have a conceptual reference the linguistic structures maintain a label, related to the specification of the linguistic features on the dimensions. The differentiation of the linguistic features in **Table 1** shows that the first dimension for all solution extends over all types of linguistic features. Because of that, we have labelled this structure *structure mix* (ST mix). The second structure are existing for all dimensional solutions in the second dimension. This dimension subsumes linguistic feature mostly from the mathematics and academic language. Therefore, we have labelled this linguistic structure as *mathematics, academic language mix* (MAL mix). In some dimensions mathematics language features appears alone, like in dimension five for the five-dimensional solution, or are the only positive related linguistic features to a dimension like in the third dimension in the three-dimensional solution. For this reason, we have labelled that structure *mathematics language mix* (MA mix). Other linguistic structures are related to the academic or everyday language. They exist for the third and fourth dimension for the five-dimensional solution and third dimension for the four-dimensional solution each time with a negative correlation with linguistic features for mostly mathematical purposes. Because of the occurrence of the academic and everyday language in this structure and the always negative relationship to mathematics language we have labelled this structure *academic, everyday language and negative mathematics language mix* (AEL-NM mix). Furthermore, the connection between features from the academic or everyday language are also related to text relation for the four-dimensional solution in the last dimension. Therefore, we have labelled this structure as *academic, everyday language and text relation mix* (AELR mix).

Table 2 depicts the summary of the results of the dimensional reduction of the linguistic features.

Classification of the Mathematical Word Problems

After the analysis of the different dimensional solution and the labelling of relevant structures, the next part of the examination is the classification of the word problems with the dimensional structure with a cluster analysis. Likely as for the factor analysis, the first step is to explore how much cluster are reasonable to select. Therefore, we have performed 14 different Gaussian models with a variation of possible cluster ellipses which differed in distribution, volume, shape, and orientation. Numbers of clusters were select with a minimal BIC as best solution. The results of the comparison of the different models are shown in **Figure 3**.

The optimal number of clusters for the five-dimensional solutions is according to the BIC criteria a three-cluster solution (VEE (*ellipsoidal equal shape*) BIC=-3,849.843). Other possible numbers of clusters are three (VVE-BIC=-3,868.625) and four (EEE=-870.034).

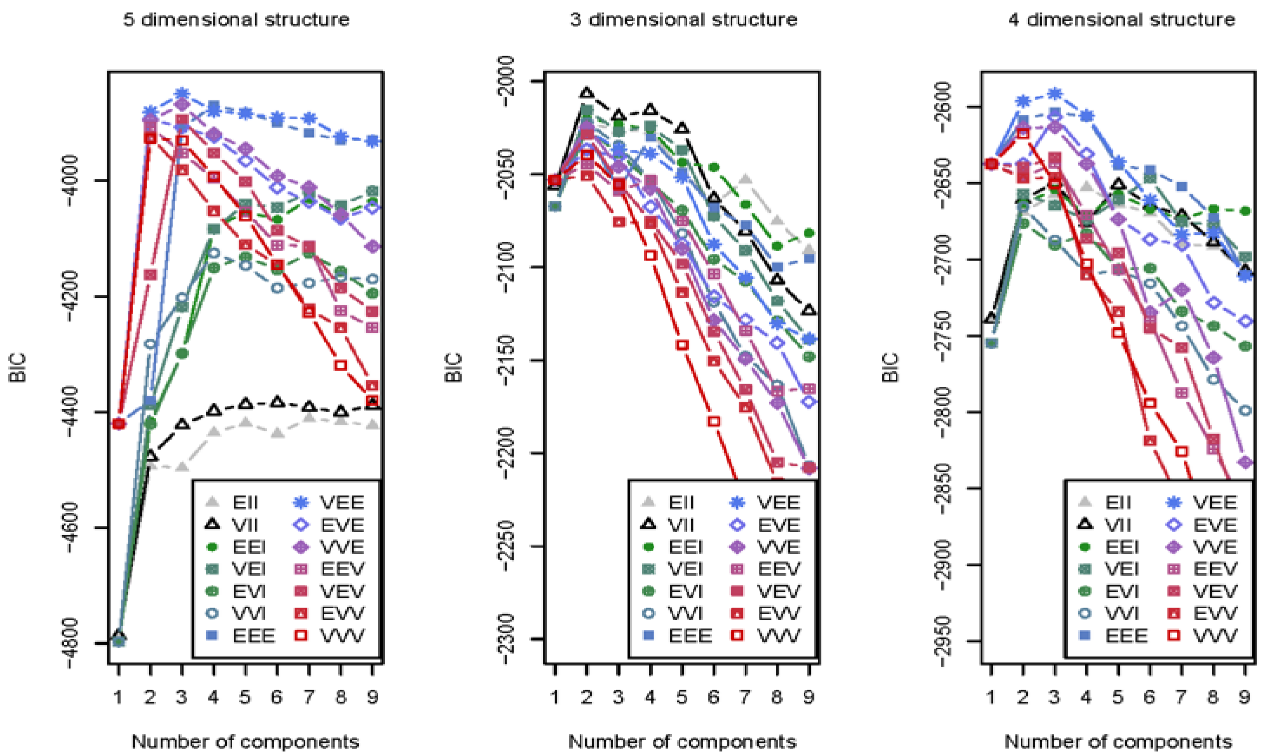

Figure 3. Representation of ellipses of isodensity for 14 Gaussian models receive by Eigen-decomposition for the five, three, and four dimensional solution for different number of components (clusters), to select the optimal number of clusters. Ellipses vary in distribution, volume, shape, and orientation (Source: Authors' own illustration)

The BIC criteria indicate for the three-dimensional solution an optimal number of two cluster (VII-[*spherical varying volume*] BIC=-2,006.633). Other options are also two (VEI-BIC=-2,015.356) or four cluster (VII-BIC=-2,015.672). For the four-dimensional solution the BIC criteria indicate a three cluster solution (VEE-[*ellipsoidal equal shape*] BIC=-1,591.188). Further possible cluster solutions, with a clearly higher BIC, are two clusters (VEE-BIC=-2,596.199), and three clusters (EEE-BIC=-2,602.898).

With the feasible best number of clusters for all dimensional solution we have calculated the frequency distribution to classify the different clusters. **Figure 4** shows the results of the calculation. Because the linguistic feature variables were z-scored after the dimensional reduction, the scores of the y-axis are also z-scored. With the z-scored distribution of the factor scores we have weighted the difference of one standard deviation, as a big difference to the mean. We weighted a half standard deviation (0.5) as an average difference to the mean and we have weighted a quarter standard deviation (0.25) as a slightly difference to the mean. If it possible for the classification we have only considered big and average differences. If we only can distinguish slightly differences, we have considered slightly difference too.

Classification of word problems according to the three-dimensional solution

The clusters of the three-dimensional solution have two strongly different distribution of frequencies on the clusters. The first cluster is characterized with a slightly to almost average negative difference to the mean. Whereby the first ST mix dimension has the highest, the second MAL mix dimension has the medium high, and the third MA mix dimension the lowest differences. For the second cluster the differences to the mean are overall positive with differences to the mean around a quarter standard deviation. The distribution is similar, the only difference is that the second MAL mix dimension has a slightly lower difference to the mean as the third MA mix dimension. Because of the explained distributions mathematical word problems from the first cluster, we have classified as low level dimensional structured (LL). Mathematical word problems from the second cluster we have classified as slightly level dimensional structured (SL).

Classification of word problems according to the four-dimensional solution

The clusters of the four-dimensional solution have also two clusters (one and three) with a similar distribution of frequencies but with positive (one) and negative (three) orientation. Additionally, to those clusters, the second cluster have for all dimensions a low differences to the mean. The first cluster have a big difference to the mean for the first ST mix dimension. Also, an average difference has the second MAL mix dimension. The third AEL-NM mix dimension just reach a slightly high difference to the mean. For the second cluster the first ST mix dimension reach a negative quarter differences to the mean and almost to a quarter negative differences for the third AEL-NM mix dimension. The other dimensions are average characteristic for the second cluster. The third cluster are differing from the other clusters with the negative distribution of frequency to the mean for the first ST mix dimension, with an average to big negative difference to the mean, and the second MAL mix dimension, with slightly over average negative difference to the mean. Because of that result, mathematical word problems from the first cluster we have classified as high ST MAL dimensional structured (HSM). Word problems from the second cluster we have classified as medium dimensional structures (MED). We classified mathematical word problems form the third cluster, as low ST MAL dimensional structured (LSM).

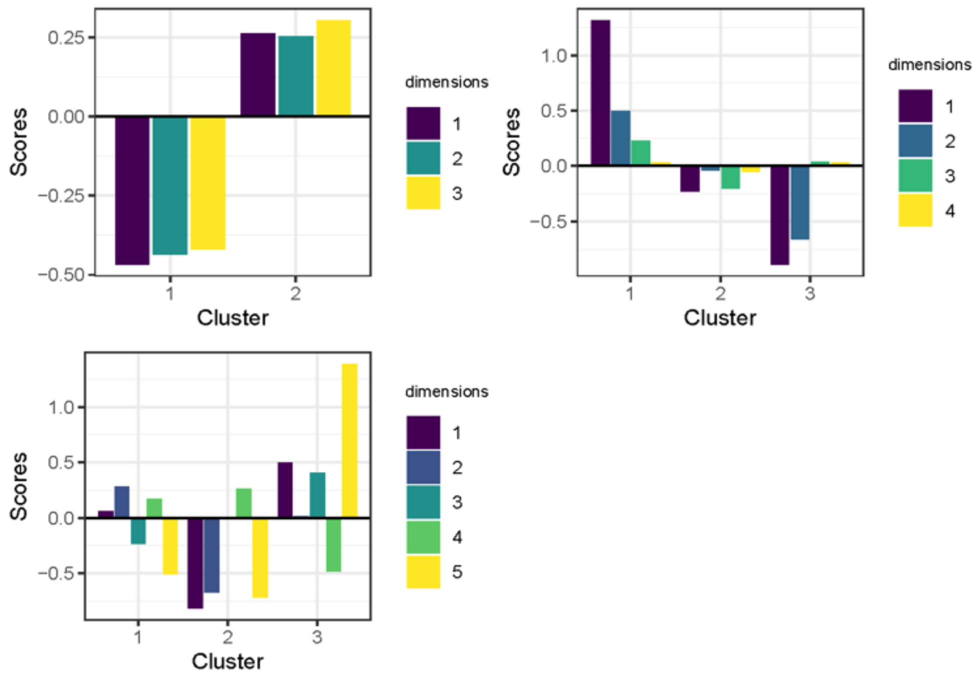

Figure 4. Frequency distribution of the dimensions for each cluster. On the upper left corner for the three dimensional solution. On the upper right corner for the four dimensional solution. On lower left corner for the five dimensional solution. The distribution shows the z-scored frequencies of the dimensional structure (Source: Authors' own illustration)

Table 3. Summary of the classification of the word problems according to the dimensional structures

Three-dimensional solution	Four-dimensional solution	Five-dimensional solution
Low level dimensional structured word problems (LL)	High ST MAL dimensional structured (HSM)	Low average MA dimensional structure (LAM)
Slightly level dimensional structured (SL)	Medium dimensional structures (MED)	Low ST, MAL, & MA dimensional structure (SMM)
	Low ST MAL dimensional structured (LSM)	High MA mix to average ST and third AEL NM mix dimensional structure with low average fourth AEL NM mix dimensional structure (HMASTA+LA)

Classification of word problems according to the five-dimensional solution

The results of the cluster analysis for the five-dimensional solutions shows a diverse explanation of the frequency distribution of the dimensional structure for the word problems. The first cluster have slightly high, for the fourth AEL-NM mix dimension, to average positive differences to the mean for the second MAL mix dimension. Because the AEL-NM occurs for the five-dimensional solution twice, in different manifestation of linguistic features, in the third and fourth dimension, it has also had a slightly negative differences to the mean. Furthermore, an average negative difference to the mean has the fifth MA mix dimension. The second cluster have average to a big negative difference to the mean for the first ST mix dimension, for the second MAL mix dimension, and for the fifth MA mix dimension. For the positive difference only the fourth AEL-NM mix dimension have a slightly high positive difference to the mean. The third cluster have average positive differences to the mean for the first ST mix dimension and the third AEL-NM mix dimension. Moreover, the cluster have a big positive differences to the mean for the fifth MA mix dimension. The fourth AEL NM mix dimension reached average negative differences to the mean. Because of these results of the dimensional structures word problems that can be located in the first cluster we have classified as low average MA mix dimensional structured (LAM). Furthermore, we have classified the second cluster as low ST, MAL, and MA mix dimensional structured (SMM).

In the third cluster MA mix dimensions highly occurs. Also, ST and the third AEL-NM mix dimensions are common dimensional structures for this cluster. The fourth AEL-NM mix dimensions occurs in lower average frequencies for this cluster. Because of the results of the distributed frequencies of the dimensional structures the word problems in this cluster are classified as HMASTA+LA mathematical word problems (**Table 3**).

Possible Implication for the Linguistic Complexity of Mathematical Word Problems

Finally, the classified mathematical word problems for the different dimensional solution should be the basis for implication for the linguistic complexity of mathematical word problems. According to the literature academic and mathematics language feature should increase the linguistic complexity of mathematical word problems (Abedi, 2006; Schleppegrell, 2012; Schmitt et al., 2011; Snow & Uccelli, 2009). The only linguistic structures that could reduce the linguistic complexity are structures with linguistic features from the everyday language (Cummins, 2017; Gogolin & Lange, 2011). However, for this study linguistic structures with only features from the everyday language cannot been observed, because of that the conclusion can be made, that all linguistic structure in this study potentially increase the linguistic complexity.

Table 4. Summary of the result of the possible implication of linguistic complexity of the classified mathematical word problems

	Classified mathematical word problems	Plausible linguistic complexity
Three-dimensional solution	LL	Low
	SL	Slightly/medium
Four-dimensional solution	LSM	Low
	MED	Medium
	HSM	High
Five-dimensional solution	LAM	Low
	SMM	Low
	HMASTA+LA	High

For word problems with three-dimensional structure

The three-dimensional solution is characterized by the ST, MAL, and MA mix dimensions. The classified LL mathematical word problems are a group of mathematical word problems with overall a low frequencies of all dimensions. That means that in relationship to the whole data set of mathematical word problems, for these kinds of mathematical word problems the ST, MAL, and MA mix dimensions occur less frequently for the three-dimensional solution. Because of that, it is plausible that LL mathematical word problems have a lower linguistic complexity as SL word problems. In contrast to LL mathematical word problems SL mathematical word problems have a slightly higher occurrence of the ST, MAL, and MA mix dimensional structures. Therefore, is likely that SL mathematical word problems have a higher linguistic complexity than for the LL word problems. However, because of the only slightly higher occurrence, the linguistic complexity should be only slightly/medium intense.

For word problems with four-dimensional structure

For the four-dimensional solution the dimensions ST, MAL, AEL-NM, and AELR mix dimension are typically. The HSM mathematical word problems have a high occurrence of ST and MAL mix dimensions and because of that they should have a higher linguistic complexity than the other classified word problems. MED mathematical word problems don't have a noticeable distributed frequency over zero, so it is likely that they have a middle range linguistic complexity. LSM mathematical word problems have a low occurrence of ST and MAL mix dimensions, because of that it is plausible that they have a lower linguistic complexity in contrast to the other groups for this solution.

For word problems with five-dimensional structure

The dimensional structure of the five-dimensional solution is characterized by ST, MAL, AEL-NM, and MA mix dimensions. For LAM mathematical word problems, it exists a lower average occurrence of MA mix dimensions. Furthermore, the third AEL-NM mix dimensions also occurs slightly lower. LAM mathematical word problems have slightly higher occurrence of MAL mix dimensions. In summary it is most likely that LAM mathematical word problems could have a lower linguistic complexity, whereby it should be noted that LAM word problems have a slightly higher occurrence of MAL mix dimensions which can have an influence on the linguistic complexity of LAM word problems. For SMM mathematical word problems ST, MAL, and MA mix dimension occurs in average to highly lower as in other mathematical word problems. Also, the fourth AEL-NM mix dimension have a slightly higher occurrence in this mathematical word problems. Because of the occurrence of the dimensions, it is mostly plausible that this mathematical word problems have a lower linguistic complexity. However, it should be noted that SMM mathematical word problems have a slightly higher occurrence of the fourth AEL-NM mix dimension. HMASTA+LA word problems have a high occurrence of MA mix dimension and an average occurrence of ST and the third AEL-NM mix dimensions. Therefore, it is plausible to assume that the HMASTA+LA mathematical word problems have a high linguistic complexity. Although it is notable that HMASTA+LA mathematical word problems have a lower average occurrence of the fourth AEL-NM mix dimensions which can have an impact on the linguistic complexity of this word problems. **Table 4** shows the summarized results of the possible implication of the linguistic complexity of the classified mathematical word problems.

DISCUSSION

This study focuses on the exploratory analysis of the linguistic complexity of mathematical word problems. Different studies shows that the linguistic features have an influence on the solution of mathematical word problems (Abedi et al., 2020; Martiniello, 2009; Plath & Leiss, 2018). Especially linguistic features that primarily are assigned to the academic language or mathematical language (in generally content dependent linguistic features) causes possible language obstacles and increasing the linguistic complexity of a text (Schleppegrell, 2005; Snow & Uccelli, 2009). The approach of this study focuses on unsupervised ML techniques to explore latent linguistic dimensions and to classify mathematical word problems according to the frequencies of the dimensional structures.

The results of the dimensional reduction procedure indicated that three- and five-dimensional solution should have the highest explanatory power. To contrast the dimensional structure more precisely, we have considered the four-dimensional structure too. The classification of mathematical word problems with the linguistic dimensions shows for the solution's different groups of word problems with differential plausible linguistic complexity. For the three-dimensional solution two groups of mathematical problem can be differentiated. The first group *LL mathematical word problems* have probably a low linguistic complexity. The second group *SL mathematical word problems* have according to the results a slightly/medium linguistic complexity. The results indicate for the four-dimensional solution a distinction of three groups. Moreover, the results shows that

HSM mathematical word problems seems to have a low, the *MED mathematical word problems* a medium and the *LSM mathematical word problems* a high linguistic complexity. In respect to the five-dimensional solution three different groups can be distinguished. The *LAM* and *SMM mathematical word problems* should have a low linguistic complexity. Furthermore, the result indicates that *HMASTA+LA mathematical word problems* should have a high linguistic complexity. Following, the results of the different analysis are discussed.

In the theoretical and empirical analysis of language in educational research, researcher often considered language in dimensions. Likewise in mathematics education research language dimensions are used and are mostly differentiation on linguistic features of word, sentence, and text basis (Prediger et al., 2019). The structure of this dimensions is in the most cases for mathematics education research on a theoretical level. The results of the study extended the picture of the existing mostly theoretical patterned language dimensions, with the empirical exploration of language dimensions for mathematical word problems especially because of specificity of one domain (only mathematical word problem) of language dimensions in this study.

Compared to studies that construct the linguistic dimensions theoretically, there are significant differences with those that were constructed empirically in this study. For example, Heine et al. (2018) use three dimensions that they use as a basis of variation for many different text features. Another common classification of language dimensions involves the word, sentence, and text levels (Prediger et al., 2019). These structures are not reflected in the empirical-exploratory procedure for forming the language dimensions. Thus, the dimensions result in a mixture of the different levels. For example, for the first factor it is shown that the word level (e.g., nominalization, filler words), the sentence level (e.g., conjunctions, prepositions) and the text level (for the three-dimensional solution lexical variety) occurs. It also shows that there is no clear structuring with respect to an everyday language, academic language, and mathematical language dimension. The dimensions rather show that the different text features of everyday, academic, and mathematical language occur together in different dimensions. However, it makes a big difference for the linguistic complexity of a text whether only one academic feature occurs in a text or several, or whether the academic feature even occurs with other text features which can even simplify the linguistic complexity like for everyday language features or they occur with mathematical language features. There seems to be an intrinsic connection between text features based on functional aspects (Halliday, 2014) of the text (domain), which can only be uncovered through empirical analysis of many text examples (Ure & Ellis, 2014). Which means the analysis reveals the linguistic structures which are a result of language in use for the domain of mathematics. At the same time, this also results in the necessity to differently (re)design interventions and materials for language education in mathematics for teaching, which at the moment refer especially to the theoretically formed dimensions.

CONCLUSION

The study showed that ML techniques, using NLP as an example, can be used for educational purposes. With the help of these techniques, large text data can be analyzed, which otherwise could only be examined with the help of qualitative analyses. This usually limits the number of cases. With the analysis it was possible to empirically determine language dimensions that occur as latent linguistic structures in mathematical word problems. For educational researchers and educators, the language dimensions may have important implications for the research and design of mathematics education considering the relevance of language for mathematics learning, such as in approaches of task-based language teaching (Ellis, 2020; Lambert, 2020; Lambert & Oliver, 2020; Newton et al., 2020). To that extent, language dimensions could serve as a scaffold for students for strategy training for learning language (content related) in mathematics. Boonen et al. (2016) plea for reading comprehension skill training, the dimensional linguistic structure could be a possible opening for such a skill training in mathematics instruction. Regarding to this the language dimension of mathematical word problems and the classification have a direct implication on linguistic complexities, in that respect the knowledge of low, medium, or high linguistic complex mathematical word problems educators has the foundation for possible scaffold strategies and to structure instructional sequences. An open question for further investigation is a practice-based instrument or tool that educators can use for a language task-based teaching. A possible outcome could be an automatically analyze and specification of mathematical word problems showing which probable language complexity the word problem could have. For this purpose, machine learning methods could be again a promising application for educational purposes to administrate individual word problem task or to give individual feedback as response.

Due to the study design, there are limitations to this study that must be considered. The word problems focus different instructional situations (performance or learning situation). The simultaneous consideration word problem for different instructional situation in this study may have resulted in structural relationships between text features in the linguistic dimensions that would not occur in the same dominant way for a group of tasks (performance or learning situation). Furthermore, the instrument considered in this paper for the formation of the language dimension is domain-specific, for this reason the results are not suitable for global statements about the structure of word problems or similar tasks in other domains, e.g., for sciences.

Moreover, the study could not present any conclusions about an empirical relationship between the language dimensions and the associated linguistic complexity and task difficulty of word problems. Further evaluation studies could estimate effects of the language dimension to task difficulties of mathematical word problems. However, this paper could give a first insight into the research field of the analysis of language dimensions. In future studies, the proportion of text features and the text data could be increased (big data) by further analyses with NLP as well as ML (Balyan et al., 2020) and possibly provide even more precise information on the true structure of language dimensions of mathematical word problems.

Author contributions: All authors have sufficiently contributed to the study and agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: Authors stated that ethical review and approval was not required for the study on participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Declaration of interest: No conflict of interest is declared by authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Abedi, J. (2006). Language issues in item developemt. In S. M. Downing, & T. M. Haladyna (Ed.), *Handbook of test development* (pp. 377-398). Lawrence Erlbaum Associates.

Abedi, J., & Gándara, P. (2006). Performance of english language learners as a subgroup in large-scale assessment: Interaction of research and policy. *Educational Measurement: Issues and Practice, 25*(4), 36-46. https://doi.org/10.1111/j.1745-3992.2006.00077.x

Abedi, J., & Herman, J. (2010). Assessing english language learners' opportunity to learn mathematics: Issues and limitations. *Teachers College Record, 112*(3), 723-746. https://doi.org/10.1177/016146811011200301

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*(3), 219-234. https://doi.org/10.1207/S15324818AME1403_2

Abedi, J., Leon, S., Wolf, M. K., & Farnsworth, T. (2008). Detecting test items differentially impacting the performance of ell students. In M. K. Wolf, J. L. Herman, J. Kim, J. Abedi, S. Leon, N. Griffin, & P. L. Bachman (Eds.), *Providing validity evidence to improve the assessment of English language learners* (pp. 55-81). National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, J., Zhang, Y., Rowe, S. E., & Lee, H. (2020). Examining effectiveness and validity of accommodations for English language learners in mathematics: An evidence-based computer accommodation decision system. *Educational Measurement: Issues and Practice, 39*(4), 41-52. https://doi.org/10.1111/emip.12328

Abidi, S. M. R., Hussain, M., Xu, Y., & Zhang, W. (2018). Prediction of confusion attempting algebra homework in an intelligent tutoring system through machine learning techniques for educational sustainable development. *Sustainability, 11*(1), 105. https://doi.org/10.3390/su11010105

Aleven, V. A. W. M. M., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26*(2), 147-179. https://doi.org/10.1016/S0364-0213(02)00061-7

Alexopoulou, T., Michel, M., Murakami, A., & Meurers, D. (2017). Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing natural language processing techniques. *Language Learning, 67*(S1), 180-208. https://doi.org/10.1111/lang.12232

Almatrafi, O., Johri, A., & Rangwala, H. (2018). Needle in a haystack: Identifying learner posts that require urgent response in MOOC discussion forums. *Computers & Education, 118*, 1-9. https://doi.org/10.1016/j.compedu.2017.11.002

Amir, O., Gal, K., Yaron, D., Karabinos, M., & Belford, R. (2014). Plan recognition and visualization in exploratory learning environments. *Studies in Computational Intelligence, 524*(212), 289-327. https://doi.org/10.1007/978-3-319-02738-8_11

Balyan, R., McCarthy, K. S., & McNamara, D. S. (2020). Applying natural language processing and hierarchical machine learning approaches to text difficulty classification. *International Journal of Artificial Intelligence in Education, 30*(3), 337-370. https://doi.org/10.1007/s40593-020-00201-7

Bednorz, D. (2021). *Sprachliche Variationen von mathematischen Textaufgaben. Entwicklung eines Instruments zur Textanpassung für Textaufgaben im Mathematikunterricht* [*Linguistic variations of mathematical text tasks. Development of a text adaptation tool for text tasks in mathematics education*]. Springer. https://doi.org/10.1007/978-3-658-33003-3

Biber, D. (1985). Investigaging macroscopic textual variation through multifeature/multidimensional analyses. *Linguistics, 23*, 337-360. https://doi.org/10.1515/ling.1985.23.2.337

Biber, D. (2006). *Dimensions of register variation. A cross-linguistic comparison*. Cambridge University Press.

Biber, D., & Conrad, S. (2019). *Register, genre, and style*. Cambridge University Press. https://doi.org/10.1017/9781108686136

Biber, D., & Egbert, J. (2018). *Register variation online*. Cambridge University Press. https://doi.org/10.1017/9781316388228

Biber, D., & Gray, B. (2013a). Identifying multi-dimensional patterns of variation across registers. In M. Krug, & J. Schlüter (Eds.), *Research methods in language variation and change* (pp. 402-420). Cambridge University Press. https://doi.org/10.1017/CBO9780511792519.026

Biber, D., & Gray, B. (2013b). Nominalizing the verb phrase in academic science writing. In B. Aarts, J. Close, G. Leech, & S. Wallis (Eds.), *The verb phrase in English. Investigating recent language change with corpora* (pp. 99-132). Cambridge University Press. https://doi.org/10.1017/CBO9781139060998.006

Biber, D., & Gray, B. (2016). *Grammatical complexity in academic english. Linguistic change in writing*. Cambridge University Press. https://doi.org/10.1017/CBO9780511920776

Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition, 24*(2), 199-208. https://doi.org/10.1017/S0272263102002048

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics. Investigating language structure and use*. Cambridge University Press. https://doi.org/10.1017/CBO9780511804489

Biber, D., Egbert, J., Gray, B., Oppliger, R., & Szmrecsanyi, B. (2016). Variationist versus text-linguistic approaches to grammatical change in English: Nominal modifiers of head nouns. In M. Kytö, & P. Patha (Eds.), *The Cambridge handbook of English historical linguistics* (pp. 351-375). Cambridge University Press. https://doi.org/10.1017/CBO9781139600231.022

Boonen, A. J. H., de Koning, B. B., Jolles, J., & van der Schoot, M. (2016). Word problem solving in contemporary math education: A plea for reading comprehension skills training. *Frontiers in Psychology, 7*(February), 1-10. https://doi.org/10.3389/fpsyg.2016.00191

Boonen, A. J. H., van der Schoot, M., van Wesel, F., de Vries, M. H., & Jolles, J. (2013). What underlies successful word problem solving? A path analysis in sixth grade students. *Contemporary Educational Psychology, 38*(3), 271-279. https://doi.org/10.1016/j.cedpsych.2013.05.001

Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, 1-20. https://doi.org/10.1016/j.caeai.2020.100002

Conati, C., & Merten, C. (2007). Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation. *Knowledge-Based Systems, 20*(6), 557-574. https://doi.org/10.1016/j.knosys.2007.04.010

Conrad, S. (2015). Register variation. In D. Biber, & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 309-329). Cambridge University Press. https://doi.org/10.1017/CBO9781139764377.018

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology, 20*, 405-438. https://doi.org/10.1016/0010-0285(88)90011-4

Cummins, J. (2017). BICS and CALP: Empirical and theoretical status of the distinction. In B. V Street, & S. May (Eds.), *Literacies and language education* (pp. 59-72). Springer. https://doi.org/10.1007/978-3-319-02252-9_6

Daroczy, G., Meurers, D., Heller, J., Wolska, M., & Nürk, H. C. (2020). The interaction of linguistic and arithmetic factors affects adult performance on arithmetic word problems. *Cognitive Processing, 21*(1), 105-125. https://doi.org/10.1007/s10339-019-00948-5

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H. C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology, 6*, 348. https://doi.org/10.3389/fpsyg.2015.00348

Dittmar, M., Schmellentin, C., Gilg, E., Dittmar, M., Schmellentin, C., & Gilg, E. (2017). Kohärenzaufbau aus Text-Bild-Gefügen: Wissenserwerb mit schulischen Fachtexten [Building coherence from text-image structures: Acquiring knowledge with specialist school texts]. *Leseforum.ch*. https://www.forumlettura.ch/sysModules/obxLeseforum/Artikel/579/2017_1_Dittmar_et%20al.pdf

Ellis, R. (2020). Teacher-preparation for task-based language teaching. In C. Lambert, & R. Oliver (Eds.), *Using tasks in second language teaching: Practice in diverse contexts* (pp. 99-120). Multilingual Matters.

Gentleman, R., & Carey, V. J. (2008). Unsupervised machine learning. In W. Huber, D. Scholtens, F. Hahne, & A. von Heydebreck (Eds.), *Bioconductor case studies* (pp. 137-157). https://doi.org/10.1007/978-0-387-77240-0_7

Gobert, J. D., & Sao Pedro, M. A. (2016). Digital assessment environments for scientific inquiry practices. In A. A. Rupp, & J. P. Leighton (Eds.), *The handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 508-534). Wiley-Blackwell. https://doi.org/10.1002/9781118956588.ch21

Gogolin, I., & Lange, I. (2011). Bildungssprache und Durchgängige Sprachbildung [Academic language and end-to-end language education]. In S. Fürstenau, & M. Gomolla (Eds.), *Migration und schulischer Wandel: Mehrsprachigkeit* [*Migration and school change: Multilingualism*] (pp. 107-127). VS Verlag. https://doi.org/10.1007/978-3-531-92659-9_6

Haag, N., Heppt, B., Stanat, P., & Kuhl, P. (2013). Second language learners` performance in mathematics: Disentangling the effects of academic language features. *Learning and Instruction, 28*, 24-34. https://doi.org/10.1016/j.learninstruc.2013.04.001

Haag, N., Roppelt, A., & Heppt, B. (2015). Effects of mathematics items' language demands for language minority students: Do they differ between grades? *Learning and Individual Differences, 42*, 70-76. https://doi.org/10.1016/j.lindif.2015.08.010

Halliday, M. A. K. (1993). On the language of physical science. In M. A. K. Halliday, & J. R. Martin (Eds.), *Writing science: Literacy and discursive power* (pp. 43-51). Routledge.

Halliday, M. A. K. (2004). *The language of science*. Continuum.

Halliday, M. A. K. (2005). *Computational and quantitative studies*. Continuum.

Halliday, M. A. K. (2007). The notion of context in language education. In J. Webster (Ed.), *Language and education. The collected works of M. A. K. Halliday* (pp. 269-290). Continuum.

Halliday, M. A. K. (2014). *Halliday's introduction to functional grammar*. Routledge. https://doi.org/10.4324/9780203783771

Heine, L., Domenech, M., Otto, L., Neumann, A., Krelle, M., Leiss, D., Höttecke, D., Ehmke, T., & Schwippert, K. (2018). Modellierung sprachlicher Anforderungen in Testaufgaben verschiedener Unterrichtsfächer: Theoretische und empirische Grundlagen [Modeling linguistic requirements in test items of different subjects: Theoretical and empirical foundations]. *Zeitschrift für Angewandte Linguistik* [*Journal of Applied Linguistics*], *69*, 69-96. https://doi.org/10.1515/zfal-2018-0017

Heppt, B., Böhme, K., & Stanat, P. (2014). The role of academic-language features for reading comprehension of language-minority students and students from low-ses families. *Reading Research Quarterly, 50*(1), 61-82. https://doi.org/10.1002/rrq.83

Heppt, B., Volodina, A., Eglinsky, J., Stanat, P., & Weinert, S. (2020). Faktorielle und kriteriale Validität von BiSpra 2-4 [Factorial and critical validity of BiSpra 2-4]. *Diagnostica* [*Diagnostics*], *67*(2), 24-35. https://doi.org/10.1026/0012-1924/a000259

Hofstetter, C. (2003). Contextual and mathematics accommodation test effects for English-language learners. *Applied Measurement in Education, 16*, 159-188. https://doi.org/10.1207/S15324818AME1602_4

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher-ai complementarity. *Journal of Learning Analytics, 6*(2), 27-52. https://doi.org/10.18608/jla.2019.62.3

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255-260. https://doi.org/10.1126/science.aaa8415

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31-36. https://doi.org/10.1007/BF02291575

Kieffer, M. J., Lesaux, N. K., Rivera, M., & Francis, D. J. (2009). Accommodations for english language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research, 79*(3), 1168-1201. https://doi.org/10.3102/0034654309332490

Kiplinger, V. L., Haug, C. A, & Abedi, J. (2000). Measuring math–not reading–on a math assessment: A language accommodations study of english language learners and other special populations. In *Proceedings of the Annual meeting of the American Educational Research Association*.

Klix, F. (1995). Stabilität und Wandlungen in geistigen Dispositionen des Menschen [Stability and transformations in human mental dispositions]. In *Proceedings of the Leibniz-Sozietät* (pp. 5-40).

Lambert, C. (2020). Frameworks for using tasks in second language instruction. In C. Lambert, & R. Oliver (Eds.), *Using tasks in second language teaching: Practice in diverse contexts* (pp. 13-30). Multilingual Matters. https://doi.org/10.21832/LAMBER9448

Lambert, C., & Oliver, R. (2020). Introduction: Tasks in context. In C. Lambert, & R. Oliver (Eds.), *Using tasks in second language teaching: Practice in diverse contexts* (pp. 1-10). Multilingual Matters. https://doi.org/10.7202/1069422ar

Leiss, D., Domenech, M., Ehmke, T., & Schwippert, K. (2017). Schwer–schwierig–diffizil: Zum Einfluss sprachlicher Komplexität von Aufgaben auf fachliche Leistungen in der Sekundarstufe I [Challenging–difficult–hard: On the influence of linguistic complexity of tasks on subject-related performance in lower secondary education]. In D. Leiss, M. Hagena, A. Neumann, & K. Schwippert (Eds.), *Mathematik und Sprache. Empirischer Forschungsstand und unterrichtliche Herausforderungen* [Mathematics and language. Empirical state of research and teaching challenges] (pp. 99-125). Waxmann.

Lemke, J. L. (2012). Technical discourse and technocratic ideology. In M. A. K. Halliday, J. Gibbons, & H. Nicholas (Eds.), *Learning, keeping, and using language* (pp. 435-460). John Benjamins Publishing Company. https://doi.org/10.1075/z.lkul2.31lem

Lin, F.-R., Hsieh, L.-S., & Chuang, F.-T. (2009). Discovering genres of online discussion threads via text mining. *Computers & Education, 52*(2), 481-495. https://doi.org/10.1016/j.compedu.2008.10.005

Lukin, A., Moore, A., Herke, M., Wegener, R., & Wu, C. (2011). Halliday's model of register revisited and explored. *Linguistics and the Human Sciences, 4*, 187-213. https://doi.org/10.1558/lhs.v4i2.187

Maier, H., & Schweiger, F. (1999). *Mathematik und Sprache. Zum Verstehen und Verwenden von Fachsprache im Mathematikunterricht* [*Mathematics and language. On the understanding and use of technical language in mathematics teaching*]. obv und hpt.

Martiniello, M. (2008). Language and the performance of english-language learners in math word problems. *Harvard Educational Review, 78*(2), 333-368. https://doi.org/10.17763/haer.78.2.70783570r1111t32

Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for English language learners in math tests. *Educational Assessment, 14*(3-4), 160-179. https://doi.org/10.1080/10627190903422906

McEnery, T., & Hardie, A. (2012). *Corpus linguistics*. Cambridge University Press. https://doi.org/10.1093/oxfordhb/9780199276349.013.0024

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes, 22*(3), 247-288. https://doi.org/10.1080/01638539609544975

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*(1), 1-43. https://doi.org/10.1207/s1532690xci1401_1

McNamara, D. S., Ozuru, Y., & Floyd, R. G. (2011). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International Electronic Journal of Elementary Education, 4*(1), 229-257.

Michalke, M. (2018). *Package korpus*. https://reaktanz.de/?c=hacking&s=koRpus

Morek, M., & Heller, V. (2012). Bildungssprache–Kommunikative, epistemische, soziale und interaktive Aspekte ihres Gebrauchs [Academic language–communicative, epistemic, social, and interactive aspects of its use]. *Zeitschrift für angewandte Linguistik* [*Journal of Applied Linguistics*], *57*(1), 67-101. https://doi.org/10.1515/zfal-2012-0011

Mousavinasab, E., Zarifsanaiey, N., R. Niakan Kalhori, S., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments, 29*(1), 142-163. https://doi.org/10.1080/10494820.2018.1558257

Murayama, K., Pekrun, R., & Vom Hofe, R. (2013). Predicting long-term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child Development, 84*(4), 1475-1490. https://doi.org/10.1111/cdev.12036

Nelson, L. K. (2020). Computational grounded theory: A methodological framework. *Sociological Methods and Research, 49*(1), 3-42. https://doi.org/10.1177/0049124117729703

Nelson, L. K., Burk, D., Knudsen, M., & McCall, L. (2018). The future of coding: A comparison of hand-coding and three types of computer-assisted text analysis methods. *Sociological Methods and Research, 50*(1), 202-237. https://doi.org/10.1177/0049124118769114

Newton, J., Le, T., & Bui, D. (2020). Low-proficiency learners and task-based language teaching. In C. Lambert, & R. Oliver (Eds.), *Using tasks in second language teaching: Practice in diverse contexts* (pp. 33-48).

Ortmann, K., & Dipper, S. (2019). Variation between different discourse types: Literate vs. oral. In *Proceedings of the NAACL-Workshop on NLP for Similar Languages, Varieties and Dialects* (pp. 64-79). https://doi.org/10.18653/v1/W19-1407

Plath, J., & Leiss, D. (2018). The impact of linguistic complexity on the solution of mathematical modelling tasks. *ZDM-Mathematics Education, 50*, 159-171. https://doi.org/10.1007/s11858-017-0897-x

Pongsakdi, N., Kajamies, A., Veermans, K., Lertola, K., Vauras, M., & Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *ZDM-Mathematics Education, 52*(1), 33-44. https://doi.org/10.1007/s11858-019-01118-9

Prediger, S., Erath, K., & Opitz, E. M. (2019). The language dimension of mathematical difficulties. In A. Fritz, V. Geraldi Haase, & P. Päsänen (Eds.), *International handbook of mathematical learning difficulties* (pp. 437-455). Springer. https://doi.org/10.1007/978-3-319-97148-3_27

Reid, T. B. W. (1956). Linguistics, structuralism, and philology. *Archivum Linguisticum*.

Revelle, W. (2021). *Procedures for personality and psychological research*. https://cran.r-project.org/package=psych

Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedeure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research, 14*(4), 403-414. https://doi.org/10.1207/s15327906mbr1404_2

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 26*(2), 582-599. https://doi.org/10.1007/s40593-016-0110-3

Rosenberg, J. M., & Krist, C. (2021). Combining machine learning and qualitative methods to elaborate students' ideas about the generality of their model-based explanations. *Journal of Science Education and Technology, 30*(2), 255-267. https://doi.org/10.1007/s10956-020-09862-4

Schleppegrell, M. J. (2005). The challanges of academic language in school subjects. In I. Lindberg, & K. Sandwall (Eds.), *Spraket och kunskappen–att lara pa sitt andrasprak i skola och hogskola* [*Language and knowledge–learning in your second language at school and university*] (pp. 47-70). The Institute for Swedish as a Second Language.

Schleppegrell, M. J. (2006). The linguistic features of advanced language use: The grammar of exposition. In H. Byrnes (Ed.), *Advanced language learning. The contribution of Halliday and Vygotsky* (pp. 134-146). Continuum.

Schleppegrell, M. J. (2012). Academic language in teaching and learning. *The Elementary School Journal, 112*(3), 409-418. https://doi.org/10.1086/663297

Schleppegrell, M. J., & O'Halloran, C. (2011). Teaching academic langauge in L2 secondary settings. *Annual Review of Applied Linguistics, 31*, 3-18. https://doi.org/10.1017/S0267190511000067

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal, 95*(1), 26-43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Scrucca, L., Fop, M., Murphy, T., B., & Raftery, Adrian, E. (2016). Mclust 5: Clustering, classification and density estimation using gaussian finite mixture models. *The R Journal, 8*(1), 289-317. https://doi.org/10.32614/RJ-2016-021

Snow, C. E., & Uccelli, P. (2009). The challange of academic language. In D. R. Olson, & N. Torrance (Eds.), *The Cambridge handbook of literacy* (S. 112-133). Cambridge University Press. https://doi.org/10.1017/cbo9780511609664.008

Tomasello, M. (2008). Acquiering linguistic constructions. In W. Damon, & R. M. Lerner (Eds.), *Child and adolescent devolpment. An advanced course* (pp. 263-297). Wiley. https://doi.org/10.1002/9780470147658.chpsy0206

Ufer, S., & Bochnik, K. (2020). The role of general and subject-specific language skills when learning mathematics in elementary school. *Journal für Mathematik-Didaktik* [*Journal for Mathematics Didactics*]*, 48*, 81-117. https://doi.org/10.1007/s13138-020-00160-5

Ure, J., & Ellis, J. (2014). Register in descriptive linguistics and linguistics sociology. In O. von Uribe-Villegas (Ed.), *Issues in sociolinguistics* (pp. 197-243). De Gruyter. https://doi.org/10.1515/9783110806687.197

Verschaffel, L., Schukajlow, S., Star, J., & van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM-Mathematics Education, 52*(1), 1-16. https://doi.org/10.1007/s11858-020-01130-4

Vilenius-Tuohimaa, P. M., Aunola, K., & Nurmi, J. E. (2008). The association between mathematical word problems and reading comprehension. *Educational Psychology, 28*(4), 409-426. https://doi.org/10.1080/01443410701708228

Vom Hofe, R., Pekrun, R., Kleine, M., & Goetz, T. (2002). Projekt zur Analyse der Leistungsentwicklung in Mathematik (PALMA): Konstruktion des Regensburger Mathematikleistungstests für 5. bis 10 Klassen [Project for the analysis of performance development in mathematics (PALMA): Construction of the Regensburg mathematics achievement test for 5th to 10th grades]. *Zeitschrift für Pädagogik* [*Journal for Pedagogy*]*, 45*, 83-100. https://doi.org/10.25656/01:3940

Weiss, Z., & Meurers, D. (2019). Analyzing linguistic complexity and accuracy in academic language development of german across elementary and secondary school. In *Processings of the 14th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 380-393). Association for Computational Lingustics. https://doi.org/10.18653/v1/w19-4440

Wu, J. Y., Hsiao, Y. C., & Nian, M. W. (2020). Using supervised machine learning on large-scale online forums to classify course-related Facebook messages in predicting learning achievement within the personal learning environment. *Interactive Learning Environments, 28*(1), 65-80. https://doi.org/10.1080/10494820.2018.1515085