The Language of Learner Proof Discourse: A Corpus Study on the Variety of Linguistic Forms

Magdalena Wolska[1]

Abstract. The paper presents an analysis of linguistic diversity in learner language used in argumentative tutorial dialogues on mathematical proofs conducted in German. The analysis is based on two corpora of dialogues with a tutoring system simulated in a Wizard of Oz setup. The purpose of the analysis is to inform and motivate the choice of computational input processing methodology for an intelligent tutoring system for proofs. After lexical normalisation of mathematical domain-specific vocabulary, learner utterances are classified with respect to, first, linguistic "modality" (natural language vs. symbolic notation) and second, their dialogue function. Proof-contributing utterances are further classified with respect to their function in the proof under construction (proof steps, declarations of proof strategy to be adopted, etc.) and the type of content verbalised in natural language (logical connectives only, domain-specific vocabulary, etc.) Linguistic diversity is quantified in terms of type-token ratios over the normalised linguistic patterns, frequency spectra, and pattern-vocabulary growth curves. The analysis shows that even this seemingly linguistically predictable argumentative domain of mathematical proofs is characterised by a large variety of linguistic patterns of expression along all the above dimensions and by a large number of idiosyncratic verbalisations. Interesting is, moreover, a conversational character of the non-proof-contributing utterances, suggesting learners' informal attitude towards the computer-based dialogues and high expectations on the input interpretation resources. This calls for a combination of shallow and deep semantic processing methods for the discourse in question: shallow pattern-based approaches for contributions which do not add to the proof and deep lexicalised grammars for the proof-relevant content, in order to optimise coverage.

1 MOTIVATION

Mathematical proof can be considered *the* argumentative discourse par excellence: premises must be stated, claims justified, hypotheses discharged, only valid rules of inference followed. Narrative flair is of secondary importance; rigorous argumentation in mathematical proofs is characterised by a highly stylised language which combines formal symbolic expressions and worded natural language structures.

While proofs are central to mathematics, learners often lack skill in constructing proofs or even lack understanding of the need for proof in the first place [12, 1, 28]. Since proofs cannot be learnt other than by practice, the idea of building automated proof tutoring systems is appealing. Indeed, a number of mathematical assistance systems have been adapted for teaching proofs [30, 7, 17]. These systems, however, rely on controlled template-like input of proof structuring language and a formal language for mathematical expressions. EXCHECK [24] was a notable example of a system in which learners could use some natural language, however, its successor, EPGY TPE [25] uses menu-based input and a formula editor. This tendency toward controlled formal input as an interaction mode goes against findings on cognitive difficulties experienced by students while learning to do formal mathematics which show that the formal language and notation are among the major obstacles in proofs [26].

Support for open-ended natural language in a proof tutoring system requires that the language understanding component be capable of translating the learners' input into a symbolic representation required by a deduction system responsible for reasoning. With the view to provisioning such input processing capabilities we collected corpora of learner proofs constructed in a flexible natural language interaction (in German) with an anticipated dialogue-based tutoring system, simulated by a human. In this paper we present an analysis of linguistic diversity of the language the learners used in the course of the interaction. The purpose of the analysis is to inform and motivate the choice of computational input processing methodology for an intelligent tutoring system for proofs.

Outline The paper is organised as follows: In Section 2 the proof corpora are briefly presented. Section 3 describes data preparation: encapsulation of mathematical symbolic content, turn, utterance and word tokenisation, and textual normalisations. Section 4 presents a classification of utterance types. Section 5 presents the analyses: Linguistic diversity has been quantified in terms of (i) type-token ratios over normalised linguistic patterns along different dimensions, (ii) frequency spectra, and (iii) pattern-vocabulary growth curves. The results are discussed in Section 6.

2 PROOF TUTORING CORPORA

Our analysis of proof tutoring is based on two corpora of tutorial dialogues on mathematical theorem proving collected in Wizard of Oz experiments [19]. The domain of mathematics in the first corpus, C-I, was naive set theory and in the second corpus, C-II, binary relations.

In both experiments dialogues were conducted in German using the keyboard and a graphical user interface. The subjects were instructed to enter proof steps, rather than complete proofs at once, to encourage interaction with the system. The set theory corpus contains dialogues conducted in three experimental tutoring conditions: minimal feedback, didactic, or socratic tutoring strategy. Tutor's verbosity of the minimal feedback condition was limited, while in both other conditions as well as in the second experiment, the subjects and the tutors were unconstrained in terms of the linguistic realisation of their turns. The binary relations corpus contains dialogues conducted in two experimental study-material conditions: subjects

[1] Computational Linguistics, Saarland University, Saarbrücken, Germany, email: magda@coli.uni-saarland.de

Table 1. Basic descriptive statistics on the two corpora.

	Set theory (C-I)	Binary relations (C-II)
Proof tasks	3	4
Tutors	1	4
Subjects/Sessions	22	37
Turns	775	1906
Mean No. of turns per session (SD)	35 (12)	51 (19)
Subjects' turns (% of No. turns)	332 (43%)	927 (49%)
Mean No. of subjects' turns per session (SD)	15 (6)	25 (10)
Mode No. of attempted proofs per subject	3	2

received background reading on binary relations presented in either a verbose or a formal variant. In both experiments, the simulated systems followed strict turn-taking rules on the subject's end of the interaction: the interface did not allow the subjects to contribute a new turn until the wizard completed their turn.

The graphical user interface of the simulated system enabled button- and/or keyboard-based insertion of symbolic mathematical expressions. Unlike in the experiments described in [13] no structured editor for the symbolic expressions nor a dedicated area for mathematical formula entry was provided; mathematical expressions could be smoothly interleaved with natural language, as in mathematical narrative discourse in textbooks or scientific publications. The interactions were logged in plain ASCII format. Mathematical symbols were logged as their corresponding unicode numeric tokens (in C-I) or as their LaTeX commands (in C-II).

To illustrate the type of data under analysis here, in Figure 1 we give excerpts from both corpora which are illustrative of the type of language used.[2] C-I comprises 775 turns (332 student and 443 tutor turns, respectively), C-II has 1906 turns (927 student and 979 tutor turns). Table 1 summarises basic descriptive information on the experiments and the collected corpora. More details on the proof tutoring corpora and the experiments can be found in [35, 5].

3 PRE-PROCESSING

3.1 Pre-processing mathematical expressions

In both corpora, mathematical expressions were identified semi-automatically, using a regular-expression grammar. The grammar comprised a vocabulary of letters, mathematical symbols (unicode or LaTeX), brackets, braces, delimiters, etc. The parser's output was manually verified and corrected where necessary.[3] The quantitative analyses were conducted based on turns and utterances in which the identified mathematical expressions have been substituted with a symbolic token MATHEXPR.[4]

[2] Here and in further examples, German utterances have been translated into English preserving sense and grammatical structure as close as possible.

[3] We do not report precision results on mathematical expression identification and parsing as this is not the focus of this paper. It is assumed that an end-to-end system provides an entry method for mathematical expressions which would enable clear, possibly real-time, identification of mathematical expressions. This could be accomplished by explicitly defining "math mode" delimiters, for instance, as key combinations indicating the start and end of mathematical expression strings or as textual delimiters analogous to the $-symbols in LaTeX.

[4] As shown in [33] utterances normalised this way can be parsed using a lexicalised grammar if the information on the expression's type – term or formula – is known. With this in mind, we therefore also classify the symbolic expressions into one of the following categories: i) atomic terms: VAR, for set, relation, or individual variables, ii) non-atomic terms: TERM or _TERM_ (object-forming operation symbols appearing in isolation (as in

C-I

S1: Wenn $A \subseteq K(B)$, dann $A \cap B = \emptyset$
(en. *If* $A \subseteq K(B)$*, then* $A \cap B = \emptyset$)
...
S5: in $K(B)$ sind alle x, die nicht in B sind
(en. *in* $K(B)$ *are all* x *which are not in* B)
S6: Da $A \subseteq K(B)$ gilt, alle x, die in A sind sind auch nicht in B
(en. *Since* $A \subseteq K(B)$ *holds, all* x *which are in* A *are also* not in B)
...
S8: Dann gilt auch: Alle x, die in B sind, sind nicht in A
(en. *Then it also holds: All* x *which are in* B *are not in* A)

C-II

S1: Ich moechte zunaechst $(R \circ S)^{-1} \subseteq S^{-1} \circ R^{-1}$ beweisen
(en. *First I would like to prove* $(R \circ S)^{-1} \subseteq S^{-1} \circ R^{-1}$)
S2: Sei $(a, b) \in (R \circ S)^{-1}$
(en. *Let* $(a, b) \in (R \circ S)^{-1}$)
...
S6: Nach der Definition von $\circ$ folgt dann (a, b) ist in $S^{-1} \circ R^{-1}$
(en. *By definition of* $\circ$ *it follows then that* (a, b) *is in* $S^{-1} \circ R^{-1}$)
...
S8: Der Beweis geht genauso wie oben , da in Schritt 2 bis 6 nur Aequivalenz umformungen stattfinden
(en. *The proof goes exactly as above since in step 2 to 6 there are only equivalences*)
S9: wie kann ich jetzt weitermachen?
(en. *how can I continue now?*)
...
S11: 1. Fall: Sei $(a, b) \in R$
(en. *1. Case: Let* $(a, b) \in R$)
S12: Ich habe mich vertippt. Korrektur: Sei $(a, z) \in R$
(en. *I made a typo. Correction: Let* $(a, z) \in R$)
...
S17: Ich habe gezeigt: $(a, b) \in (R \cup S) \circ T \Rightarrow (a, b) \in R \circ T \vee (a, b) \in S \circ T$
(en. *I have shown:* $(a, b) \in (R \cup S) \circ T \Rightarrow (a, b) \in R \circ T \vee (a, b) \in S \circ T$)
...
S24: Dann existiert ein z, so dass $(a, z) \in (R \cup S)$ und $(z, b) \in T$
(en. *Then there exists an* z *such that* $(a, z) \in (R \cup S)$ *and* $(z, b) \in T$)
S25: Nach Aufgabe A gilt $(R \cup S) \circ T = (R \circ T) \cup (S \circ T)$
(en. *By Exercise A* $(R \cup S) \circ T = (R \circ T) \cup (S \circ T)$ *holds*
...
S29: Da die Mengenvereinigung kommutativ ist, koennen wir dieses in student 25 einsetzen und erhalten die Behauptung
(en. *Since set union is commutative, we can use what's in student 25 and obtain the theorem*)
S30: nach Aufgabe W und dem Beispiel-Beweis gilt ...
(en. *By Exercise W and the example proof it holds ...*
...

Figure 1. Examples of learner utterances from both corpora

Examples of utterances from Figure 1 before and after mathematical expression pre-processing are shown below:

(1) Da $A \subseteq K(B)$ gilt, alle x, die in A sind sind auch nicht in B [C-I S6]

Da $\mathrm{MATHEXPR}_{\mathrm{FORMULA}}$ gilt, alle $\mathrm{MATHEXPR}_{\mathrm{VAR}}$, die in $\mathrm{MATHEXPR}_{\mathrm{VAR}}$ sind sind auch nicht in $\mathrm{MATHEXPR}_{\mathrm{VAR}}$

(2) Nach der Definition von $\circ$ folgt dann (a, b) ist in $S^{-1} \circ R^{-1}$ [C-II S6]

Nach der Definition von $\mathrm{MATHEXPR}_{\mathrm{\_TERM\_}}$ folgt dann $\mathrm{MATHEXPR}_{\mathrm{TERM}}$ ist in $\mathrm{MATHEXPR}_{\mathrm{TERM}}$

3.2 Turn and utterance pre-processing

Turns in both corpora were sentence-tokenised based on a standard set of end-of-sentence punctuation marks. The output of the sentence tokeniser was manually verified and corrected where necessary. Word-tokenisation was performed using a standard tokeniser.

Turns were then segmented into utterances. While a sentence is typically defined as a unit of speech containing a subject and a predicate, there is no precise linguistic definition as to what constitutes an utterance. Broadly understood, an utterance is an intentional, meaningful communicative act in an interaction. An utterance may consists of a word, a phrase, or a complex sentence with embedded clauses. It may form a complete turn, but a turn may also consist of more than one utterance. For the purpose of this study the notion of an utterance was operationalised as follows:

- An utterance never spans more than one turn or one sentence;
- Multiple clauses conjoined with conjunctions ("und" (en. *and*), "oder" (en. *or*), "aber" (en. *but*), "weil" (en. *because*), "für (en. *for*), "also" (en. *so*), "wenn" (en. *if*), "als"/"wann" (en. *when*), etc.) were considered one utterance;
- Multiple clauses conjoined without conjunction words were considered separate utterances;
- "If-then" constructions, also those omitting the words "if" and "then", were considered a single utterance;
- The following non-sentential fragments, not containing a subject, were considered utterances: noun phrases, discourse markers (also inserts, such as "acha", "oh", "naja", "schoen" (en. *nice*)), colloquial subject-drop phrasings in indicative and interrogative mood, single question words and ellipted questions (for instance, "Fertig?" (en. *Done?*)), politeness phrases (such as "sorry", "Danke"), exclamatives (for instance, "Weitere Hilfe!" (en. *Further help!*)), non-sentential answers to questions, including acknowledgments ("ok", "klar" (en. *that's clear*)), yes/no answers.

Examples of tokenised multi-utterance turns from Figure 1 are shown below:[5]

(3) $\langle u \rangle$|Dann|gilt|auch|:|Alle|x|,|die|in|B|sind|,|sind|nicht|in|A|$\langle /u \rangle$ [C-I S8]

(4) $\langle u \rangle$|1.|Fall|:$\langle /u \rangle$
$\langle u \rangle$Sei|MATHEXPR|$\langle /u \rangle$ [C-II S11]

(5) $\langle u \rangle$|Ich|habe|mich|vertippt|.|$\langle /u \rangle$
$\langle u \rangle$|Korrektur|:|$\langle /u \rangle$
$\langle u \rangle$|Sei|MATHEXPR|$\langle /u \rangle$ [C-II S12]

the example utterance (2)), etc. and iii) formulas, FORMULA, for truth-valued statements, _FORMULA_ (statement-forming operators appearing in isolation), etc.

[5] | marks token boundaries. $\langle u \rangle$ and $\langle /u \rangle$ mark utterance boundaries.

3.3 Textual normalisations

Following extensive research into the properties of spoken and written discourse [10, 6], recent studies of computer-mediated communication (CMC) – or electronic discourse more generally – have shown that, much like spoken language differs from written language, the language of type-written computer-mediated communication shares some properties with spoken language, however, it also possesses textual and linguistic characteristics which are not typical for standard written language [23, 11, 18, 3]. Among those non-standard characteristics are the frequent use of abbreviations and acronyms, words and phrases written in all capitals or all lower-case, extensive use of certain punctuation marks and lack or incorrect (random) use of other punctuation (for instance, excessive use of the exclamation mark, lack of or incorrect use of commas, lack of valid end-of-sentence punctuation), and the use of emoticons. Also type-written tutorial dialogue shows qualities which are found both in spoken and written language and those of CMC. It is prone to textual ill-formedness due to the informal setting and the telegraphic nature of the linguistic production.

In order to avoid the effects of CMC-specific qualities of the learners' productions at the utterance-level, prior to the quantitative analysis learners' utterances were normalised with respect to certain writing mechanics phenomena (alternative spelling variants, capitalisation, punctuation) and with respect to the wording of common abbreviations. Moreover, lexical normalisations were performed on lexemes and phrases in order to avoid spurious diversity due to domain-specific terminology and context-specific references. Different lexical realisations of single and multi-word domain terms and conventional speech acts were substituted with symbolic tokens representing their lexical, in case of the former, or communicative, in case of the latter, types. Discourse-specific references were likewise normalised. Details of textual normalisations are summarised below.

Spelling The German umlaut diacritics were replaced with their underlying vowels and an "-e". The *eszett* ligatures were replaced with double "s". Spelling mistakes were identified and corrected using the German aspell, a Linux spell-checker, whose general dictionary has been extended with a custom dictionary of relevant domain terms (e.g. "Distributivität"/"Distributivitaet" (en. *Distributivity*));

Punctuation Repeated consecutive occurrences of the same punctuation symbols were replaced with a single occurrence ("!!!" → "!"; "...." → ".", etc.) Punctuation in abbreviations, missing or incorrect, has been normalised (e.g. "b..zw" → "bzw.", "d.h" → "d.h."). In the final analysis inter-sentential and end of sentence/utterance punctuation was ignored;

Abbreviations Upon correcting punctuation different correct and incorrect lexical variants of common abbreviations were substituted with symbolic tokens. These included, BSP for different spelling and capitalisation variants of "z.B." (en. *e.g.*), BZW for "bzw." (en. *respectively*), OBDA for "o.B.d.A." (en. *without loss of generality*), DH for "d.h." (en. *that is*), QED for "q.e.d.", ST for "s.t." (en. *such that*), OK for "ok", "oki", "Okay", etc.

Common speech acts and inserts Conventional expressions of gratitude, such as "Danke", "VIELEN DANK" and apologies, for instance, "Tut mir leid", "Sorry", "Verzeihung", were substituted with tokens THANKYOU and APOLOGY respectively. “Ja”/“Nein” responses were substituted with the token YESNO. Conversational inserts and other discourse markers such as “So”, “Na ja” were substituted with the token DISCOURSEMARKER.

Domain terms and domain-specific references Different lexical variants of nominal and adjectival domain terms which were included in the preparatory material have been mapped to a single form, DOMAINTERM. If single-word domain terms were part of a multi-word term which can be considered a named entity, the multi-word term was normalised. For instance, “DE-MORGAN-1”, “DeMorgan-1”, “DeMorgan-Regel-1”, “de morgan regel 2” all mapped to DOMAINTERM, as did “Distributivitaet von Vereinigung ueber den Durchschnitt” as a multi-word term (a name of a statement/theorem), as well as “symmetrisch” as a single-word term.

Non-deictic references to proof exercises, such as “Aufgabe W” (en. *Exercise W*), theorems provided in the preparatory material, such as “Theorem 9” or “9”, parts of proof structure, such as “Schritt 1” (en. *Step 1*), or turns in the dialogue history, such as “Student 25”[6], were mapped to the token REFERENCE.

Different conventional wordings used to signal the end of a proof, such as “quod erat demonstrandum”, “was zu zeigen war” (en. *which was to be shown*), “woraus der beweis folgt” (en. *from which the proof follows*), “Damit ist der Beweis fertig” (en. *which completes the proof*), etc., were mapped to the token corresponding to the “q.e.d.” abbreviation, QED.

Capitalisation The analyses presented in Section 5 were performed on corpus utterances normalised as above with case-insensitive matching.

Examples of utterances from Figure 1 pre-processed as outlined in this section are shown below:

(6) dann existiert ein MATHEXPR so dass MATHEXPR und MATHEXPR [C-II S24]

(7) nach REFERENCE gilt MATHEXPR [C-II S25]

(8) da DOMAINTERM DOMAINTERM ist koennen wir dieses in REFERENCE einsetzen und erhalten die Behauptung [C-II S29]

(9) nach REFERENCE und REFERENCE gilt MATHEXPR [C-II S30]

4 CLASSIFICATION OF UTTERANCE TYPES

Learner contributions in a tutoring interaction may fulfill several functions. As illustrated in the dialogue excerpts in Figure 1, learners contribute not only proof steps – complete or incomplete (C-I S5: a justification of the statement is not given), explicit or implicit (as in C-II S8: a high-level description of a set of steps is given rather than explicit proof steps) – but also other content which adds to the solution indirectly (as in C-II S1: a solution strategy to be adopted is described or C-II S11: a proof structure to follow – case distinction – is signalled) or which does not add to the solution at all (C-II S9: help is requested).

[6] References of this form are artifacts of our dialogue display interface. In the dialogue history, student turns were numbered and labelled “Student 1”, “Student 2”, etc. while tutor turns were labelled “Tutor 1”, etc.

In order to investigate linguistic diversity of learner proof discourse at a level corresponding to the different functional contribution types, we designed a typology of learner utterances based on the corpus data at hand. The present classification builds on previously proposed dialogue move taxonomies for tutorial dialogue [22, 32, 9, 4] and has been adapted specifically for the proof tutoring domain based on the analysis of our data. The classification, shown in Table 2, has a shallow hierarchical structure focusing on *Solution-contributing* content. All utterances which do not contribute solution proposals are grouped into one category, *Other*, with an extra class. *Uninterpretable* for utterances whose semantics or pragmatic intent could not be interpreted; for instance, because they were cut off mid-utterance.

The distinction between the *Solution-contributing* class and *Other* is that with *solutions* a learner is adding information to the solution he is constructing, be it by contributing an explicit or implicit solution step or steps, changing the meta-level status of the solution (for instance, stating that a new attempt at a solution will be made) or by signalling a revision or an evaluation of an already contributed solution part. The *Other* class may comprise utterances which express learner's knowledge, but only those explicitly elicited by the tutor and which do not add to the solution being constructed. Since in the scope of this paper we are mainly interested in the analysis of argumentative language of mathematical proofs and so focus on contributions with solution-relevant content, the classification of utterances which do not contribute solution steps is coarse-grained.[7]

Note that the present classification can be mapped to previously proposed classifications of dialogue actions in tutoring. For instance, the category *Proof contribution* corresponds to *Contribute domain content* in the classification proposed in [32], to *Information Exchange : Assert* in [4] and *Assertions* in [22], and comprises the categories *Solution-step* and *Solution-strategy* from [8]. Following the general scheme proposed in [9] our class of *Proof contributions* further coded in the *Novelty* dimension for steps which contribute new content (C-II S17 is a counter-example) and in the *Motivation* dimension as *Internal* or *External*, depending on whether they have been elicited by the tutor. Utterances in the *Motivation : External* category would be found, among others, in our *Answer* category.

5 QUANTITATIVE ANALYSIS OF THE LINGUISTIC FORMS

We begin the quantitative analysis with a high-level overview of the amount of natural language verbalisation in the learner language by looking at the distribution of turns and utterances formulated using mathematical symbols alone, using natural language alone, and using natural language interleaved with mathematical symbols. Following this overview, we focus on the latter two categories; that is on utterances formulated using *some* natural language. We first look at the distribution of utterance types, as defined in Section 4, in the two corpora. Then we take a closer look at the *Proof contribution* utterances, in particular at the *Proof step* category in terms of the type of content that is verbalised. We summarise the most frequently encountered linguistic forms – linguistic *verbalisation patterns* – by category, and analyse the growth of the diversity of forms with the increasing corpus size. In all analyses we consider the two corpora in isolation (C-I and C-II) and also a larger corpus consisting of the two corpora combined into one data set (C-I $\cup$ C-II).

[7] We provide the full utterance classification, including the non-solution-related categories, for the sake of completeness.

Table 2. Classification of learner utterances

Category	Description	Examples
Solution-contributing		
Proof contribution		
Proof step	Contributes a proof step or part of a proof step	“From $A \cap B = \emptyset$ follows: $A \subseteq K(B)$” “Justification: $A \subseteq (U \setminus B)$”
Proof strategy	States a solution strategy to be adopted	“I’m using the Axiom of Extensionality” “Proof by $\subseteq$ and $\supseteq$”
Proof structure	Signals solution structure	“I’m making a case distinction:” “Forward direction:”
Proof status	Signals the status of the (partial) solution	“And so one subset relation is shown.” “q.e.d.”
Meta-level		
Self-evaluation	States an evaluation of own step	“I’ve made a typo.” “Correction:”
Restart	Signals a new attempt at a proof being started	“new start” “Once again from the beginning.”
Give up	Signals abandoning the solving task	“I would like to know the solution” “I’m giving up”
Other		
Request help	Requests assistance	“I need a hint” “How is $R \circ S$ defined?” “am I on the right track?”
Answer	Provides a non-Yes/No answer to a question posed	*T: What are the possible properties of binary relations?* “symmetry” *T: What does the variable x mean?* $\langle u \rangle$”x has two meanings”$\langle /u \rangle$ $\langle u \rangle$”it occurs in two different sets”$\langle /u \rangle$
Address	Provides a non-elicited reaction to a previous contribution	“This answers my question only halfway!” “The bracket could just as well be in a different place if you ask me!”
Agree	Expresses agreement with a statement	“indeed you’re right”
Cognitive state	Expresses the state of knowledge or understanding	“i don’t know what i can do with this hint!” “I know that.”
P/E/A	Politeness/Emotion/Attitude	“Sorry!” “I will exchange you at the shop!”
Session	Expresses a meta-level session-related statement	“Actually Exercise E (as you call it) is called Exercise A here!” “how about postponing Exercise W and starting with A?”
Self talk	Expresses an unelicited comment	“The difference between $=$ und $\cap$ is questionable” “Must have something to do with the difference.”
DM	Discourse Marker	“Right...” “Good then.”
OK	Simple acknowledgment	
Yes/No	“yes” or “no” answer	

Table 3. Descriptive information on learner proof discourse in terms of content modality: symbolic (ME), natural language (NL), and natural language interleaved with symbolic expressions (ME & NL)

		C-I Unique / Total	C-II Unique / Total	C-I ∪ C-II Unique / Total
Turns		147 / 332	497 / 927	628 / 1259
	ME	2 / 153	2 / 274	2 / 427
	NL	34 / 51	134 / 162	163 / 213
	ME & NL	111 / 128	361 / 491	463 / 619
Utterances[1]		200 / 443	531 / 1118	702 / 1561
	ME	2 / 189	1 / 300	2 / 489
	NL	64 / 92	185 / 278	240 / 370
	ME & NL	134 / 162	345 / 540	460 / 702

[1] Non-empty utterances after removing punctuation (see pre-processing in Section 3; A single occurrence of an utterance consisting of a question mark alone (in C-II) is included in the NL category.

Two frequency counts are given in the descriptive statistics tables throughout the rest of this paper: "Total" denotes the number of turn/utterance instances (tokens or "vocabulary size"; where by "vocabulary" here we mean linguistic patterns). "Unique" denotes the number of *distinct* types (unique pattern types). The proportion of these two measures is known as "type-token ratio". The two raw frequencies rather than the summarised measure are provided because the number of tokens is different for each cell in the tables, so the raw counts are more informative.

Aside from the frequency distributions, we plot graphs of frequency spectra. Spectrum visualisations are typically used with word frequencies. They show a frequency distribution in terms of number of types by frequency class, where a frequency class is a set of (sets of) instances with the same number of occurrences in the data. In other words, it shows how many *distinct types* (y-axis) occur once, twice, and so on (x-axis), thus revealing the degree of skewedness of the types distribution; the earlier the tail with y around 1 starts, the more idiosyncratic types are likely to exist in the data. We use verbalisation patterns – pre-processed utterances – as units of analysis.[8]

5.1 Mathematical symbols vs. natural language

The most prominent surface characteristic of mathematical discourse is that it is the familiar mixture of symbols and natural language, the mother tongue of the author or, in case of most of scientific publications, English, which has become the de facto language of science. While, in principle, proofs can be presented using the symbolic language of mathematics alone – as in formal logic, for instance – this presentation style is not common in communicating mathematics. In fact, it has been argued that symbolic notation does not have to dominate in a proof for it to make a "better" proof [16]. There is, however, no "prescribed" presentation style other than guidelines, and even on those authors differ (see [29, 15, 20, 21], to mention just a few).

In the context of learning mathematics, mathematical notation, its mastery, has been shown to be one of the major obstacles in learning to do proofs [26]. Interestingly, the presentation style of the study-material – mainly formal vs. mainly natural language, verbose proof presentation – has an influence on the learners' use of natural language in computer-based tutoring [34]; that is, learners mimic the linguistic style they are presented with. As the first approximation of linguistic variety in learner proof discourse, we therefore analyse the

[8] R [27] was used to create for the plots and the zipfR package [14] for the frequency spectra. Only the first 15 frequency classes are shown since in all cases the frequency of the larger classes oscillated between 0 and 5.

Table 4. Distribution of utterance types

	C-I Total	C-II Total	C-I ∪ C-II Total
Solution-contributing	187	548	735
Proof contribution	180	539	719
Proof step	171	469	640
Proof strategy	4	30	34
Proof status	5	24	29
Proof structure	-	16	16
Meta-level	7	9	16
Self-evaluation	2	5	7
Restart	2	3	5
Give up	3	1	4
Other	64	267	331
Request help	16	154	170
Yes/No	18	24	42
Cognitive state	15	16	31
Politeness/Emotion/Attitude	3	21	24
Discourse marker	1	21	22
Answer	5	15	20
OK	1	6	7
Address	1	5	6
Session	-	4	4
Agree	2	1	3
Self talk	2	-	2
Uninterpretable	3	4	7

learners' contributions in terms of the two types of content modalities: natural language and symbolic expressions.

Table 3 shows the distribution of turns and utterances in both corpora with respect to natural language and symbolic content. ME denotes turns and utterances consisting of symbolic expressions alone, NL those consisting of natural language alone (as in C-II S8), and ME & NL those consisting of natural language interleaved with mathematical expressions (C-I S1 or C-II S29).

In both corpora the majority of turns and utterances contain some natural language (turns: 54% NL/ME & NL vs. 46% ME in C-I and 70% vs. 30%, respectively, in C-II; utterances: 57% NL/ME & NL vs. 43% ME in C-I and 73% and 27%, respectively, in C-II). Only 14 NL/ME & NL turn-level patterns and only 28 utterance-level patterns occur both in C-I and C-II (turn-level: 640 NL/ME & NL patterns in C-I and C-II considered in isolation vs. 626 in C-I ∪ C-II; utterances: 728 in C-I and C-II in isolation vs. 700 in C-I ∪ C-II). There is proportionally more natural language in C-II even though, as shown in [34], the participants in the formal material condition were less verbose than those in the verbose material condition.

Overall, 69% of the utterances in C-I ∪ C-II contain some linguistic material, among which there are 700 distinct utterances (verbalisation patterns). From this point on we focus on a subset of the data: we look at utterances only and only those which do contain natural language. We start by looking at the distribution of utterance types.

5.2 Distribution of utterance types

Table 4 shows the distribution of utterance types, as defined in Section 4, in both corpora.[9] The majority of utterances in both corpora are solution-contributing, 74% of all utterances in C-I and 67% in C-II, and most of them proof steps. This is not surprising of course. The second experiment involved more complex proofs requiring, for

[9] Only the utterance types with more than five occurrences will be discussed here. Utterance types with lower frequency of occurrence appear too sparse for any conclusions about their wording.

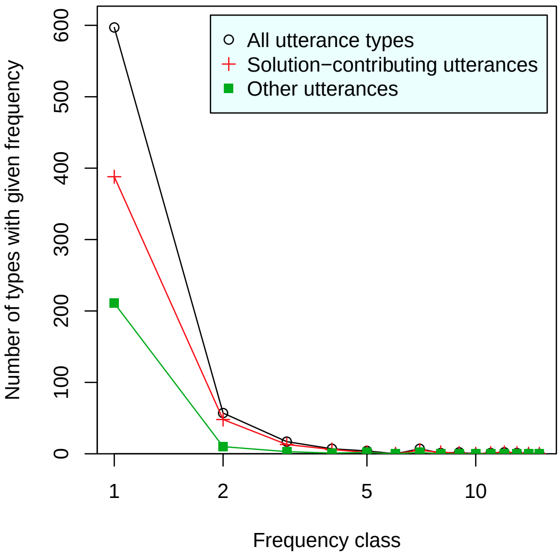

Figure 2. Frequency spectra: Utterance types (x-axis log-scaled)

instance, considering cases and proving both directions of a bi-conditional, which resulted in explicit verbalisations of the proving strategy, the proof structure, and in learners signalling that a complex proof (or its part; e.g. one direction of a bi-conditional) is completed.

Among the non-solution-contributing utterances, the largest class, 51%, are help requests of different specificity; from general requests (such as "Hilfe!" (en. *Help!*) or "Einfaches Beispiel wuerde mir weiter helfen" (en. *A simple example would help me*)) to specific requests of a definition (such as "Wie lautet die Definition der Operation ${}^{-1}$?" (en. *What's the definition of* ${}^{-1}$*?*) or "Erklaere die Definition $R \circ S$ in Worten!" (en. *Explain the definition of* $R \circ S$ *in words!*)), or enquiries whether propositions hold (such as "Ist (a, z) in R?" (en. *Is* (a, z) *in* R*?*) or "Elemente von $(R \circ S) \circ T$ sind Tripel der Form (x, y, z), oder?" (en. *Elements of* $(R \circ S) \circ T$ *are triples of the form* (x, y, z)*, right?*)) The second largest category are closed-class types, Yes/No and OK, which together make up 15% of all the non-solution-contributing utterances.

The second largest category of open-ended verbalisations are meta-cognitive statements on the state of knowledge (or, for the most part, of the *lack* of knowledge), 31 occurrences. Statements such as "Keine Ahnung mehr wie der Nachweis korrekt erbracht werden kann" (en. *No idea how the proof can be correctly produced*) or "Verstehe die definition nicht" (en. *Don't understand the definition*), can be interpreted as indirect requests of help. Interestingly, only one wording appeared more than once, "Dann weiss ich nicht weiter" (en. *So I'm lost*).

Aside from the two common variants of expressions of gratitude ("Danke"/"Vielen Dank" (en. *Thank you/Thank you very much*)) and the four common German variants of apologies ("Tut mir leid"/"Entschuldigung"/"Verzeihung"/"Sorry"), the remaining expressions of emotions and attitude (Politeness/Emotion/Attitude class) were idiosyncratic and unpredictable, and spanned both positive polarity emotions, for instance, "Das macht Spass mit Dir" (en. *It's fun with you!*) and negative polarity ("Wollen Sie mir nun Mathematik beibringen oder wollen Sie mich pruefen???" (en. *Do you want to teach me math now or do you are you giving me a test???*), "NERV!!" (en. *[anger]*)). Not surprisingly, idiosyncratic were also the occurrences of the remaining open-ended classes, answers and addresses, whose content is entirely determined be the preceding context, i.e. the tutor's contribution which triggered them.

What is interesting is that there were 22 occurrences of discourse markers, the kinds typical of spoken language: "na ja" (en. *oh well*), "oh", "hm". The variety of discourse markers suggests that computer-mediated dialogue was treated by the subjects much like natural spoken interaction, even though it was type-written.

Figure 2 shows the frequency spectra of all the utterance types and of the two major utterance classes. It is clear from the plot that the distribution of distinct verbalisations is heavily skewed. For all sets of utterance types, already the number of patterns with at least between three to five occurrences is less than 10. The tail of patterns with frequency 1 starts between 5-10 or more occurrences.

Frequency spectra also show that the data is sparse and even though some utterance types have a high frequency of occurrence (Table 4) they consist of mainly idiosyncratic linguistic patterns. Of course, most interesting from the point of view of formalisation are the core argumentative utterances which build up a proof. Thus, we now take a closer look at the verbalisations of proof contributions.

5.3 Proof contributions

Since we are interested in the diversity of wording, we first consider the type of content that proof contributions verbalise. Considering that the *ultimate* goal of this work is to computationally translate the natural language verbalisations into a formal language of a deduction system, aside from the three classes of proof-level descriptions – proof strategy, proof structure, and proof status (see Table 2) – three classes of proof steps are distinguished in the analysis that follows. The sub-categorisation of proof steps takes into account, on the one hand, the type of content the natural language expresses and, on the other hand, the type of linguistic knowledge which needs to be encoded in order for formalisation to be possible.

The simplest case for translation are steps in which natural language is used only for logical operators (connectives and binders/quantifiers), to signal proof step components, and where no discourse context nor domain-specific linguistic information is needed. By proof step components we mean elements of a deduction system's proof language such as the declarative proof script language presented in [2]. In order to formalise proof steps of this kind, the only linguistic knowledge needed is the natural language vocabulary and syntax of logical connectives and of the proof structural components (proof discourse connectives); that is, only a basic interpretation lexicon. Examples of this class of proof steps include:[10]

(10) Wenn $A \subseteq K(B)$, dann $A \cap B = \emptyset$ [C-I S1]

(11) Sei $(a, b) \in (R \circ S)^{-1}$ [C-II S2]

We will refer to this class of steps as *NL logic & proof step components* which stands for "natural language logical connectives and proof step components".

The second and third class of proof steps are those which require context and linguistic domain knowledge for interpretation and formalisation: if beyond the type of content described above, only domain concepts from the domain(s) to which the proof refers (here: set theory and binary relations) and discourse-specific references have to be translated, then the proof step belongs to the second category,

[10] The example sentences are worded here as they occur in the corpus. For the analysis, they have been pre-processed as described in Section 3.

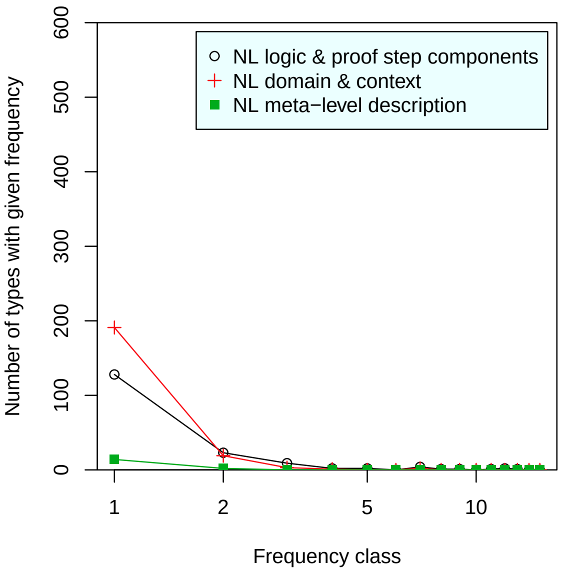

Figure 3. Frequency spectra: Proof step types (x-axis log-scaled; y-axis range extended to match Figure 2 for comparison)

to which we will refer to as *NL domain & context*. The verbalised domain concepts may be single and multi-word domain terms[11] but also informal verbalisations of domain relations, such as the locative prepositional phrase with "in" for set membership. Examples of the second class of proof steps include:

(12) in $K(B)$ sind alle x, die nicht in B sind [C-I S5]

(13) Nach der Definition von $\circ$ folgt dann (a, b) ist in $S^{-1} \circ R^{-1}$ [C-I S6]

(14) Nach Aufgabe A gilt $(R \cup S) \circ T = (R \circ T) \cup (S \circ T)$ [C-II S25]

In C-II S25 the reference "Aufgabe A" needs to be resolved. Note, however, that the utterance "Es gilt nach Definition ausserdem $S^{-1} \circ R^{-1} = (x, y) | \exists z(z \in M \wedge (x, z) \in S^{-1} \wedge (z, y) \in R^{-1})$" (en. *By the definition it moreover holds that* ...) belongs to the first class, *NL logic & proof step components*: no domain-specific vocabulary is used; the word "definition" is in the basic lexicon of mathematics.[12]

Finally, the third class comprise those steps which are not specified explicitly, but rather indirectly as high-level meta-descriptions of a (possibly complex) transformation which needs to be performed in order to reconstruct the intended step. An example of such as complex proof step is C-II S8. Other examples include:

[11] See the paragraph on normalisation of domain terms and domain-specific references in Section 3.3

[12] The verbalisation-oriented proof step classification proposed in [31], while similar to ours and designed with a similar motivation, is imprecise. First, it is not clear whether the class *simple connections* would accommodate utterances with adverbs or adverbial phrases, such as "Moreover, as previously shown, it follows that ..." Second, and more importantly, the distinction between *weakly verbalised* and *strongly verbalised* formulas is unclear based on the definitions given. *Weakly verbalised* formulas are defined as those "where some relations or quantifiers are partly verbalised", while *strongly verbalised* formulas as those "where all relations and quantifiers are fully verbalised". Based on these definitions it is not clear why the example "a is the limit of $(a_n)_{n \in N}$", given in the paper, should be classified as *weakly verbalised*, whereas "For all ϵ holds: there exists a $n_0(\epsilon) \in N$ with ..." as *strongly verbalised*; clearly, the set membership relation in $n_0(\epsilon) \in N$ is not verbalised.

Table 5. Descriptive information on proof contributions

	C-I Unique / Total	C-II Unique / Total	C-I ∪ C-II Unique / Total
Proof step	138/ 171	287 / 469	407 / 640
NL logic & proof step components	54 / 80	136 / 286	175 / 366
NL domain & context	78 / 85	140 / 171	216 / 256
NL meta-level description	6 / 6	11 / 12	16 / 18
Proof strategy	4 / 4	25 / 30	29 / 34
Proof structure	- / -	7 / 16	7 / 16
Proof status	1 / 5	7 / 24	7 / 29

(15) Analog geht der Fall, wenn $(a, z) \in S$.
(en. *The case for* $(a, z) \in S$ *is analogous*)

(16) de morgan regel 2 auf beide komplemente angewendet
(en. *de morgan rule 2 applied to both complements*)

(17) $(S \circ T)$ ist genauso definiert
(en. $(S \circ T)$ *is defined the same way*).

Complex proof steps of this kind will be referred to as *NL meta-level description*.

Table 5 shows descriptive statistics on proof contributions with proof steps sub-classified as described above. Not surprisingly, the wording of two types of proof contributions which refer to the proof-level concepts – proof strategy and proof structure – is diverse. Wording of proof status information is repetitive; indeed, most often only the end of the proof is signalled explicitly and most often using the abbreviation "q.e.d." Now, also not surprisingly, within the class of proof steps, the more complex the content, the more varied the wording. Meta-level descriptions of proofs are almost entirely idiosyncratic. Only two utterance patterns occurred more than once: "MATHEXPR ist analog definiert" (en. MATHEXPR *is defined analogously*) and "das gleiche gilt fuer MATHEXPR" (en. *The same holds for* MATHEXPR). The wording of proof steps in the *NL domain & context* category is also diverse: 92% of instances are distinct in C-I, 82% in C-II, and 84% overall. Most repetitive patterns are found in the *NL logic & proof step components* class: 67% of all utterance instances in this category are distinct in C-I, only 47% in C-II, and 48% in both corpora combined. Overall, 63% of proof steps (from the three categories) are distinct.

Figure 3 shows the frequency spectra of the three proof step categories in C-I ∪ C-II. Again, the distribution of verbalisation patterns is heavily skewed. In the largest category, *NL domain & context*, 210 out of the 216 unique patterns occur only once or twice; that is 97%. In the *NL logic & proof step components* category, around 150 out of the 175 unique patterns, 73%, occur once or twice. However, within this class there are 8 patterns with at least five instances of occurrence. Table 6 shows the top-10 most frequent linguistic patterns in the three classes of proof steps from the combined corpus, C-I ∪ C-II, with their frequency of occurrence.

5.4 Growth of the diversity of forms

Finally, we are interested in how the diversity of forms evolves with the number of conducted dialogues. Specifically, we would like to know how many dialogues are needed to have observed most of the verbalisation patterns. Figure 4 shows a plot of a variant of the type-token (vocabulary growth) curve [36]. On the x-axis is the number of dialogues seen. Rather than the raw type count, the y-axis shows the proportion of observed pattern types out of all pattern types in

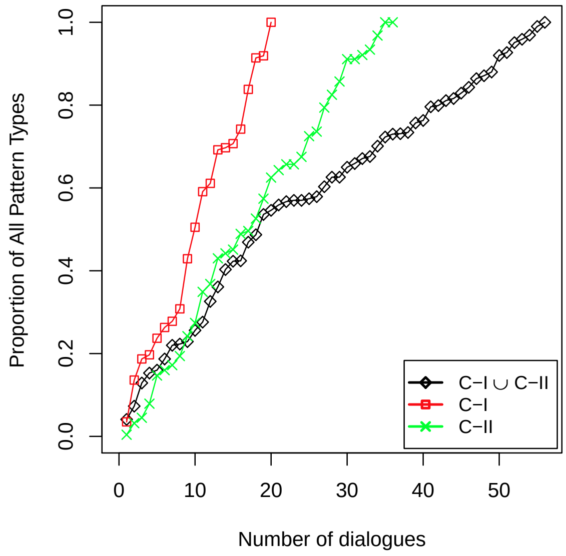

Figure 4. Growth of the utterance patterns over random dialogue sequences

Table 6. Top-10 most frequent utterance patterns expressing proof steps

Linguistic pattern	Frequency
Proof step	
NL logic & proof step components	
sei MATHEXPR	54
es gilt MATHEXPR	13
wenn MATHEXPR dann MATHEXPR	12
also MATHEXPR	12
dann ist MATHEXPR	11
also ist MATHEXPR	9
MATHEXPR und MATHEXPR	8
MATHEXPR ist dann MATHEXPR	7
daraus folgt MATHEXPR	7
daraus folgt dass MATHEXPR	7
NL domain & context	
nach REFERENCE MATHEXPR	7
DOMAINTERM	7
nach REFERENCE ist MATHEXPR	4
MATHEXPR nach REFERENCE	3
DOMAINTERM von MATHEXPR ist DOMAINTERM MATHEXPR	3
aus REFERENCE folgt MATHEXPR	3
wegen der formel fuer DOMAINTERM folgt MATHEXPR	2
oder MATHEXPR wegen DOMAINTERM von MATHEXPR	2
nach REFERENCE gilt MATHEXPR	2
nach DOMAINTERM gibt es ein MATHEXPR mit MATHEXPR	2
NL meta-level description	
MATHEXPR ist analog definiert	2
das gleiche gilt fuer MATHEXPR	2
gleiches gilt mit MATHEXPR	1
DOMAINTERM auf beide DOMAINTERM angewendet	1
der fall MATHEXPR verlaeuft analog	1
der beweis von MATHEXPR ist analog zum beweis von MATHEXPR	1
beweis geht genauso wie oben da in REFERENCE bis REFERENCE nur DOMAINTERM umformungen stattfinden	1
analog geht der fall wenn MATHEXPR	1
andersrum	1
die zweite DOMAINTERM ergibt sich aus der umkehrung aller bisherigen beweisschritte	1

the given corpus.[13] The order of dialogues in C-I and C-II has been randomised. For the C-I ∪ C-II plot, the corpora were combined and a random sequence drawn from the combined set.

What can be seen from the graphs is that the pattern vocabulary grows linearly (given the random sample drawn). The tendency is similar in both corpora: half of the patterns have been seen at about 40% of the data sets and 80% of the patterns at about 77% into the data set in C-I (ca. 17 dialogues) and 70% in C-II (ca. 26 dialogues). In the combined corpus, however, half of the patterns have been seen already about 32% into the data set. 80% of the patterns have been seen about 70% into the data set (ca. 41 dialogues).

6 DISCUSSION

First, it is clear from the results that the language of learner discourse in proofs is not as repetitive as one might expect. Learners use complex natural language utterances not only during meta-communication with the tutor, but also when contributing proof steps. 57% of all utterances in C-I and 73% in C-II contained some natural language. The fact that natural language was more often used in the C-II corpus may be explained by the fact that the binary relations proofs were more complex than the set theory proofs. However, set theory is very naturally expressed in natural language, so the reason why this was the case needs further investigation.

Second, the wording of proof steps is surprisingly diverse and the language used in the two corpora is different. The fact that there were only 28 utterance verbalisations which occurred in both data sets is surprising.[14] This low number of common patterns is reflected in the type-token plot (Figure 4) which exhibits a steady increase with only one area of slower growth in the combined corpus, about 20-25% into the randomly-ordered data set.

The difference in the linguistic diversity of the proof language (the proof contributions class) in the two corpora can be also seen in the different distributions of distinct linguistic patterns (Table 5). Among the *logic & proof step components* class, 67% of the verbalisations were distinct in C-I and 47% in C-II. In the *domain & context* class, 92% of all the verbalisations were distinct in C-I and 82% in C-II. That is, the language in C-II appears more repetitive. In both corpora, however, the language in the latter class of proof steps is more heterogeneous than in the former. The frequency spectra and the pattern growth curves show further the degree to which the language is indeed diverse. In the *logic & proof step components* class, 81% of the distinct types were single-occurrence utterances (81% in C-I and 72% in C-II). In the *domain & context* class, 90% of the types were single-occurrence (96% in C-I and 85% in C-II).

Not surprisingly, the majority of the meta-level communication are the learners' requests for assistance: requests for hints, definitions, explanations, etc. As these are not the core argumentative utterance types, we did not present a detailed analysis here, however, to roughly illustrate the diversity of wording it is enough to mention that out of the 170 help requests, 149 (88%) were distinct verbalisations. 136 of these were single-occurrence patterns. A further subclassification of help requests might reveal more homogeneity in the wording within the subcategories.

The relatively large number of discourse markers, typical of spoken interaction, suggests that participants had an informal approach

[13] 198, NL + ME & NL, utterance patterns in C-I, 530 in C-II, and 700 in C-I ∪ C-II; see Table 3.

[14] 8 were from the non-solution contributing class and 20 were proof step verbalisations, the majority from the *logic & proof step components* class.

to dialogue style and treated it much like a chat, adapting spoken language they would have otherwise used in a natural setting to the experiments' type-written modality; this is a known phenomenon [18]. The diversity of verbalisations may be partly due to this.

7 CONCLUSIONS AND FURTHER WORK

We have shown that even this seemingly linguistically predictable argumentative domain of mathematical proofs is characterised by a large variety of linguistic patterns of expression and by a large number of idiosyncratic verbalisations and that the meta-communicative part of discourse which does not directly contribute to the solution has an conversational character, suggesting learners' informal attitude towards the computer-based dialogues and high expectations on the input interpretation resources. This calls for a combination of shallow and deep semantic processing methods for the discourse in question: shallow pattern-based approaches for contributions which do not add to the proof and deep lexicalised grammars for the proof-relevant content, in order to optimise coverage. At the time of writing a parsing grammar for German we have been developing is capable of analysing all the linguistic proof-contribution structures which occurred more than twice in the data. Future work will proceed in two directions: (i) we will continue to improve the grammar coverage and (ii) we have started pre-processing proofs from a corpus of open-access scientific publications (in English) in order to perform an analogous analysis of language variety in expert proof discourse.

REFERENCES

[1] D. Almeida, 'A survey of mathematical undergraduates' interaction with proof: Some implications for mathematics education', *International Journal of Mathematical Education in Science and Technology*, **3**(6), 869–890, (2000).

[2] A. Autexier, D. Dietrich, and M. Schiller, 'Towards an Intelligent Tutor for Mathematical Proofs', in *Proceedings of the 1st Workshop on CTP Components for Educational Software*, eds., Pedro Quaresma and Ralph-Johan Back, volume 79 of *EPTCS*, pp. 1–28, (2012).

[3] N. Baron, 'Language of the Internet', in *The Stanford Handbook for Language Engineers*, ed., Ali Farghali, 59–127, Stanford, CA, (2003).

[4] L. Becker, W. Ward, S. Van Vuuren, and M. Palmer, 'DISCUSS: A dialogue move taxonomy layered over semantic representations', in *Proceedings of the 9th International Conference on Computational Semantics (IWCS-11)*, pp. 310–314, (2011).

[5] C. Benzmüller, H. Horacek, H. Lesourd, I. Kruijff-Korbayová, M. Schiller, and M. Wolska, 'A corpus of tutorial dialogs on theorem proving; the influence of the presentation of the study-material', in *Proceedings of the 5th LREC*, pp. 1766–1769.

[6] D. Biber, *Variation across speech and writing*, CUP, 1988.

[7] W. Billingsley and P. Robinson, 'Student Proof Exercises Using MathsTiles and Isabelle/HOL in an Intelligent Book', *Journal of Automated Reasoning*, **39**(2), 181–218, (2007).

[8] M. Buckley and M. Wolska, 'A classification of dialogue actions in tutorial dialogue', in *Proceedings of the 22nd International Conference on Computational Linguistics (Coling-08)*, pp. 73–80, (2008).

[9] G.C. Campbell, N.B. Steinhauser, M.O. Dzikovska, J.D. Moore, C.B. Callaway, and E. Farrow, 'DeMAND coding scheme: A "common language" for representing and analyzing student discourse', in *Proceedings of 14th AIED Conference (AIED-09)*, pp. 665–667, (2009).

[10] W. Chafe and Deborah T., 'The Relation between Written and Spoken Language', *Annual Review of Anthropology*, **16**, (1987).

[11] D. Crystal, *Language and the Internet*, CUP, 2001.

[12] T Dreyfus, 'Why johnny can't prove', *ESM*, **38**, 85–109, (1999).

[13] M. Dzikovska, D. Reitter, J. Moore, and C. Zinn, 'Data-driven Modeling of Human Tutoring in Calculus', in *Proceedings of the Language-enhanced Educational Technology Workshop*, pp. 22–28, (2006).

[14] S. Evert and M. Baroni, '*zipfR*: Word frequency distributions in R', in *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pp. 29–32, (2007). (R v. 0.6-6 of 2012-04-03).

[15] L. Gillman, *Writing Mathematics Well: A Manual for Authors*, Mathematical Association of America, 1987.

[16] Paul R. Halmos, 'How to write mathematics', *L'Énseignement Mathématique*, **16**, (1970).

[17] M. Hendriks, C. Kaliszyk, F. van Raamsdonk, and F. Wiedijk, 'Teaching logic using a state-of-art proof assistant', *Acta Didactica Napocensia*, **3**(2), 35–48, (2010).

[18] Y. Hård af Segerstad, *Use and Adaptation of Written Language to the Conditions of Computer-Mediated Communication*, Ph.D. dissertation, Göteborg University, 2002.

[19] J.F. Kelley, 'An iterative design methodology for user-friendly natural language office information applications', *ACM Transactions on Office Information Systems*, **2**(1), 26–41, (1984).

[20] D.E. Knuth, T. Larrabee, and R.M. Roberts, *Mathematical Writing*, MAA Notes, Mathematical Association of America, 1989.

[21] Steven G. Krantz, *A Primer of Mathematical Writing*, AMS, 1997.

[22] J. Marineau, P. Wiemer-Hastings, D. Harter, B. Olde, P. Chipman, A. Karnavat, V. Pomeroy, S. Rajan, and A. Graesser, 'Classification of speech acts in tutorial dialogue', in *Proceedings of the ITS-00 Workshop on Modeling Human Teaching Tactics and Strategies*, pp. 65–71, (2000).

[23] N. Maynor, 'The Language of Electronic Mail: Written Speech?', in *Centennial Usage Studies*, 48–54, (1994).

[24] J. McDonald, 'The EXCHECK CAI system', in *University-level computer-assisted instruction at Stanford: 1968–1980*, ed., Patrick Suppes, 765–790, Stanford University, Institute for Mathematical Studies in the Social Sciences, Stanford, CA, (1981).

[25] D. McMath, M. Rozenfeld, and R. Sommer, 'A computer environment for writing ordinary mathematical proofs', in *Proceedings of the 8th Conference on Logic for Programming*, pp. 507–516, (2001).

[26] R.C. Moore, 'Making the transition to the formal proof', *ESM*, **27**, 249–266, (1994).

[27] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

[28] A. Selden and J. Selden, 'Errors and misconceptions in college level theorem proving', Technical Report 2003–3, Tennessee Technological University, Cookeville, TN, (2003).

[29] N.E. Steenrod, P.R. Halmos, M.M. Schiffer, and J.A. Dieudonné, *How to write mathematics*, Americal Mathematical Society, 1981.

[30] Andrzej Trybulec and Piotr Rudnicki, 'Using Mizar in Computer Aided Instruction of Mathematics', in *Norwegian-French Conference of CAI in Mathematics*, (1993). Retrieved on May 17 from `http://www.mizar.org/project/oslo.pdf`.

[31] M. Wagner and H. Lesourd, 'Using TEX macs in Math Education: An exploratory Study', in *Proceedings of MathUI-08*, (2008). Retrieved on May 27, 2012 from `http://www.activemath.org/workshops/MathUI/08/proceedings/iMathEU.html` (Online proceedings).

[32] M. Wolska and M. Buckley, *Text Resources and Lexical Knowledge, Selected Papers from the 9th Conference on Natural Language*, volume 8 of *Text, Translation, Computational Processing*, chapter A Taxonomy of Task-related Dialogue Actions: The Cases of Tutorial and Collaborative Planning Dialogue, 105–118, Mouton De Gruyter, 2008.

[33] M. Wolska and I. Kruijff-Korbayová, 'Analysis of Mixed Natural and Symbolic Language Input in Mathematical Dialogs.', in *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL-04)*, pp. 25–32, (2004).

[34] M. Wolska and I. Kruijff-Korbayová, 'Factors influencing input styles in tutoring systems: the case of the study-material presentation format', in *Proceedings of the ECAI-06 Workshop on Language-enabled Educational Technology*, pp. 86–91, (2006).

[35] M. Wolska, I. Kruijff-Korbayová, and H. Horacek, 'Lexical-semantic interpretation of language input in mathematical dialogs', in *Proceedings of the ACL 2nd Workshop on Text Meaning and Interpretation*, pp. 81–88, (2004).

[36] G. Youmans, 'Measuring Lexical Style and Competence: The Type-Token Vocabulary Curve', *Style*, **24**, 584–0599, (1990).