British Journal of Educational Technology Vol 50 No 6 2019 3047–3063
doi:10.1111/bjet.12875

Exploring the potential of natural language processing to support microgenetic analysis of collaborative learning discussions

Florence R. Sullivan and P. Kevin Keith

Florence R. Sullivan, PhD (Cognitive Studies in Education), is a professor of Learning Technology at the University of Massachusetts, Amherst, Department of Teacher Education and Curriculum Studies. Her research interests include collaborative learning with computational media, the development of computational thinking and issues of equity in computer science-based education. P. Kevin Keith, MEd (Mathematics Education), is a professor of Computer Science and Mathematics at Landmark College, Putney, VT, USA, and a graduate student in the College of Education at the University of Massachusetts, Amherst. His dissertation investigates how emergent roles enacted by students influence collaboration in an educational robotics learning environment. Address for correspondence: Florence R. Sullivan, College of Education, University of Massachusetts, 813 North Pleasant, Amherst, MA 01003, USA. Email: fsullivan@educ.umass.edu

Abstract

In this study, we explore the potential of a natural language processing (NLP) approach to support discourse analysis of in-situ, small group learning conversations. The theoretical basis of this work derives from Bakhtin's notion of speech genres as bounded by educational robotics activity. Our goal is to leverage computational linguistics methods to advance and improve educational research methods. We used a parts-of-speech (POS) tagging program to automatically parse a transcript of spoken dialogue collected from a small group of middle school students involved in solving a robotics challenge. We grammatically parsed the dialogue at the level of the trigram. Then, through a deliberative process, we mapped the POS trigrams to our theoretically derived problem solving in computational environments coding system. Next, we developed a stacked histogram visualization to identify rich interactional segments in the data. Seven segments of the transcript were thus identified for closer analysis. Our NLP-based approach partially replicated prior findings. Here, we present the theoretical basis for the work, our analytical approach in exploring this NLP-based method, and our research findings.

In this paper, we present our exploration of the use of a natural language processing (NLP) method called parts-of-speech (POS) tagging to support discourse analysis in a collaborative learning setting. NLP methods such as POS are used widely in data science applications to help create predictive and descriptive models of behavior as embedded in textual sources. NLP is one of a plethora of computational methods that fall under the umbrella of artificial intelligence. Here, we define artificial intelligence (AI) in education as computational entities (technologies or methods) that support student learning and/or research on learning through adaptable machine intelligence.

Examples of AI approaches that support student learning include intelligent tutoring systems, personalized learning systems, agent-based software systems and systems that have a more

Practitioner Notes

What is already known about this topic

- Over the last 10 years, several educational research papers indicate that natural language processing (NLP) techniques can be used to help interpret well-structured, written dialogue, eg, conversations in online class discussions.
- Two recent papers indicate that NLP techniques can also be used to help interpret well-structured, spoken dialogue, eg, replies to interview questions and/or comments made during think aloud protocols.
- Multimodal learning analytic techniques are being used to investigate collaborative learning. These studies use non-verbal features of data (gaze, gesture, physical actions), prosodic features of verbal data (pitch and tone) and/or turn-taking and duration of talk per speaker data, as means of predicting group success. None of the MMLA studies attempt semantic analysis of student talk in collaborative settings.

What this paper adds

- A theoretical framework for why and how an automated NLP approach can support discourse analysis research on co-located, computer-based, collaborative problem solving interactions. This framework, entitled the Problem Solving in Computational Environment Speech Genre, links children's physical interactions with computational devices to their verbal exchanges and presents a theoretical rationale for the use of NLP methods in educational research.
- Description of an interdisciplinary method that combines NLP techniques with qualitative coding approaches to support analysis of student collaborative learning with educational robotics.
- Identification of student learning outcomes derived from the semantic, PSCE Speech Genre and NLP approach.

Implications for practice and/or policy

- Educational researchers will be able to expand upon our findings towards the goal of using computation and automation to support microgenetic analysis of large datasets.
- Robust microgenetic learning findings will provide curriculum developers, educational technology developers and teachers with guidance on how to construct and or create learning materials and environments.
- From an interdisciplinary perspective, this research can support more interdisciplinary exploration of conversational dialogues that are ill-structured, indexical and referential. This research will support the further development of machine learning techniques and neural network models by computational linguists.

physical component including robotics (Dillenbourg, 2016). AI approaches that support research on learning are usually termed learning analytic approaches. In addition to NLP-based text data mining approaches, learning analytics may focus on gaze, gesture and other modal enactments (Worsley *et al.*, 2016). Key to all AI applications in education (those that support learning and those that support research on learning) is the reliance on computational entities that interpret and respond to human intelligence.

The goal of our work is the development of a theoretically based, computational method for assisting in the microgenetic analysis of speech data (specifically, copresent, collaborative problem-solving group talk in a robotics learning environment). Microgenetic analysis is an extremely robust form of learning research (Kuhn, 2002) that focuses on the development of conceptual understanding. However, due to the intensive nature of data collection and analysis required, microgenetic research is typically performed with very small numbers of participants, which limits the application of the findings (Pressley, 1992). In our work, we seek to address this limitation of the microgenetic approach by exploring how computational means can be developed and meaningfully deployed to assist researchers in microgenetic analysis.

Microgenetic analysis

Microgenetic analysis is an observational research technique in which the researcher attends closely to the social interactions, speech acts and the use of tools within the learning environment in order to understand the genesis (or the origins) of conceptual development. Siegler (2006) has described three essential properties of microgenetic analysis thusly: "(1) observations span the period of rapidly changing competence; (2) within this period, the density of observations is high, relative to the rate of change; and (3) observations are analyzed intensively, with the goal of inferring the representations and processes that gave rise to them" (p. 469). Collecting and analyzing *all* interactions over a given period of time gives the researcher the advantage of understanding the trajectory of the cognitive change. As a result, empirical findings arrived at through microgenetic analysis are remarkably robust (Kuhn, 2002).

The two primary constraints of the microgenetic technique include the amount of time it takes to conduct close analysis of the data (Siegler, 2006), and secondarily the lack of the generalizability of findings derived from microgenetic case studies featuring only a few participants (Pressley, 1992). Here, we examine if and how computational methods can be meaningfully deployed to mitigate the time and generalizability constraints of microgenetic techniques.

Learning analytics

Learning analytic techniques have been used with three types of data that can be collected from people while they are learning: behavioral, physiological and representational. Approaches that focus on representational data include those that examine speech, text, drawing and other externalizations of cognitive activity. Written text analysis is, by far, the most prevalent form of learning analytics aimed at representations. This is due, in part, to the number of natural language processing tools currently available to perform such analysis, eg, Coh-Metrix (Dowell, Graesser, & Cai, 2016), Netlytic, Linguistic Inquiry and Word Count (LIWC), Rapid Miner, LightSIDE and WEKA (Gruzd, Paulin, & Haythornthwaite, 2016). These tools make possible a number of descriptive and analytic activities, eg, word, sentence and paragraph counts, word cloud visualization, sentiment analysis, lexical diversity type–token ratio calculations to determine text cohesion, as well as examination of the relatedness of words through cluster analysis. This last technique is useful for delineating topics of discussion.

Another approach, multimodal learning analytics (MMLA), seeks to capture and synthesize some combination of all three types of learning data: behavioral, physiological and representational, towards the goal of developing comprehensive models of student learning (Worsely *et al.*, 2016). Recent studies related to colocated, computer-based, collaborative problem solving (CPS) using the MMLA approach, focus on understanding how gaze, gesture and physical actions in the computer environment predict group success (Cukurova, Luckin, Millán, & Mavrikis, 2018; Schneider & Blikstein, 2015; Schneider, Sharma, Cuendet, Aufferey, Dillenbourg, & Pea, 2018; Spikol, Ruffaldi, Landolfi, & Cukurova, 2017). These studies have demonstrated the utility of these types of data as a means of convergent triangulation in correlating non-verbal elements with learning gains.

Other co-located computer-based, CPS studies using the MMLA approach have included verbal data in the analysis, but these studies focus not on the semantic meaning of the utterances, rather they focus on prosodic elements (pitch, tone), duration of speaking time and/or turn-taking to help interpret group functioning (Lubold & Pon-Barry, 2014, Praharaj, Scheffel, Drachsler, & Specht, 2018). While all of these studies are important and are contributing to our understanding of body language and turn taking in colocated, computer-based, CPS, none of them are seeking to use learning analytics to semantically analyze human dialogue.

MMLA appears to be a very promising method; however, the research undertaken in this vein has a number of limitations that have yet to be fully resolved. For instance, in a review of the MMLA literature Worsley (2018) notes that 35 of 46 empirical studies published from 2012 to 2018 have taken place in laboratory settings; this is so because the reality of collecting multiple streams of data from numbers of individuals in actual learning settings is not feasible. Indeed, two recent field-based studies in the MMLA literature report on the difficulty of recording audio data in particular (Echeverria, Falcones, Castells, Granda, & Chiluiza, 2017; Liu & Stamper, 2017). However, while laboratory studies are easier to set up and control, they lack external validity, thereby attenuating the utility of findings. A further problem with the MMLA approach is the interpretability of multiple data streams within a unifying theoretical framework, such that the "the output of MMLA [becomes] more actionable." (Worsley, p. 8). These challenges remain for MMLA researchers to solve.

Our work focuses on verbal interactions as representations of cognitive activity. We seek to build on the work of researchers who have used learning analytics to identify meaningful textual exchanges in learning situations. For example, Ferguson and Shum (2011) used a technique that focused on identifying specific discourse features in online discussions as well as time stamp data and counts of participants to allow them to pinpoint particularly meaningful discussions in the context of a daylong online workshop for teachers. Thus discovered, these meaningful discussions could then be submitted to deeper analysis. Our work builds on this work, but does so with transcripts of copresent, indexical conversations, a qualitatively different data corpus, that presents special problems for analysis.

Text mining and the analysis of co-present data

Our research focuses on analyzing conversational data about educational robotics captured on video and audio tape in face-to-face classrooms and informal learning environments. Our work is similar to that of other educational researchers who have focused analysis on spoken language including Worsley and Blikstein (2011) and Sherin (2013). However, it differs from these projects in that we seek to understand in-situ, problem-solving conversations, not responses to think-aloud prompts (Worsley & Blikstein) or interview questions (Sherin; Worsely & Blikstein). Our colleague's data are qualitatively different to our data, in that think-alouds and interviews are aimed at eliciting targeted and specific thoughts or responses, whereas, the dynamic, ongoing, in-situ interchange of collaborative problem-solving conversations is noticeably different in linguistic character. Our data is highly contextualized, referential and indexical talk that is often partial, fragmented and/or overlapping.

Dialogue models, semantic and syntactic analysis

To guide us in analysis of our dataset, we turned to the work of Ginzburg and Fernandez (2010), Bakhtin (1986), and Goffman (1974). Ginzburg and Fernandez have contributed, theoretically, to the development of computational models of dialogue for automated agents in teleservice settings; eg, booking an airline flight over the phone. From their work, we derived ideas for analyzing sentence fragments, indexical and pronominal terms. Specifically, we realized we needed to work at both the utterance (semantic) and the grammatical (syntactic) levels, focusing on *what* was said and *how* it was said. To get at the latter we used parts of speech analysis (eg, noun, verb, adverb, preposition). In this way, we could meaningfully examine partially expressed thoughts and sentence fragments, as well as more fully formulated utterances. This conceptualization of our analysis task led us to thinking more about the nature of the robotics activity, itself. We realized that the very bounded nature of the activity may, similarly, bound the types of speech acts students might utter. To develop these ideas further, we turned to Bakhtin's (1986) work on speech genres and Goffman's (1974) theory related to social frameworks.

Speech genres and social frameworks

According to Bakhtin (1986), speech genres are characterized by relatively stable types of utterances occurring within a particular sphere of human activity. There are many and varied types of speech genres from everyday talk—"short rejoinders in everyday dialogue" (p. 60)—to various forms of writing (eg, the novel, scientific reports) to verbal military commands to poetry. The social and symbiotic nature of speech genres may be regarded as tools that help us act in and make sense of the world, and as products of our acting in and making sense of the world (Varelas, Becker, Luster, & Wenzel, 2002). Speech genres serve to organize a sequence of interactions in a culturally recognizable situation (Wells, 1999). This culturally recognizable situation may best be thought of as Goffman's (1974) social interaction frame.

Goffman (1974) argues that all social interactions are framed by the sociocultural context and an individual's understanding and interpretation of that context. Social frameworks "provide background understanding for events that incorporate the will, aim, and controlling effort of an intelligence, a live agency, the chief one being the human being" (p. 22). Varenne (1998) adds to Goffman's frame theory by discussing the "always already there" (p. 185) impact of historically situated cultural and social facts. However, both theorists stress the idea that individuals—while influenced by the cultural and social frames they are born into—have the ability to act independently to achieve their own specific goals.

Therefore, specific sociocultural contexts, such as working in a small collaborative group to solve a robotics problem in a sixth-grade science class, invoke a social interaction frame for students and evinces the relatively stable use of particular utterances in speech interactions. In this paper, we define a problem solving in a computational environment (PSCE) speech genre that refers to talk that occurs among middle school students within the context of solving a robotics problem in class (a particular sphere of human activity). The categories created for this analysis are reported in the methods section. In adopting this speech genre approach, we are *not* attempting to map out the entire domain of possible speech acts that may occur in the setting, rather we are seeking to identify the *regularities* in the speech genre that may point, over time, to the microgenetic development of conceptual understanding.

Methods

Speech genre analysis and qualitative models of activity

Our method includes a speech genre analysis in which we seek to understand the *work* that particular types of utterances are doing in a given student interaction while solving a robotics problem. We utilize qualitative models of student activity in the problem-solving environment to help us contextualize student utterances and better understand the possible meaning of an utterance. To explore this method, we are using a dataset we collected as part of a prior study. For a complete description of the research site, participants and data collection procedures from this prior study, see Sullivan (2011). In this prior analysis, we examined a focal student groups' collaborative development of a creative idea in solving a robotics challenge. The teaching goal of the robotics challenge set by the curriculum designers and enacted by the teacher, was to acquaint the students with the functioning of the light sensor. A key element of the creative solution development was the change in student conceptual understanding of how the light sensor functions.

As part of this prior, by-hand analysis, we developed a qualitative model of student problem-solving activity with robotics. This model consisted of a troubleshooting cycle, which includes the following activity: "(1) writing and testing the program, (2) diagnosing problems with the program or structure of the device, (3) proposing and arguing for specific changes to the program/structure, (4) making changes to the program/structure, and (5) testing the device again" (Sullivan, 2011, p. 57). The troubleshooting cycle is a relatively regular and stable feature of student activity while solving robotics problems.

The temporally sequential nature of the troubleshooting cycle is ideal for microgenetic analysis and it strongly informed our development of the PSCE coding system (presented below). Through our current computational analysis of this same dataset, we sought to identify regularities in speech (the problem solving in computational environments speech genre) that might map to the identified regularities in student troubleshooting activity. Drawing on Bakhtin's (1986) speech genre theory, we argue that while students are engaged in the troubleshooting activity characteristic of group CPS with robotics, their speech will, likewise, be directed towards problem solving. Figure 1 presents a theoretical model of our speech genre approach.

Computational analysis

The computational aspect of our approach relies on linguistic regularity and the *function* of particular grammatical constructions as spoken by students in the robotics learning problem space. Due to the relatively stable character of troubleshooting cycle activity, we hypothesized

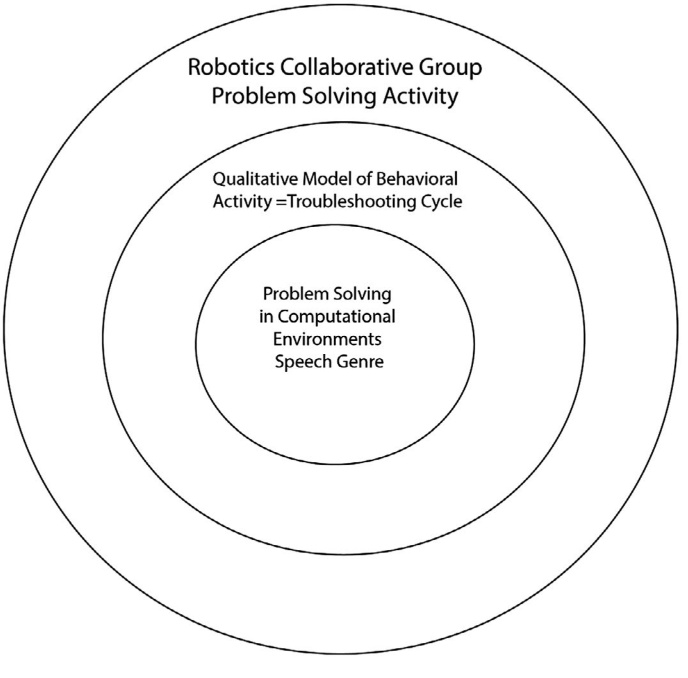

Figure 1: Theoretical model of PSCE speech genre development

a relatively stable character to the domain of utterances that may be offered during these times: in short, we hypothesized an identifiable speech genre that we could then explore computationally. As noted above, speech genres are bounded domains of potential utterances that make up the genre. Further, the utterance themselves will consist of a range of specific words (the vocabulary associated with the activity) and they will have a particular grammar. As Halliday and Matthiessen (2014) have noted, it is both the vocabulary and the grammar that are the meaning making resource for those involved in the activity and the discussion. While text-mining keyword searches are useful for delimiting the vocabulary usage, we found that they were insufficient for analysis of learning processes. Therefore, we sought to also examine the grammar of utterances.

Towards that end, we employed a natural language processing library created by colleagues at Stanford University (Toutanova, Klein, Manning, & Singer, 2003), featuring a parts-of-speech (POS) Treebank developed at the University of Pennsylvania by Santorini (1990) to allow us to identify important grammatical constructions that make up the PSCE speech genre. We reasoned that the POS tagger would begin to help us identify types of utterance that may all be doing the same kind of *work* in terms of the troubleshooting cycle; eg, we sought to linguistically identify periods of diagnostic activity and periods of argumentation. These activities are foundational to the troubleshooting cycle. We reasoned that if we could identify cycles of diagnostic and argumentative activity in the dataset, we could find conceptually rich segments of the data. We also sought to identify comments aimed at group regulation (eg, who will do what) as these types of comments are foundational to the functioning of collaborative learning groups (Sullivan & Wilson, 2015).

Ngram unit of analysis

In developing our computational POS tagging approach, we needed to decide upon a useful unit of analysis. We chose to work with utterances at the level of the bigram (two words) and trigram (three words). We selected these ngram configurations because, arguably, they are the smallest levels at which complete utterances might be made. Halliday and Matthiessen (2014) point out that while the clause is the smallest semantic unit in the English language, clauses are made up of smaller grammatical units that also have meaning, including the nominal group, the verbal group, the adverbial group and the prepositional phrase. Importantly, the *theme* of a clause will be carried by one of these smaller structural elements (p. 92). In this way, Halliday and Matthiessen's (2014) system of classifying utterances into these smaller units provides a basis for interpreting the partial contributions of grammatical elements in determining the meaning of an utterance in context. For example, nominal groups that include deictic terms (eg, "that," "this," "these," "those," "here," "there") indicate contextually specific Things. An instance of this from our trigram analysis is *that black light*. In this nominal group, the *light* is the Thing that the students are indicating. It is subcategorized by the modifier *black* and differentiated from other Things by the use of the determinative *that*.

Other examples of nominal groups that feature deictic terms from our trigram analysis were: *all those lines, the ruler there.* The work that these trigrams performed was to focus the attention of the group on a specific Thing in the robotics learning environment. These Things were, almost always, one of the objects that help make up the robotics challenge problem space (eg, the microcomputer, the sensors, the motors, the Lego pieces, the laptop, the Robolab software program, the space on the classroom floor designated as the testing zone for completed programs). In this way, particular POS analysis of the trigrams that included deictic terms and nouns could be interpreted within the PSCE coding framework as having to do, specifically, with objects in the problem space.

Meanwhile, verbal groups contain a word whose primary classification is a verb. Examples of verbal groups from our trigram analysis include: *have to go, have to put* and *go to step.* By definition, verbal groups deal with the action or states of being. In the context of robotics problem solving, these verb groups often dealt with what action the group needed to take to solve the problem or the state of the robot in relation to solving the challenge; hence, the interpretability of the verb group POS trigram. Likewise, adverbial groups and prepositional phrases were also interpretable.

The data used to pilot our method consisted of the interactions of three students during a 30-minute videotaped classroom activity devoted to learning how to use the light sensor. To accomplish the PSCE speech genre analysis, the transcripts of the words uttered by the students were broken down into bigram or trigram word segments which were named in such a way as to retain temporal differentiation. For example, an excerpt from the transcription reads:

1. *I: okay*
2. *J: oh okay*
3. *I: but we need a ruler to make it go far away*

Single-word utterances were not considered for this analysis; therefore, the utterance in line one was not included. An utterance of two words was included in the analysis if and only if the entire utterance consisted of two words. Since line two consists of only two words, it was included in the analysis. Any utterance of three or more words was then divided into multiple overlapping three-word segments and included for analysis, the formula for developing the trigram segments was (n-2). Therefore, the 11-word utterance in line three would have been divided into nine segments:

but we need

we need a

need a ruler

a ruler to

ruler to make

to make it

make it go

it go far

go far away

The line number of the original utterance was preserved along with each unique segment to retain temporality. The dataset produced 2,627 unique ngram segments of text. The vast majority of ngram segments were trigrams (as opposed to bigrams), in the following pages we refer to POS trigrams as a default term for both POS bigrams and trigrams.

Parts of speech tagging

These trigram segments were then processed through the Java-implemented Stanford log-linear POS tagger (Toutanova, *et al*., 2003). POS taggers tokenize individual words and then utilize computational methods to assign a POS (such as noun, verb, coordinating conjunction) to each word. The Stanford POS tagger utilizes the Penn Treebank tag set (Santorini, 1990). Marcus, Santorini, and Marcinkiewicz (1993) note that the computational Penn Treebank approach outperforms manual methods on three dimensions "speed, consistency, and accuracy" (p. 313).

A report was then created of each unique POS trigram, along with the associated text segment and line number. Based on our domain expertise and the qualitative model of the troubleshooting cycle, we mapped each POS tag string to a code in the PSCE speech genre coding scheme (Table 1). It is important to recognize that the same POS tag string may be assigned to trigrams consisting of different words. We sought to interpret the *work* each POS tag string was doing by looking across the trigrams that garnered the same string. In other words, it was not the specific words that mattered, rather it was the grammatical role the words played in the overall structure of the utterance that mattered. The mapping of POS trigrams to PSCE codes was a deliberative process. The process included reading the trigrams in context and discussing each one in terms of the troubleshooting cycle qualitative model. Based on this model and the specificity of this text, we both deductively and inductively created the PSCE coding system used here.

In order to meaningfully manage the 2,627 unique trigram segments of text, we elected to only map frequently occurring POS trigram segments to the PSCE codes. For the purposes of this analysis, if 5 or more segments of text were associated with the same POS trigram, we mapped that POS configuration to the PSCE codes. Exceptions were made when a partial POS tag could be coded to the same code. For example, in Table 2, the trigrams were all coded as *activity negotiation*. Therefore, any segments with a partial match of RB VBP were coded as *activity negotiation*. That being said, trigrams with partial matches to a specific PSCE code were mapped only to that code and to no other code. All trigrams were only counted once, and only assigned to one PSCE code (if warranted by frequency count). Some trigrams were not coded beyond being assigned the POS tag string, even if there were five instances or more. This was the case if it was clear that the trigram consisted of ideas belonging to two separate sentence clauses, meaning that a noun, verb or adverb group or a prepositional phrase had been split up in the trigram segmenting process. In these instances, the trigram would be handled more appropriately in a different trigram constellation with no break in grammatical meaning.

Finally, as part of our analysis, we produced a stacked histogram (Figure 2) that indicated the number of PSCE codes generated per group of 10 utterances. For this analysis, the utterances were grouped temporally by their original location in the transcript. The number and type of PSCE-coded trigrams that occurred in each group determined the height of the bar. For example, point 0 on the x-axis represents the first 10 utterances made by the students in this 30-minute problem-solving vignette, 15 of these trigrams received a PSCE code. The stacked histogram was then analyzed to *identify clusters of high occurrences of coded trigrams*, which, theoretically, would

Table 1: Problem solving in computational environments speech genre coding scheme

Diagnosis	*Query*	*Argumentation*
Evaluation	Clarification	Group regulation
		Organization of tasks/roles
		Modal
		Activity negotiation
Confirmation		Content and concepts
		Programming elements
		Comparative
		Explanation
		Building elements
		Comparative
		Explanation
Puzzlement		Problem definition
		Familiarization

Table 2: Example of POS to PSCE coding scheme

Text segment	*POS tag string*	*Tag meaning*	*PSCE code*
Now do the	RB VBP DT	Adverb, verb, determiner	Activity negotiation
Now put it	RB VBP PRP	Adverb, verb, preposition	Activity negotiation
Hey don't play	RB VBP RB VB	Adverb, verb, adverb, verb	Activity negotiation

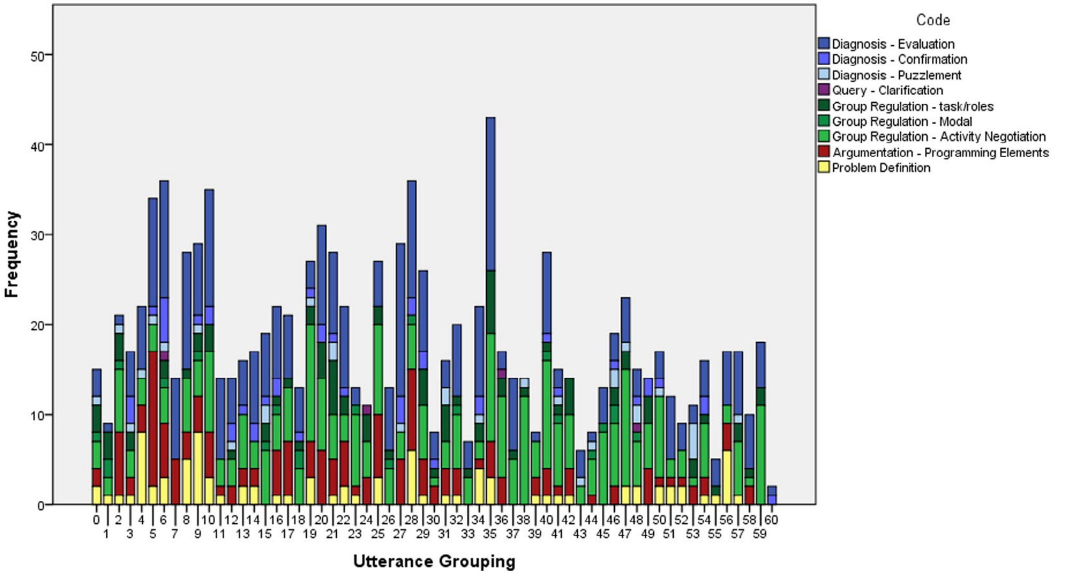

Figure 2: Stacked histogram of coded ngrams
[Colour figure can be viewed at wileyonlinelibrary.com]

indicate sections within the transcription that are conceptually rich in terms of discussion within the troubleshooting cycle. In the following section, we present the results of our analysis of these.

Results

There were 605 utterances made by the students in the 30-minute vignette analyzed here. A visual inspection of the stacked histogram (Figure 2) revealed that seven of the utterance clusters peaked above the 25 PSCE code point on the y-axis, signaling a potentially high level of problem-solving discussion in the text at those peaks. In closely analyzing these segments, we were able to identify a clear progression in student conceptual understanding related to the functioning of the light sensor, which was the goal of this particular challenge and key to their creative problem solving. In relation to Figure 2, the utterance cluster peaks can be identified as points along the x-axis as demonstrated in Table 3.

In the following, temporally sequenced excerpts (Tables 4–8), we demonstrate, with the data, the change in student's way of thinking and talking about how to solve the problem, and specifically how to use the light sensor to help solve the problem. As will be shown, the students begin with *no understanding* of how the light sensor functions in relation to solving the problem, they then progress to a view of the light sensor as a measurement device whose primary purpose is to record a light reading. This view then evolves as they work through the problem to an understanding of

Table 3: Correspondence of utterance cluster peaks to transcript lines

Cluster #	*x-axis points*	*Corresponding utterance lines in transcript*
1	5–6	51–70
2	8–10	81–110
3	19–21	191–220
4	25	251–260
5	27–29	271–300
6	35	351–360
7	40	401–410

the light sensor as a computational device, a device that not only takes a reading, but compares the current reading to a programmed threshold and executes an event triggered by reaching the threshold. This progression demonstrates how the students' conceptual understanding of the light sensor changed over time and through interaction in the learning environment. The utterances that were most relevant to our interpretation occurred in five of the seven clusters, these utterances are presented in bold text in Tables 4–8 below.

Meanwhile, two of the utterance clusters identified by the MLA technique, while evidencing a high number of PSCE codes, did not relate directly to the use of the light sensor. One of the conversations focused on a debugging discussion related to the placement of the robot on the floor (line #s 251–260), and the second conversation involved the teacher who modeled debugging activity for the students while helping them think about a problem, and, hence, his utterances increased the number of PSCE codes in that segment (line #s 351–360). Because we were interested in how the MLA technique might aid in understanding how student knowledge changed over time, we focused on the identified discussions as they related to the light sensor.

The specific robotics challenge students were trying to solve is as follows:

1. move forward until a black line on the floor is detected
2. turn 90 degrees
3. back up for 1 foot
4. stop

In the first excerpt (Table 4), students discuss programming the robot to move forward by use of a timing element. However, no timing element is required for the first step of this challenge. To solve this challenge, the first step needs to be programmed with a light sensor, not a timing element. The fact that the students were discussing how to program the forward movement of the robot with a timing element, indicates they did not, yet, understand how the light sensor functions.

Table 4: Cluster #1: programming with timing elements

Line #	*Speaker*	*Utterance*
50	J	Well I don't know why it's only doing the first of three that step let's do one cause it probably *has to be on time first*
51	S	*Yeah that has to be on time*
52	J	*Then this has to do the light sensor*
53	S	No do it
54	J	*This has to be this has to be time again*

As can be seen in line 50, student J states that the program "has to be on time first," meaning the first step in their program is to move the robot forward for a specific amount of time. This reasoning is concurred with on line 51 by student S when she states "yeah that has to be on time." In line 52, student J suggests that the second step of the program should involve the light sensor, but the choice of words belies little knowledge of how to actually do this—he says "this has to do the light sensor." Indeed, the first step in the program should set a triggering numeric threshold for the light sensor. The student's comment in line 52 does not reflect this level of understanding.

A few minutes later, the students were still using the timing element, but they were starting to discuss the idea that, in order for the program to function properly, the light sensor must see the black line, which they erroneously call the "light," in line 106. This excerpt is presented in Table 5.

Table 5: Cluster #2: students realize need to program sensor

Line #	*Speaker*	*Utterance*
95	S	*It's going forward for a time it's going to step one*
96	J	*Going forward right for one second*
97	S	*Yeah but it has to step for one second right but it has to go and touch the black line right yeah cause then the sensor*
98	S8	Anybody lose a ring yeah I know
99	S	Yeah that's good now let's send it
100	J	Wait wait you have to wait now *I have to put this is time* now I have to put this back onto this stupid line
101	S	No not that that doesn't really have to oh yea that does yeah no
102	J	Yes it does yes wait
103	S	Oh look at this
104	J	It has to go backwards
105	S	Oh it does
106	J	Yeah it has to go back it has **to** *hit the light* then go backwards

On line 97, student S begins to identify the contradiction between their existing program, which uses a 1-second timing element to move the robot forward, and the instruction to write a program that uses the light sensor to detect a black line. Here, the notion that the robot has to "touch" the black line refers to the need for it to *sense* the black line.

In Table 6, students continue to think they need to program both the light sensor and provide a timing element in order to move the robot forward. However, their understanding of the light sensor is improving as they now realize they need to set a numeric threshold with it.

Table 6: Cluster #3: continuing to think two elements are necessary

Line #	*Speaker*	*Utterance*
199	J	Watch *I think I know what the problem is the light let's put it at thirty five* cause it's on so now send wait hey hey hey
200	I	This is going to be over a foot
201	S	It's going too slow
202	J	Oh my g-d *no it doesn't want to go forwards cause there's no time limit*

On line 199, Student J is diagnosing a problem with the robotic device in terms of the light sensor program, he suggests they use a threshold reading of 35 to trigger the next element of the program. However, in line 202 the same student decides that the robot is not moving forward "cause there's no time limit." Again, we see the confusion of the role of the light sensor in solving the problem. However, we also see that the students are starting to understand that the light sensor is integral to solving the challenge.

In the next excerpt (Table 7), a few minutes later, the students articulate the idea that the robot will turn when it "sees" the "black light" and so, one cannot start the program from the "black light." This is the first indication in their discussion that the students understand that the sensor, if programmed correctly, may trigger another event.

Table 7: Cluster #4: sensor as triggering device

Line #	*Speaker*	*Utterance*
280	J	*You're not supposed to put it on the black light that's why*
281	S	There you go
282	J	*It has to go away from the black light*
283	I	It's gonna follow your
284	S	Hello black shoe
285	J	*It has to start from the you can't from the black light that's why it's not doing it*
286	S	Black shoe
287	J	No let me watch now that's to get so far then *it has to see the black light and turn* but if it, it doesn't do that well then it has to go turn and go backwards

On line 280, student J indicates that the group cannot place the robot on the black line as a way to initiate the program. In line 282, he indicates that the robot "has to go away from the black light," a comment he reiterates on line 285. Finally, on line 287, student J indicates that the robot has to "see the black light and turn." Taken together, these comments demonstrate that student J has begun to understand that a calculation is required, the group cannot start the program on the black line because the robot needs to "see" the black line as differentiated from some other color (in this case the gray carpet).

After a few more minutes, student J has a more significant breakthrough in understanding how the light sensor functions. In Table 8, student J explains the breakthrough to the other students—pointing to the displayed readings on the sensor, student J calls attention to how fluctuations in that number affect the movement of the motors on the robot. This is the essence of the sensor as a triggering device. At this point, the students are well on their way to solving the robotics challenge, having constructed an understanding of the light sensor as a computational device.

Table 8: Cluster #6: light sensor as computational device

Line #	*Speaker*	*Utterance*
400	J	Turn it off turn it off turn it off turn it off *watch see this goes the lower the number it goes straight right and then when it changes to forty three it doesn't try to,* forty one see if it does it connect, just hold it in the air see it's the light it's too big
401	S	It's going forward
402	J	*Check the black line check that black line on the on the white paper the one in the middle*

On line 400, student J remarks "the lower the number it goes straight, right? And then when it changes to forty-three it doesn't try to…" In this segment, student J has made the connection between the registered light reading (lower than 43) and the triggered action "it goes straight, right?" He then notices that when the light sensor reaches the threshold of 43, it no longer moves forward. Student J argues that a light reading of 43 is "too big" to trigger movement. While student S seems to contest that with her comment that "it's going forward," student J appears to have made a conceptual breakthrough, he realizes that specific, directional fluctuations in the light readings are triggering specific responses in the robotic device—the essence of the programmed sensor.

In these computationally identified clusters, it is possible to see how the students shifted their thinking about the first step of the problem and the use of the light sensor in solving the problem. Over time, they realize that the light sensor is not just used to sense a different color but can be used as a tool to control the movement of the robotic device. In this way, the students are deepening their understanding through interaction with the tool itself, and they achieved the goal of the lesson, to learn how to use the light sensor.

Additional observations on the method

While there were a total of 2,627 unique trigrams identified in the overall 30-minute text, only 115 unique trigrams were mapped to the high-peak clusters observed in the histogram visualization. Moreover, of these 115 unique trigrams, 110 of them had one of only 11 (out of the possible 36) POS tags in the first position of the trigram. And, of these 11, 4 fall into the verb category (each of these four parts of speech representing a different verb tense and/or a different speaker position). These parts of speech, examples of the words they denote and the trigrams they initiate, are presented in Table 9.

Table 9: First position POS tags most frequently associated with PSCE Codes

Parts of speech	*POS tag*	*Word examples (partial list)*	*Trigram examples (partial list)*
Conjunction coordinating	CC	but, and	"but we need," "but it has," "but where do," "and then it," "and what was"
Preposition, subordinate conjunction	IN	like, of, for, on, over, in, with, as	"of eighty-five," "like one foot," "for one second," "on the motor," "over a foot" "with the light," "in a circle" "as you turn"
Adjective comparative	JJR	higher, more, darker,	"higher the shorter," "more than one," "darker than the"
Noun	NN	sensor, floor, program, problem	"sensor is triggered," "floor is the," "program that robot," "problem is the"
Personal pronoun	PRP	it, I, you, we	"it just stopped," "it goes backward," "I wonder if," "I need that," "you think that," "you should use," "we need to"
Adverb	RB	backwards, exactly, slowly, up,	"backwards a foot," "exactly a foot," "slowly by one," "up in the"
Preposition (to)	TO	to	"to the robot," "to the floor," "to step one"
Verb (various tenses and speaker designations)	VB VBD VBP VBZ	get, need, hit, run, got, are, has, is, because	"get a ruler," "need a hand," "hit the back," "run the program," "got to plug," "are we going," "has to be," "has to go" "is going backwards" "because it says"

As can be seen in Table 9, the first position parts of speech tags that were most generative in terms of the PSCE codes are either words that one may easily associate with reasoning and argumentation—the word "but" may indicate refutation, the word "and" elaboration, the words "higher" and "shorter" comparison, the words "is," and "are" with questions, the word "because" with evidence or hypothesis generation—or words that identify an entity—"it" may refer to any of the objects in the problem space, "sensor" and "floor" refer to specific objects in the problem space, "I, you, we" refer to people in the problem space. Likewise, the words "backward, forward" refer to the movement of an object in the problem space.

We view this observation as supportive of our conjecture that speech in specific realms is bounded by the structure of the activity in that realm. In a different, bounded realm of activity, it might be possible to see more than 11 first position POS tags as meaningful, and/or these tags would be different to what is meaningful in our context. The physically active, tool rich, multidimensional problem space of robotics learning motivates specific speech acts and affords particular grammatical constructions, which then lend themselves to computational exploration.

Discussion

In this paper, we have presented our exploration of a natural language processing approach to assist in the microgenetic analysis of discourse data. A major issue we confront in our research, here, is the highly contextualized, indexical and fragmented nature of student talk while solving a robotics problem. The data corpus we worked with was full of partially expressed thoughts and deictic comments that referred to "it," "that" and "there," for example. Using computational means to analyze such fragmented talk is difficult. Our approach of analyzing the text at the trigram level included creating a trigram based on every word in the sentence except the last two (n-2), in this way, we gained full coverage of student expressions and were able to identify meaningful noun and verb groups. Our observation that a particular set of first position POS tags was more generative than other tags is helpful in defining an overall way forward—the mapping of POS tags to annotations, could be streamlined for genres of activity, such as active, tool-rich, multidimensional problem spaces.

Our computational method is not intended to be a fully automated approach, nor is it meant to function as a full solution to the problem of microgenetic analysis. Rather, *our approach is a powerful aid* to the educational researcher who is an expert in her or his chosen area of study. Indeed, the expertise of the researcher is key to both the development of the coding scheme that POS trigrams would be mapped to (in our case the PSCE coding system) and to the meaningful interpretation of computationally identified segments of the transcript.

We have developed our method in the context of a tool-intensive activity, a robotics learning setting. In this way, our approach would transfer well to other, similar settings. But, it is not clear how the method would work in less bounded and tool-intensive collaborative problem-solving settings. The bounded-ness of the problem-solving activity is key to our theoretical approach. And, we argue, there are many, similarly bounded educational settings, especially when technology is involved. Therefore, this approach will be useful in a number of educational research studies.

Acknowledgements

The research reported in this manuscript was supported by funding from the National Science Foundation DRL #1253250. The authors would like to thank Dr Rick Adrion for computer science mentorship in interdisciplinary research.

Statements on open data, ethics and conflict of interest

The transcript data used in this study may be obtained by written request to the corresponding author.

The Institutional Review Board (IRB) at the author's university reviewed and approved this research study following university regulations. The university IRB subscribes to the ethical standards delineated in the Belmont Report, all applicable federal, state and local regulations related to the conduct of human subject research, as well as University policies and procedures. The university IRB is registered with the US Department of Health and Human Services Office for Human Research Protections (IRB00000017) and it holds a Federal wide Assurance (FWA) which documents the University's commitment to comply with federal regulations for the protection of human subjects in research (FWA00003909).

The authors have no conflicts of interest related to the study reported here.

References

Bakhtin, M. M. (1986). The problem of speech genres. In V. W. McGee (Trans.), C. Emerson, & M. Holquist (Eds.), *Speech genres and other late essays* (pp. 60–102). Austin, TX: University of Texas Press.

Cukurova, M., Luckin, R., Millán, E., & Mavrikis, M. (2018). The NISPI framework: Analysing collaborative problem-solving from students' physical interactions. *Computers and Education*, *116*, 93–109. https://doi.org/10.1016/j.compedu.2017.08.007

Dillenbourg, P. (2016). The evolution of research on digital education. *International Journal of Artificial Intelligence in Education*, *26*, 544–560. https://doi.org/10.1007/s40593-016-1016-z

Dowell, N. M. M., Graesser, A. C., & Cai, Z. (2016). Language and discourse analysis with Coh-Metrix: Applications from Educational material to learning environments at scale. *Journal of Learning Analytics*, *3*(3), 72–95. https://doi.org/10.18608/jla.2016.33.5

Echeverria, V., Falcones, G., Castells, J., Granda, R., & Chiluiza, K. (2017). Multimodal collaborative workgroup dataset and challenges. *CEUR Workshop Proceedings*, *1828*, 94–98. https://doi.org/10.475/123

Ferguson, R., & Shum, S. B. (2011). Learning analytics to identify exploratory dialogue within synchronous text chat. *Proceedings of the LAK 2011: 1st International Conference on Learning Analytics and Knowledge*, Banff, Alberta, Canada, 99–103.

Ginzburg, J., & Fernandez, R. (2010). Computational models of dialogue. In A. Clark, C. Fox, & S. Lappin (Eds.), *Computational linguistics and natural language processing handbook*. Oxford, UK: Blackwell.

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. New York, NY: Harper and Row.

Gruzd, A., Paulin, D., & Haythornthwaite, C. (2016). Analyzing social media and learning through content and social network analysis: A faceted methodological approach. *Journal of Learning Analytics*, *3*(3), 46–71.

Halliday, M. A. K., & Matthiesen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). New York, NY: Routledge Press.

Kuhn, D. (2002). A multi-component system that constructs knowledge: Insights from microgenetic study. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 109–130). Cambridge, England: Cambridge University Press.

Liu, R., & Stamper, J. (2017). Multimodal data collection and analysis of collaborative learning through an intelligent tutoring system. *LAK Workshops*, *17*, 1–6.

Lubold, N., & Pon-Barry, H. (2014). Acoustic-prosodic entrainment and rapport in collaborative learning dialogues. In *Proceedings of the 2014 ACM WS on Multimodal Learning Analytics Workshop and Grand Challenge* (pp. 5–12). Istanbul: ACM.

Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The penn treebank. *Computational Linguistics*, *19*(2), 313–330.

Praharaj, S., Scheffel, M., Drachsler, H., & Specht, M. (2018). Multimodal analytics for real-time feedback in co-located collaboration. In *Proceedings of the 13th European Conference on Technology Enhanced Learning, EC-TEL 2018*, September 3–5, 2018. Leeds, UK. https://doi.org/10.1007/978-3-319-98572-5_15

Pressley, M. (1992). How *not* to study strategy discovery. *American Psychologist*, *47*, 1240–1241.

Santorini, B. (1990). *Parts-of-speech tagging guidelines for the Penn Treebank project* (3rd revision). University of Pennsylvania, Penn Engineering, Scholarly Commons. Retrieved from http://repository.upenn.edu/cis_reports/570/?utm_source=repository.upenn.edu%2Fcis_reports%2F570andutm_medium=PDFandutm_campaign=PDFCoverPages

Schneider, B., & Blikstein, P. (2015). Unraveling student's interaction around a tangible interface using multimodal learning analytics. *Journal of Educational Data Mining*, *7*(3), 89–116.

Schneider, B., Sharma, K., Cuendet, S., Aufferey, G., Dillenbourg, P., & Pea, R. (2018). Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *International Journal of Computer Supported Collaborative Learning*, *13*, 241–261.

Sherin, B. (2013). A Computational study of commonsense science: An exploration in the automated analysis of clinical interview data. *Journal of the Learning Sciences*, *22*(4), 600–638. https://doi.org/10.1080/10508406.2013.836654

Siegler, R. S. (2006). Microgenetic analyses of learning. In W. Damon & R. Lerner (Eds.), *Handbook of child psychology* (6th ed., pp. 464–510). Hoboken, NJ: John Wiley and Sons.

Spikol, D., Ruffaldi, E., Landolfi, L., & Cukurova, M. (2017). Estimation of success in collaborative learning based on multimodal learning analytics features. In *Proceedings of the IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, Timisoara, Romania, 269–273. https://doi.org/10.1109/ICALT.2017.122

Sullivan, F. R. (2011). Serious and playful inquiry: Epistemological aspects of collaborative creativity. *Journal of Educational Technology and Society*, *14*(1), 55–65.

Sullivan, F. R., & Wilson, N. (2015). Playful talk: Negotiating opportunities to learn in collaborative groups. *Journal of the Learning Sciences*, *24*(1), 5–52. https://doi.org/10.1080/10508406.2013.839945

Toutanova, K., Klein, D., Manning, C., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. *Proceedings of HLT-NAACL*, Edmonton, Canada, 252–259.

Varelas, M., Becker, J., Luster, B., & Wenzel, S. (2002). When genres meet: Inquiry into a sixth-grade urban science class. *Journal of Research in Science Teaching*, *39*(7), 579–605.

Varenne, H. (1998). Local construction and educational facts. In H. Varenne & R. McDermott (Eds.), *Successful failure: The school America builds*. Boulder, CO: Westview Press.

Wells, G. C. (1999). *Dialogic inquiry: Towards a sociocultural practice and theory of education*. New York, NY: Cambridge University Press.

Worsley, M. (2018). Multimodal learning analytics: Past, present and potential futures. In *Companion Proceedings 8th International Conference on Learning Analytics & Knowledge*, March 7–9, 2018. Sidney, Australia.

Worsley, M., Abrahamson, D., Blikstein, P., Grover, S., Schneider, B., & Tissenbaum, M. (2016). Situating multimodal learning analytics. In C.-K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners. Proceedings of the International Conference of the Learning Sciences (ICLS 2016)* (Vol. 2, pp. 1346–1349). Singapore: International Society of the Learning Sciences.

Worsley, M., & Blikstein, P. (2011). What's an Expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis. In *Proceedings for the 4th Annual Conference on Educational Data Mining* (pp. 235–240). Eindhoven, Netherlands.