

Towards Careful Practices for Automated Linguistic Analysis of Group Learning

Iris Howley

Human–Computer Interaction Institute
Carnegie Mellon University, USA
ihowley@cs.cmu.edu

Carolyn Penstein Rosé

Language Technologies Institute and Human–Computer Interaction Institute
Carnegie Mellon University, USA
cprose@cs.cmu.edu

ABSTRACT: The multifaceted nature of collaborative learning environments necessitates theory to investigate the cognitive, motivational, and relational dimensions of collaboration. Several existing frameworks include aspects related to each of these three. This article explores the capability of multi-dimensional frameworks for analysis of collaborative processes to isolate and assess these separate dimensions of collaboration. Much successful work has contributed towards computational modelling for automated collaborative process analysis in the past decade. In this paper, we explore the extent to which evidence points to an intertwining between dimensions, raising important caveats for careful consideration when making assessments based on the observation of codes as they are applied to collaborative discourse. We conclude with a research agenda for future investigation to address this limitation.

Keywords: Discourse analytics, transactivity, systemic functional linguistics

1 INTRODUCTION

Language behaviour is incredibly rich. When used within a protocol analysis methodology, it can be a window into the inner workings of one’s mind (van Someren, Barnard, & Sandberg, 1994). When situated within a sense-making task, language can provide the opportunity for an individual to externalize that process so that it can be inspected (Chi, 1997). Within a social setting, it provides a currency for social exchange as well as visible evidence of otherwise intangible social values and the processes through which they are exchanged (Bourdieu, 1991). In this paper, we are concerned with the latter situation, where language is the visible multi-dimensional manifestation of interaction between individuals, with cognitive, motivational, and relational aspects.

In recent years, there has been a growing interest in assessment of collaboration (PISA, 2015), and in particular, assessment of collaborative processes visible through discussion (Weinberger & Fischer,

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

2006; Strijbos, 2011). It has been argued that discussion provides one of the best windows into affective and motivational processes (D’Mello & Graesser, 2012), but it should be acknowledged that it comes with specific challenges as well. The related fields of educational data mining (EDM; Baker & Yacef, 2009) and learning analytics (LA; Siemens & Baker, 2012) have demonstrated success at modelling learning-relevant processes and even using these models to optimize technological support for learning (Koedinger, Brunskill, D’Mello, Pardos, & Rosé, 2015; Koedinger & Aleven, 2007). A key foundational component of this work is a formalization of learning objectives into knowledge components and of identification of relevant observable learning behaviours (Koedinger, Corbett, & Perfetti, 2012). Structured problem solving activities, such as in math class, provide an ideal context in which to demonstrate the success of such an approach since they afford the opportunity to control the sequencing and timing of opportunities to observe the evidence of acquired skills. Learning may frequently take place either in highly structured or far less structured environments, such as collaborative design or learning in capstone projects. One goal of the emerging area of discourse analytics is to measure learning-relevant processes as revealed through discourse (de Liddo, Shum, Quinto, Bachler, & Cannavacciuolo, 2013; Fergusson, de Liddo, Whitelock, de Laat, & Buckingham Shum, 2014). One major challenge, however, is that as the environment becomes less structured, the ability to control the sequencing and timing of opportunities to make specific observations is correspondingly reduced. Furthermore, there is a well-known complex interplay between cognitive, motivational, and relational variables even in settings involving only a single student. As the scope of the context expands from including a single student to multiple students, this complex interplay becomes still more complex and the ability to control timing and sequencing of the opportunity to observe growth is further reduced. This interplay may sometimes interfere with measurement, for example, when a student declines to perform a skill not due to a lack of ability but due to self-consciousness in a social setting.

This paper assumes that discourse analytic work, consistent with other areas of LA, begins with identification of theoretical constructs of interest, operationalization of those constructs so that they can be measured, validation of the measurements, and finally application to data and interpretation of that application. Analytic technologies applied to language data have demonstrated a certain level of success at estimation of cognitive, motivational, and relational constructs from observed language behaviour in interactive settings (McLaren et al., 2007; Rosé et al., 2008; Erkens & Janssen, 2008; D’Mello & Graesser, 2012). Nevertheless, there is a growing awareness of the extent to which context specificity of models threatens the predictive accuracy of their application across contexts (Mu, Stegmann, Mayfield, Rosé, & Fischer, 2012). The ability to analyze far larger quantities of data than would be possible by hand in part excuses the limited accuracy of state-of-the-art models. Basic research on the development of analytic tools works towards improvement of the accuracy of automated measurements from discourse data, external validation of these measurements, and application of measurements to theory building (e.g., Gweon, Jain, McDonogh, Raj, & Rosé, 2013). From a methodological perspective, however, an important question to ask is this: Given the acknowledged interplay between cognitive, motivational, and relational factors, what accuracy should we expect as an

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

upper bound on what can be achieved, both within settings and across settings? Due to the inherent inability of a computational model to detect precisely where it is failing to generalize to a new data point, it is not possible to answer this question through computational modelling alone. And thus we provide in this article an integration across studies with accompanying reflection.

The purpose of this article is to highlight these questions and challenges and suggest an agenda for the development of practices within the LA field for rigorous application of discourse analytic approaches as well as to suggest directions for addressing the issues through improvement of the approaches themselves in future research. The critical orientation encouraged by this article could be thought of as pointing towards a *verbalization theory* as found in the area of verbal protocol analysis (van Someren et al., 1994). Within that careful methodology, the purpose of a verbalization theory is to specify the limits of what can be expected to be inferred from a protocol analysis and under what circumstances. It offers caveats that should be taken into account whenever such a methodology is used.

To highlight the challenges at hand effectively, we use as an exemplar a specific process analysis framework named SouFLé that has been featured in handbook chapters defining the area of discourse analytics (DA) from multiple research communities. These communities include learning sciences (Howley, Mayfield, Rosé, & Strijbos, 2013), formal assessment (Rosé, Howley, Wen, Yang, & Ferschke, in press), computer-supported collaborative learning (CSCL; Howley, Mayfield, & Rosé, 2013), educational technology (Rosé, 2012), learning analytics (Rosé, in press-a), and learning sciences (Rosé, in press-b). SouFLé¹ is a three-dimensional categorical coding scheme, including a cognitive, motivational, and relational dimension. In this article, we integrate findings from five years of development and application with the specific purpose of identifying types of interference between dimensions that should be considered when interpreting the results of computational models that make similar measurements. In this work, manual application of definitions of coding categories are used as the basis for computational modelling. This method is used in order both to identify kinds of examples that are out of scope for automated detection, and to identify situations where language data fails to provide the opportunity to observe variables along one or more of the three dimensions. While this analysis does not provide a quantitative upper bound on the performance of computational discourse models, it takes an important step in that direction by offering the basis for approaching automated discourse analyses with an appropriate level of skepticism that is the earmark of rigorous empirical work.

In the remainder of the paper, we first describe each of the three dimensions of the SouFLé framework in order to offer the definitions of the strands we will then discuss as intertwining. In each case, we will highlight progress towards computational modelling of each dimension as well as present findings from application studies that illustrate specific, concrete instances of the more general issues raised above. This recounting offers some validation of the value of the constructs as separate strands. Then we

¹ SouFLé was named in such a way to highlight its close connection with Systemic Functional Linguistics.

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

discuss evidence from an integration across studies, showing that these strands are in fact intertwined. Finally, we address conclusions from this integration across studies and propose an agenda for work in the area of discourse analytics (DA) to work towards disentanglement in future work.

2 SOUFLÉ: ASSESSMENT OF COLLABORATIVE LEARNING IN SYNCHRONOUS COMMUNICATION

Howley, Mayfield, and Rosé (2013) first introduced the SouFLé framework as a linguistic analysis approach for studying small groups. The intention was to define contribution level codes in terms of basic language processes without reference to theoretical constructs specific to a particular theory of learning or collaboration, but instead grounded in linguistics (Martin & Rose, 2003; Martin & White, 2005) and very broadly accepted learning-relevant constructs from the learning sciences, such as transactivity (Berkowitz & Gibbs, 1983; Teasley, 1997; Suthers, 2006; Resnick, Asterhan, & Clark, 2015). More specifically, the aim was to provide a neutral way of describing collaborative processes that might serve as a boundary object for researchers from different theoretical perspectives within the learning sciences. Here we define its cognitive, motivational, and relational dimensions in turn (Strijbos, 2011; Howley, Mayfield, Rosé, & Strijbos, 2013; Howley, Kumar, Mayfield, Dyke, & Rosé, 2013). The cognitive dimension is the one most closely connected with learning processes (Berkowitz & Gibbs, 1983; Weinberger & Fischer, 2006). It is designed to identify contributions that can be considered signposts for sociocognitive conflict. The other two dimensions are meant to trace social positioning processes within conversation that move learners in and out of an appropriate social proximity to one another for the purpose of facilitating engagement in the valued sociocognitive processes highlighted by the cognitive dimension.

In this section, we begin with an example analysis that will make the coding approach concrete. Next, we will explain each of the three dimensions in turn. As we present results and findings related to each dimension, we will illustrate both how each dimension has support, and yet how the integration across studies begins to point to the problem that these dimensions are intertwined. This lays the foundation for the next section where the intertwining is addressed more directly.

2.1 Example Analysis

We illustrate the three dimensions of SouFLé using an example discussion from a collaborative design task in an undergraduate thermodynamics course. Codes for each of the three dimensions are indicated, but not precisely defined until the sections that follow. Reading through the example in Table 1, we see that all three dimensions point to a shift in positioning of the speakers in the conversation in the second half. A tutorial dialogue agent acts as a facilitator in the discussion. The agent, referred to as Dr. Bob, begins as the authoritative source of knowledge, but near the end of the dialogue, student sa08 and sa04 have begun to position themselves as more authoritative, as can be seen in the negotiation dimension, which is used to compute an authoritativeness score for each student. In the

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

authoritativeness coding scheme, the issue is a speaker’s positioning with respect to being a source either of action or knowledge to contribute in a discussion. Since positioning with respect to authoritativeness can be thought of as positioning higher or lower within a social hierarchy, we refer to this as vertical positioning. Next is the relational dimension, which offers a coding scheme referred to as engagement. This is used to measure the communicated openness to other perspectives in framing assertions. The initial discussion consists of contributions that would be coded on the relational dimension mostly as given, and thus not up for negotiation. The issue here is the acknowledgement (or not) that a proposition could be thought of in more than one way. We will refer to this style of assertion as monoglossic if it does not acknowledge that the proposition can be shared. We refer to this as horizontal positioning because in acknowledging that a proposition may not be shared, it becomes more comfortable for others to disagree, and thus come to a more intimate positioning. In the example with Dr. Bob, these two dimensions shift in parallel over the course of the interaction. As the students begin to position themselves more authoritatively, they also acknowledge the possibilities for other views of the knowledge they share. This more open style is referred to as heteroglossic in contrast. Just as the students’ behaviour shifts along these two dimensions, they also begin displaying more reasoning and building on previously contributed ideas. These are the two requirements that define what are called transactive contributions. Each of the three dimensions of SouFLé is designed to probe a separate dimension of collaborative learning; however, as we discuss below, our finding is that these threads influence one another and are therefore not fully separate even if their operationalizations refer to conceptually separate constructs. In particular, in this example, we thus observe in the social positioning shifts conducive to productive exchange in the cognitive dimension.

Table 1. A sample dialogue between two students and an automated tutor (Doctor Bob), coded for our three SouFLé dimensions of negotiation, engagement, and transactivity

#	Author	Text	Negot	Engagement	Transa
1	Dr Bob	Hi — I’m Doctor Bob, your instructor for today’s lab.	K1	Monoglossic	
2	sa04	Well bob, I hope you don’t mind that I’m actively hostile to robots.	K1	Heteroglossic Expand	
3	Dr Bob	You’re working together, but you’ve each got a different design goal — it’s at the top of your worksheet.	K1	Monoglossic	
4	Dr Bob	Take a moment to share your goal with your partner.	A2		
5	sa08	most environmentally friendly	K1	Monoglossic	
6	sa04	my goal is to maximize power. yours?	K1/K2	Monoglossic	
7	Dr Bob	To help you meet your design goals, we are going to discuss some of the parameters given on the worksheet.	K1	Monoglossic	
8	Dr Bob	Let’s start with Tmax.	A1		

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

9	sa08	so basically, we comprmise on greeness and power	A1	Monoglossic	
10	Dr Bob	The heat rejected by the cycle increases by increasing Tmax.	K1	Monoglossic	Ext
11	Dr Bob	We see this in Graph 3 of the worksheet.	K1	Heteroglossic Expand	
12	Dr Bob	BTW: Is it safe to keep increasing Tmax?	O		
13	sa08	No	K2	Monoglossic	
14	sa04	Nope. You'll blow out your turbine	K2	Monoglossic	Ext
15	Dr Bob	Good. Material properties constrain the maximum temperature we can use in a cycle. For our cycle, Tmax cannot be more than 570°C	K1	Monoglossic	T
16	Dr Bob	Try using this understanding that our team now shares to come up with a potential value of Tmax (T @ S2) that will help you meet your objectives.	A2		T
17	sa08	We don't want it to be at 570 both for the material and	K1	Heteroglossic Expand	T
18	sa08	well, for power and efficiency, we want a high tmax, but environmentally, we want a lower one. It has to be higher than 410 for steam quality	K1	Heteroglossic Expand	T
19	sa08	so somewhere between 410 and 570	K1	Monoglossic	
20	sa08	what about right in the middle, what about 500?	K1	Heteroglossic Expand	
21	sa04	seems reasonable	O	Heteroglossic Expand	
22	sa08	We choose 500 degrees C	A1	Heteroglossic Expand	
23	sa04	however, environmental friendliness can be increased by either increasing efficiency or by reducing waste heat, so maybe it's better to just max out our temperature.	K1	Heteroglossic Expand	Ext

2.2 Modelling Approach

Our approach to automating collaborative process analysis has focused on the development of reliability in our coding schemes to provide training and test sets for supervised machine learning models. In our methodology, we revalidate our coding schemes each time we move on to a significantly different domain or student population. This enables us to notice whether the definitions of codes become less appropriate due to the changes in contextual factors. It is well known that predictive models fall prey to over-specificity to the contexts in which the models were trained. A whole area of machine learning research referred to as domain adaptation or multi-domain learning focuses specifically on addressing this issue (Daumé, 2007; Joshi, Dredze, Cohen, & Rosé, 2012), and yet it is far from a solved problem. An important aspect of our computational work, highlighted in the sections below, is that we have made

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

extensive use of a wide variety of text features made available through the LightSIDE toolbench (Mayfield & Rosé, 2013) and other means, including a wide variety of structural features that make heavy use of part-of-speech, word categories, syntactic structure, and local rhetorical structure.

This approach to revalidation for each new corpus is different from some other automated linguistics methods commonly employed to analyze discourse where a set of all-purpose scales may be provided based on a specific corpus and then applied to every corpus. Examples include Linguistic Inquiry and Word Count (LIWC; Chung & Pennebaker, 2014) and Coh-Metrix (McNamara, Graesser, McCarthy, & Cai, 2014).

LIWC analyzes a body of text and outputs the degree to which the author expressed a variety of psychological processes, relativity, and personal concerns. Psychological processes commonly include positive and negative emotions, sensory processes, as well as some cognitive (i.e., causation, insight, certainty) and social (i.e., family, friends) processes. Each category includes a collection of words that count towards the sum of that category. For example, positive emotion words include happy, pretty, joy, win, and many more. However, without a more complex linguistic and discourse model, such counts have a tendency to confuse words with more than one meaning. For instance, the word “pretty” may be used both as an adjective, but also as a modifier to an adjective or as an adverb as in “pretty sure” or “pretty bad.” This becomes a larger concern when looking at considerably more informal discourse, such as collaborative learning conversations. Furthermore, while LIWC has a social processes category, the words included are nouns and verbs used to reference communication or people (i.e., mom, co-worker, man). Interpersonal communication contains a range of nuanced linguistic behaviours indicating social positioning and other processes that a generalized grouping for “references to family” will not provide satisfactory insight.

A more sophisticated computational linguistic model might be better suited for the complexity of conversational discourse, and one such system is Coh-Metrix. Coh-Metrix is a more theoretically grounded system, which among other uses automatically selects and scopes reading materials for students based upon their abilities (McNamara et al., 2014). It has also been applied to the analysis of collaborative learning (Dascalu, Trausan-Matu, McNamara, & Dessus, 2015). The system is built upon a multilevel framework to improve student reading through exposure to linguistic features that have been closely associated with deeper comprehension, rather than basic reading comprehension. These five levels include words (such as grammatical categories and word frequency), syntax (via sentence level structure), text base (diversity of vocabulary and referring to objects through pronouns, etc.), situation model (computed as aspects of cohesion), and genre & rhetorical structure (referred to as narrativity) (Graesser & McNamara, 2011). The system employs third-party tools such as Latent Semantic Analysis and WordNet as part of its approach.

Coh-Metrix extracts a comprehensive set of linguistic features at many levels, but it is important to note that it is not comprehensive in terms of the types of linguistic structures that may be relevant for a

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

specific process analysis task on each of those levels, especially at the situation level. Careful application of these train-once approaches (e.g., through hierarchical models that account for variation related to non-independence between data points and systematic variation due to subpopulation variables), guard against some issues with spurious correlations and domain transfer problems discussed above. However, not all researchers utilizing these tools have applied them in these careful ways. An equally worrisome situation is that validations of the inferences from such frameworks in corpora where they are applied are rarely provided. These are important limitations not in the resources themselves but in how the community of LA and related fields have taken them up.

2.3 Cognitive Dimension

While the SouFLé framework draws heavily from linguistics, the cognitive dimension of SouFLé is distinct from the social and motivational dimensions in that its definition is not strictly linguistic. However, the values underlying the construct of transactivity (Berkowitz & Gibbs, 1983) are not controversial.

2.3.1 Operationalization

The simple idea behind the concept of transactivity is a value placed on making reasoning explicit and elaborating on previously expressed reasoning. Transactive contributions build on or evaluate instances of expressed reasoning that came earlier in the discussion. The unit of analysis adopted in SouFLé was first established for analysis of a related construct referred to as Social Modes of Co-construction (Weinberger & Fischer, 2006). In particular, one unit is the minimal amount of text required to express reasoning. In the Weinberger and Fischer formulation, this is enough text to express a connection between some detail from the given task (which in their case is the object of the case study analyses their students are producing in their studies) with a theoretical concept (which comes from the attribution theory framework, which the students are applying to the case studies). When analyzing for transactivity, researchers segment linguistic contributions based on this unit of analysis. When the researchers have seen enough text that expresses a case study detail, a theoretical concept, and a connection between the two, they place a segment boundary. The simple way of thinking about what constitutes a reasoning display is that it has to communicate an expression of some causal mechanism or express an evaluation or comparison. The basic premise was that a reasoning statement should reflect the process of drawing an inference or conclusion using reason.

Statements that display reasoning can be coded as either externalizations, which represent a new direction in the conversation, or transactive contributions, which operate on or build on prior contributions. In our distinction between externalizations and transactive contributions, we have attempted to take an intuitive approach by determining whether a contribution refers linguistically in some way to a prior statement, such as using a pronoun or deictic expression, or using clearly related ideas.

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

2.3.2 Computational Modelling

In our prior work, we developed and applied machine learning techniques for automatic analysis of transactivity in discussion forums (Rosé et al., 2008), chat transcripts (Joshi & Rosé, 2007), transcribed group discussions (Ai, Sionti, Wang, & Rosé, 2010), and speech recordings of dyadic discussions (Gweon et al., 2013). When we attempt to build computational models of this and other dimensions, we learn from inspecting the models we build from our data, and those insights contribute back to our understanding of the constructs themselves.

Transactivity requires identifying instances building on prior contributions. Thus, it is not surprising that we identified that features computed to measure commonality between a new contribution and those of different speakers contributed previously in the conversation improve predictive accuracy. This is not surprising given that the definition of transactivity refers to integration of or connection between the ideas of different speakers. What this means, though, is that approaches that attempt to predict whether a contribution is transactive or not by extracting features for prediction only from the contribution itself, without reference to the prior context, will be less successful than those that leverage context. A simple way of applying the idea is to include one or more features that represent evidence of connection between a turn and earlier turns by other speakers, such as a measure of lexical cohesion between the current turn and previous turns contributed by different speakers in the same thread (Rosé et al., 2008).

Another way of leveraging interconnectedness between the turns of speakers is to monitor a specific sociolinguistic process in a discussion that suggests an effort to make such connections. Here we use as our example an effort to model speech style accommodation using unsupervised Dynamic Bayesian Networks (Jain, McDonogh, Gweon, Raj, & Rosé, 2012) as one step towards automating analysis of transactivity in speech (Gweon, Jain, McDonogh, Raj, & Rosé, 2012; Gweon et al., 2013). Research on speech style accommodation has found that conversants may shift their speaking style within an interaction, becoming either more similar or less similar to one another. By examining speech style accommodation as a social cue, we can better determine if conversational participants are working to build common ground with one another, which should also be reflected in the prevalence of transactive statements building on others' ideas (Gweon et al., 2013). Indeed, our work has shown that our automatic measures of speech style accommodation are significantly positively correlated with other-oriented transactive statements.

The concept of transactivity originally grows out of a neo-Piagetian theory of learning where this conversational behaviour is said to reflect a balance of perceived power within an interaction. Earlier research in the area of speech style accommodation suggests that it should be possible to find evidence of power differentials as well as adjustments in these differentials through shifts in language usage patterns. It can be expected, then, that linguistic accommodation would predict the occurrence of transactivity. Therefore, language representation for evidence of such language usage shifts should be useful for predicting occurrences of transactivity.

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

This hypothesis has been confirmed through a demonstration that speech style accommodation as measured by the Jain et al. (2012) unsupervised model has a significant positive correlation with the prevalence of transactive contributions (Gweon et al., 2012). The authors examined pairs of undergraduate students engaged in a debate about the fall of the Ottoman Empire. Externalizations and transactive statements were manually coded in the transcribed dialogues. Speech style accommodation was measured by the model described in Jain et al. (2012), but focused on prosodic features such as pitch, energy, and speaking rate. In these debates, our unsupervised approach to measuring speech style accommodation correlated with the manual transactive codes ($R=.4$; Gweon et al., 2012). This positive correlation is on the one hand interesting and useful from the standpoint of automated assessment. However, it comes with the troubling side effect that if the prevalence of observed transactivity is influenced through social factors, then we cannot rely on our observation of its occurrence in interaction as a verifiable assessment of ability to produce that form of argumentation. Students may be fully capable of the reasoning and articulation skills required to produce transactive contributions but may simply refrain from doing so because of the social environment.

2.3.3 Findings

As expected, our work on analysis of transactivity in connection with learning is consistent with prior work (Joshi & Rosé, 2007; Teasley, 1997). Beyond its usual role as a mediating variable related to sociocognitive conflict and learning, in a lab study representing an assembly line task we confirmed that it is also associated with effective knowledge sharing when newcomers join a new working group (Gweon et al., 2011). Nevertheless, as acknowledged above, we see evidence of the intertwining of cognitive and social factors here. In the example above, social positioning was associated with a conducive environment for transactivity; below we will see instances where the opposite is the case.

2.4 The Motivational Dimension

The motivational dimension in SouFLé is meant to capture conversational behaviour that reflects the self-efficacy of students related to their ability to participate meaningfully in the collaborative learning interaction (Howley, Mayfield, & Rosé, 2011).

2.4.1 Operationalization

This dimension is rooted in Martin and Rose's (2003) Negotiation Framework, from the systemic functional linguistics community. We use codings at this level to compute a relative authoritativeness score for students within an interaction. Because of this, we sometimes refer to this coding scheme as the Authoritativeness Framework.

This coding highlights the moves made in a dialogue that reflect the authoritativeness with which those moves were made, and gives structure to exchanges between participants. Our formulation of the Authoritativeness Framework is comprised of two axes with six and three codes, respectively, and incorporates structural and pragmatic knowledge of language. At its core for flows of knowledge are two

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

moves in particular. The first is K1, or “primary knower,” and the second is K2, or “secondary knower.” A “primary knower” move includes a statement of fact, an opinion, or an answer to a factual question, such as “yes” or “no.” It only counts as “primary knower” if it is not presented in such a way as to elicit an evaluation from another participant in the discussion. In other words, the speaker is positioned as the source of this knowledge in the flow of conversation. Conversely, a “secondary knower” move includes statements where the speaker is positioned as recipient and therefore not the authoritative source of this knowledge. This occurs when asking a question eliciting information, or presenting information in a context where evaluation is the expected response (such as when it is formulated specifically to elicit feedback). In flows of action, there are corresponding moves of A1, or “primary actor,” and A2, or “secondary actor.”

There is no strict form–function relationship between these codes and the text being analyzed. The simplest example of this is a line such as “yeah,” which could be authoritative in response to a question or could be non-authoritative in response to someone else’s evaluation. Additionally, factual statements where the speaker is uncertain of the correctness and is explicitly looking for approval from a listener would be coded as a K2 move, even if structurally similar to most K1 moves. The roles that speakers take in enacting these codes can shift rapidly within a conversation, and are dynamic, being heavily based on the context of what has happened leading up to an utterance, and how that utterance is responded to by other participants.

When the Negotiation Framework is applied to a corpus, each turn gets one code. Sequences of codes form flows of information or action within an interaction. Each complete flow contains exactly one primary core move and at most one secondary core move. Other preparatory and follow up moves may be included. While the source of each flow of knowledge and action in a conversation is negotiated locally, the overall level of authoritativeness in a person’s stance is related to the proportion of time during which the speaker adopted an authoritative (i.e., primary knower or primary actor) stance. Thus, in order to compute an authoritativeness score for a student within an interaction, we first count the number of flows that student participated in. We then count the number of these flows in which the student contributed the primary core move, which positions them as the source within that flow. The proportion of authoritative source moves over total number of flows is the student’s authoritativeness score.

2.4.2 Computational Modelling

Application of this dimension has been automated in synchronous chat environments (Howley, Adamson, Dyke, Mayfield, Beuth, & Rosé, 2012), transcribed doctor–patient interactions (Mayfield, Laws, Wilson, & Rosé, 2014), and transcribed collaborative discussions (Mayfield & Rosé, 2011).

In our computational work (Mayfield & Rosé, 2011), we draw insights from the theoretical foundation for the coding scheme that imposes sequencing constraints on patterns of codes within an interaction. While the codes are assigned to individual contributions in a conversation, we are able to encode the

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

sequencing constraints within an Integer Linear Programming framework. The best performing model included these constraints imported directly from the theory foundation for the coding scheme, and significantly outperformed an otherwise equivalent model without the constraints. The model achieved high correlation with authoritativeness ratings from human assigned codes in a corpus of direction giving dialogues ($R=.97$) as well as a corpus of doctor–patient interactions ($R=.96$). Work on automation of this coding scheme has been one of our strongest demonstrations of the application of insights from linguistics for computational modelling at the discourse level.

2.4.3 Findings

In our prior work, we have seen correlations between self-report measures of collective self-efficacy from collaborative groups and measures of authoritativeness of stance derived from our coding in this dimension (Howley et al., 2012). On this dimension, we consider that an authoritative presentation of knowledge is one presented without seeking external validation for that knowledge.

One of our long time such efforts has been using our negotiation coding as a way of estimating self-efficacy in collaborative learning encounters (Howley et al., 2011). We have already described how we are able to use our negotiation coding to assign an authoritativeness measure to students by counting the number of flows of information or action within an interaction in which they are positioned as the source. This enables us to transform the turn-by-turn coding into a scale. In transforming the pattern of codes to a scale, we are then able to examine the extent to which this positioning on the vertical social dimension correlates with extra-linguistic variables. We initially expected to see positive correlations between authoritativeness and extra-linguistic variables associated with a value placed on capability in connection with the specific knowledge and action associated with the threads used in the computation. Our initial interpretation suggested that we could leverage authoritativeness as a potential behavioural measure of academic group self-efficacy. However, application of the same coding scheme to data in strikingly different contexts challenges an overly simplistic interpretation of the significance of the authoritativeness rating.

For example, authoritativeness correlates both with domain related academic self-efficacy and learning in collaborative problem solving settings (Howley et al., 2011; Howley et al., 2012). This relationship is reasonable since the ability to provide knowledge and act in task-relevant ways is what academic self-efficacy measures in these contexts, and the tasks are designed in such a way that meaningful task engagement is meant to produce learning. What is even more interesting is that it also sheds light on the interplay between social and cognitive factors in learning, and points to opportunities for impacting engagement in important learning behaviours by addressing social problems such as bullying (Cui, Chaudhuri, Kumar, Gweon, Rosé, 2008; Howley et al., 2012). In this work, we saw that students respond to aggressive behaviour by reducing their level of authoritativeness in an interaction. At the extreme end of the spectrum, this reduced authoritativeness resulted in a reduction of learning-relevant responses to impasses in problem solving, and ultimately a reduction in learning. While it would be possible to explain this reduction in learning through a purely cognitive means, exploring the situation

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

more broadly in terms of both social and cognitive factors, we see that the reduction in learning-relevant behaviours from a cognitive perspective had a social cause. In this context, authoritativeness is a reflection of a student's estimation of their ability to contribute to the joint problem solving. In the absence of such confidence, a student would reasonably abdicate to the student deemed more capable. This anticipated correlation between authoritativeness and self-efficacy appears in additional work (Howley et al., 2011; Howley et al., 2012).

It is consistent with this interpretation to expect different correlations in contexts where the expectations associated with task roles are different, such as in doctor–patient interactions where the doctor is expected to have special knowledge not possessed by the patient. As an evaluation of the predictive validity of our authoritativeness metric in a health context, we have applied the authoritativeness metric to the analysis of doctor–patient communication (Mayfield, Laws, Wilson, & Rosé, 2014). We measured the predictive validity of this metric in connection with validated measures related to trust in doctor–patient communication. In particular, we tested five specific trust-related constructs selected by colleagues at Brown University who specialize in trust in doctor–patient communication. We determined that over a corpus of 450 doctor–patient interactions paired with questionnaire data, four out of five constructs were significantly correlated with authoritativeness, with R-values ranging from .25 to .35 using authoritativeness scores computed from hand-coded negotiation codes. A construct related to patient health efficacy from the same questionnaire data did not correlate with patient authoritativeness, which is expected in this context since the role of patient comes with different expectations regarding expertise than a collaborative problem solving session.

In addition to providing the basis for the authoritativeness scale, the negotiation codes more generally have been valuable for structuring multi-threaded conversational interactions in preparation for subsequent analysis. For example, analysis of task-relevant differences in information sharing practices between military and civilian pairs performing the same task in a lab study (Mayfield, Garbus, Adamson, & Rosé, 2001; Mayfield & Rosé, 2011) as well as conversational strategies associated with stress reduction in online cancer support chats (Mayfield, Adamson, & Rosé, 2012; Mayfield, Wen, Rosé, & Golant, 2012). There we have also found that positioning with respect to knowledge transfer is predictive of stress reduction in these chats; however, it does not appear to be directly related to self-efficacy. In particular, a closely related notion is empowerment, which we have found is related to aspects of our negotiation coding, but not to the summative authoritativeness ratio.

As the connection between authoritativeness and external variables in different domains plays out differently, we realized that our original conception of the negotiation codes as representing a motivational dimension related to self-efficacy was too simplistic. Across all of the contexts, we see an explanation for its significance in terms of positioning for active contribution. But the implications of that contribution in terms of what it presupposes from the speakers and how it affects them and others appears to be quite context dependent. In our more recent work, we have characterized it more directly in terms of knowledge transfer (Mayfield, Laws, Wilson, & Rosé, 2014). However, we cannot deny that

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

within a learning context, the ability, opportunity, and success at contributing knowledge actively within an interaction has tremendous significance in terms of self-efficacy, and other constructs related to self-esteem and engagement. Here we begin to see the lines between cognitive and motivational dimensions begin to blur.

2.5 The Relational Dimension

The relational dimension in SouFLé is meant to capture the level of openness to the ideas of others communicated in a student's framing of assertions.

2.5.1 Operationalization

Whereas in the cognitive dimension we adopted an approach to identify expressions of reasoning and transactivity, in the relational dimension, we base our work on the earlier Systemic Functional Linguistic (SFL) work of Martin and White (2005), whose theoretical approach explicitly mandates not going beyond the evidence explicit in a text. The important distinction in our application of the Martin and White framework is the distinction between monoglossic and heteroglossic assertions. A monoglossic assertion is one framed as though it leaves no room for questioning. Monoglossic contributions are in contrast to those framed in a heteroglossic manner, where the assumed perspective of others is explicitly acknowledged within the framing. There are two types of contributions we code as heteroglossic: 1) one showing openness to other perspectives, which we refer to as "Heteroglossic Expand," and 2) one that explicitly expresses a rejection of some other perspective, which we refer to as "Heteroglossic Contract."

2.5.2 Computational Modelling and Findings

In our published work, we have analyzed heteroglossia in interaction analysis by hand, but not automatically. In that study, we found a significant, strong correlation between displayed openness in a discussion group and the prevalence of reasoning displays (Howley, Kumar, Mayfield, Dyke, & Rosé, 2013). In our computational work related to this dimension, we implemented a conversational computer agent such that we manipulated the style, in one condition as Heteroglossic Expand, and in another condition as Heteroglossic Contract. In the Expand condition, we observed significantly more inclination to make ideas explicit (Kumar, Beuth, & Rosé, 2011). This again highlights the importance of the intertwining of dimensions for the purpose of assessment. In this case, we see how social factors affect our ability to observe a student's ideation in a discussion context.

2.6 Reflecting on Intertwining and Looking Towards Disentanglement

Reflecting on the above discussion of the three dimensions of SouFLé, one thing learned is that although collaborative learning researchers typically think of transactivity from a cognitive perspective, at a deep level, it has social implications. Authoritativeness is not just a reflection of the impetus to contribute to a conversation, but also a reflection of a particular quantity of knowledge for which a person is willing to take responsibility. Finally, to round out this picture, we are reminded that the relational dimension of

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

SouFLé has as its strongest result its correlation with contribution of reasoning and ideas in interactions. We must conclude that while our initial goal was to separate the cognitive, motivational, and relational dimensions of collaboration for our verbalization theory, our work on computational modelling shows us how strongly intertwined these dimensions actually are.

Outside of collaborative learning linguistic analyses, there is ample evidence that the cognitive, motivational, and social dimensions of learning are intertwined as well. Baker, D’Mello, Rodrigo, and Graesser (2010) showed that student affect in individual learning situations had an impact on cognitive outcomes. That is boredom, and especially persistent boredom, is associated with poorer learning and more off-task behaviour. So, even in an individual interactive learning environment, the affective and cognitive threads are intertwined. Joksimović, Gašević, Kovanović, Adesope, and Hatala (2013) looked at the relationship of cognitive processes and language in a collaborative learning environment and discovered a similar intertwining. Their results showed that different phases of Communities of Inquiry cognitive presence (i.e., triggering, exploration, integration, and resolution) are associated with different words reflective of different thinking depths. Words of interest were determined based upon the LIWC categories such as causal (e.g., because, hence) and insight (e.g., consider, think, know) words, among others.

While our SouFLé analysis has shown us that the cognitive, motivational, and social processes are intertwined, prohibiting a simple path to achieving a verbalization theory that would license drawing simple conclusions from applications of individual dimensions, our framework has also shown its strength as a lens from which to better understand key moments in the collaborative learning process. The first step when working with a new microscope is to adjust the height of the lens above the specimen of interest. An ideal environment in which to engage in such an effort is in the midst of a multivocal analysis where researchers steeped in alternative methodologies each analyze the same dataset using their own approach, and then challenge one another’s assumptions and interpretations. We have had the valuable opportunity to see what SouFLé is able to elucidate in comparison with two alternative approaches, one much higher level (a social network analysis) and one much more detailed (a qualitative analysis) on two different data sets as part of a large-scale investigation into multivocality as a new approach for analyzing collaborative learning (Suthers, Lund, Rosé, Teplovs, & Law, 2013). In connection with one of these data sets, we also had the opportunity to contrast SouFLé with an alternative three-dimensional coding scheme.

In the first of the two data sets (Rosé, 2013), four different analytic teams analyzed data from a study where 9th grade biology students worked on a virtual lab related to diffusion (Dyke, Howley, Adamson, Kumar, & Rosé, 2013). There were two qualitative analyses, a network analytic approach, and our own SouFLé analysis. In this set of analyses, both one of the qualitative analyses (Stahl, 2013) and the network analytic approach (Goggins & Dyke, 2013) adopted a network-like representation. Another qualitative analysis (Cress & Kimmerle, 2013) took a purely descriptive approach. The issue of social positioning was the focus of the SouFLé analysis (Howley, Mayfield, Rosé, & Strijbos, 2013) as well as

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

one qualitative analysis (Stahl, 2013), and the network approach (Goggins & Dyke, 2013; Stahl, 2013). The main contrast was in terms of the focus of inquiry. Both the qualitative and network analytic analyses focused on the relative level of dominance of the participants within groups. At their alternative ends of a spectrum of zooming out and zooming in, the high level network analytic approach was able to provide a summative view of the interactions, where it was clear in the end which participant within each group had communicated the most. At the other end of the spectrum, the qualitative approach was able to provide snapshots of behaviour where domineering or dominating behaviour was vividly illustrated. Interestingly, the coding and counting approach of the SouFLé analysis was able to represent the pattern of behaviour over time. Showing the different aspects of behaviour over time set the behaviours identified as domineering in the other analyses into a different context. This process analysis eventually pinpointed aspects of the intervention that triggered a ripple effect of negativity within groups that was different from the behaviour either of the other analyses brought out as potentially problematic. Thus, SouFLé provided something of a sweet spot for challenging assumptions and interpretations.

An especially interesting contrast came out in the comparison between the two coding and counting approaches, one being the SouFLé framework and the other a three-dimensional framework by Strijbos (Howley, Mayfield, Rosé, & Strijbos, 2013) in a second data set. In this data, groups of undergraduates worked together on chemistry problems (Sawyer, Frey, & Brown, 2013). The SouFLé analysis (Howley, Mayfield, & Rosé, 2013) was contrasted both with an alternative three-dimensional coding and counting approach in addition to being compared with a network analytic approach (Oshima, Matsuzawa, Oshima, & Nihara, 2013) and a qualitative approach (Sawyer, Frey, & Brown, 2013). In this case, again, both the network analytic approach and the qualitative approach focused on similar issues, namely the contrast between conceptual and procedural approaches to problem solving. In addition, all analyses touched upon the issue of leadership within teams.

Here we see some evidence of the value of a linguistic approach in making fine-grained distinctions in terms of the social significance of language choices. In particular, when comparing the relational dimensions of the two coding schemes, we see value in the linguistic formulation of engagement, where we are able to represent more of the subtlety in how openness or closedness is communicated in language. Within both frameworks, one side of the contrast is viewed as more imposing (contracting, negative) and the other less imposing (expanding, positive). In the SouFLé framework, contributions are characterized as expanding or contracting the set of ideas that remain up for consideration. In the Strijbos (2011) framework, contributions are characterized as either enacting a positive or negative polarity. In our comparison between the two separate codings, we saw a many-to-many correspondence between these distinctions. Because of the many-to-many correspondence, it is possible (and indeed happens!) that a participant may be rated as more dominant than another in one coding scheme and the reverse in the other. The SouFLé framework characterizes the way a negative phrasing can be used to remove a hindrance to the consideration of an idea. Thus, a negative phrasing does not necessarily

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

communicate lack of openness towards group members, although it necessarily shows a lack of openness towards something. However, in terms of the relationship between speakers, it is not necessarily imposing. The subtlety with which SouFLé approaches the many layers of language choices encoded in each framing of an assertion within interaction proves its value here. We value the ability to monitor the effect of the framing of contributions on the social positioning of speakers with respect to one another in a discussion.

Similar to the experience with the first dataset, the coding and counting approaches challenged the interpretation of both the more abstract and less abstract approaches. In particular, the formalization of contributions to the conversation at the cognitive level offered the opportunity to ask what it meant for the two groups to approach the content conceptually versus procedurally. The end result was a perspective that revealed both groups engaging in a mixture of both of these foci, and in some ways the biggest difference was in the way they approached these two foci rather than an actual difference in emphasis. Similarly, the two coding and counting approaches were able to pinpoint different aspects of leadership within teams that might be relative strengths and weaknesses of different students within groups. As in the earlier set of analyses, the unique contribution of the coding and counting approaches was the extent to which they enabled viewing the nature of the collaborative process as it unfolded over time. The network approach was very adept at providing a summative view of contributions at various points in time. The qualitative approach was able to provide a blow-by-blow story describing the contrasting groups and punctuating the story with vivid images from raw data snippets. The coding and counting approaches were able to illustrate the complexity of the construct of leadership and contribution within collaborative groups.

A contrasting impression came from comparing across approaches when pinpointing pivotal moments in the collaboration. In the network analysis, a moment was called out as pivotal because a change took place in the shape of the evolving network after that time. In the coding and counting approaches, pivotal moments were called out because something in the form or content of a contribution itself was striking based on the formal definitions of the codes. In the qualitative analysis, moments were called out as pivotal if they struck the analyst as such, apart from any pre-conceived definitions. From this standpoint, we are challenged to think about ways in which all of these approaches might be wielded more flexibly to provide either a summative- or process-oriented perspective.

Overall, what we conclude is that SouFLé is most valuable in terms of visualizing a process over time, especially in terms of teasing out specific details of linguistic choices and their implications on the tenor of an interaction, as well as illustrating the interplay between cognitive, relational, and motivational dimensions of collaboration. It may be less adept than a network analytic approach at providing a bird's eye view of the summative effects of behaviours that occur over time or of providing a detailed snapshot of specific behaviours that might stand out as striking to a human analyst.

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

3 CONCLUSIONS AND CURRENT WORK

In this paper, we have described work to date related to operationalization and computationalization of a multi-dimensional framework for collaborative process analysis. We have motivated our methodological approach, operationalized each dimension, described successes where they have been achieved on automated analysis, and summarized findings. An important theme is the intertwining of cognitive, social, and motivational variables as observed in interaction, which we offer as an invitation to the broader LA community to join in the work of disentangling these dimensions in research going forward.

We began this reflection by discussing the ways in which forms of analysis of learner discourse may enable us to understand more about learning processes and assess collaborative skills. However, let us now stop and reflect on the careful way in which protocol analysis is applied in order to understand thinking, problem solving, and learning processes (van Someren et al., 1994). There is an accompanying notion of a verbalization theory that stipulates the limitations of the methodology on obtaining a verifiable perspective on such processes. Now we reflect on the idea of a verbalization theory as a way of monitoring our ability to see a verifiable reflection of the speaker's internal processes in the data we collect and analyze. The role of such a theory is to serve as a caveat of what we need to be careful of when using a lens to answer a scientific question. Since we see how intertwined the cognitive, relational, and motivational dimensions are in collaborative learning, we cannot assume that utterances in our social dialogues are purely reflective of cognitive processes, but rather are a combination of all three dimensions. This highlights the importance of not reducing to a single dimension, or characterizing problems and solutions on only one dimension. However, we see in this work evidence that including multiple dimensions does not solve the problem either. The dimensions themselves may isolate behaviours specifically related to those dimensions, and yet the distributions of codes on each dimension are related to the distributions of codes on the other dimensions because of the way those dimensions themselves are entangled with one another. The challenge that remains is in moving beyond the caveats towards solving these problems and elucidating new knowledge.

Perhaps the greatest success at isolating a single dimension has come from our work on speech style accommodation. Here the success was in isolating the social dimension of an interaction specifically in very low-level linguistic choices at the phoneme level. At this level, the manner of speaking is least influenced by the content of what is spoken. While the social dimension of interaction appears to greatly influence what we are able to view on the cognitive dimension (i.e., social considerations may inhibit display of cognitive abilities and processes), the converse may not be true. Moreover, if the social dimension does indeed turn out to be more basic in this respect, then if we can progress in our attempt to translate linguistic theory about the social implications of language choices into computational models, we may be able to at least identify the places where we can and cannot see a faithful representation of what is happening at a cognitive level. This addresses the challenge that computational models in general cannot be depended upon to identify which instances they are not

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

able to classify properly. The valuable insight here is that we may be able to identify those places where social considerations might be obscuring our view on other dimensions. We may never be able to view all that we want to see in terms of the cognitive processes at work in collaborative settings if this would require removal of social factors. However, as we further elaborate our verbalization theory, we may learn how to better set up the conditions of collaboration in such a way that we are in the best possible position to isolate those aspects of cognition we want to study, and then to interpret what we see properly.

4 ACKNOWLEDGMENTS

Special thanks to all the contributors to this work, including Elijah Mayfield, Gregory Dyke, Rohit Kumar, David Adamson, Gahgene Gweon, Mahaveer Jain, Dong Nguyen, Bhiksha Raj, John McDonough, Dan Suthers, Kris Lund, Chris Teplovs, and Nancy Law. This research was supported in part by NSF grants OMA-0836012 and ACI-1443068.

REFERENCES

- Ai, H., Sionti, M., Wang, Y. C., & Rosé, C. P. (2010). Finding transactive contributions in whole group classroom discussions. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS '10)*, 29 June–2 July 2010, Chicago, IL, USA (Vol. 1. pp. 976–983). International Society of the Learning Sciences (ISLS).
- Baker, R. S. J. d., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human–Computer Studies*, 68(4), 223–241. <http://dx.doi.org/10.1016/j.ijhcs.2009.12.003>
- Baker, R. S. J. d., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.
- Berkowitz, M., & Gibbs, J. (1983). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly*, 29, 399–410.
- Bourdieu, P. (1991). *Language and symbolic power*. Cambridge, MA: Harvard University Press.
- Cress, U., & Kimmerle, J. (2013). Successful knowledge building needs group awareness: Interaction analysis of a 9th grade CSCL Biology lesson. In D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, & N. Law (Eds.), *Productive multivocality in the analysis of group ns*. New York: Springer. http://dx.doi.org/interactio10.1007/978-1-4614-8960-3_27
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences*, 63, 271–315. http://dx.doi.org/10.1207/s15327809jls0603_1
- Cui, Y., Chaudhuri, S., Kumar, R., Gweon, G., & Rosé, C. P. (2009). Helping agents in VMT. In G. Stahl (Ed.), *Studying virtual math teams*, Springer CSCL Series, New York: Springer.

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262. <http://dx.doi.org/10.18608/jla.2016.33.12>

- Chung, C. K., & Pennebaker, J. W. (2014). Using computerized text analysis to track social processes. In T. Holtgraves (Ed.), *Oxford Handbooks Online*. Oxford, UK: Oxford University Press.
- Dascalu, M., Trausan-Matu, S., McNamara, D., & Dessus, P. (2015). ReaderBench: Automated evaluation of collaboration based on cohesion and dialogism. *International Journal of Computer-Supported Collaborative Learning*, 4(10), 395–423. <http://dx.doi.org/10.1007/s11412-015-9226-y>
- Daumé III, H. (2007). Frustratingly easy domain adaptation. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL 2007)*, 23–30 June 2007, Prague, Czech Republic (pp. 256–263). Stroudsburg, PA, USA: Association for Computational Linguistics.
- de Liddo, A., Shum, S., Quinto, I., Bachler, M., & Cannavacciuolo, L. (2013). 1st International Workshop on Discourse-Centric Learning Analytics (DCLA13), held at the 3rd International Conference on Learning Analytics and Knowledge (LAK '13), 8 April 2013, Leuven, Belgium. SoLAR.
- D'Mello, S., & Graesser, A. (2012). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems*, 2(4), 23. <https://dx.doi.org/10.1145/2395123.2395128>
- Dyke, G., Howley, I., Adamson, D., Kumar, R., & Rosé, C. P. (2013). Towards academically productive talk supported by conversational agents. In D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, & N. Law (Eds.), *Productive multivocality in the analysis of group interactions*. New York: Springer. http://dx.doi.org/10.1007/978-1-4614-8960-3_25
- Erkens, G., & Janssen, J. (2008). Automatic coding of dialogue acts in collaboration protocols. *International Journal of Computer Supported Collaborative Learning*, 3, 447–470. <http://dx.doi.org/10.1007/s11412-008-9052-6>
- Ferguson, R., de Liddo, A., Whitelock, D., de Laat, M., & Buckingham Shum, S. (2014). 2nd International Workshop on Discourse-Centric Learning Analytics (DCLA14), held at the 4th International Conference on Learning Analytics and Knowledge (LAK '14), 24–28 March, Indianapolis, IN, USA. SoLAR. <https://dx.doi.org/10.1145/2567574.2567631>
- Goggins, S., & Dyke, G. (2013). Network analytic techniques for online chat. In D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, & N. Law (Eds.), *Productive multivocality in the analysis of group interactions*. New York: Springer. http://dx.doi.org/10.1007/978-1-4614-8960-3_29
- Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3(2), 371–398. <http://dx.doi.org/10.1111/j.1756-8765.2010.01081.x>
- Gweon, G., Kane, A., & Rosé, C. P. (2011). Facilitating knowledge transfer between groups through idea co-construction processes. *Proceedings of the Annual Meeting of the Interdisciplinary Network for Group Research (INGRoup)*, Minneapolis, MN. <http://dx.doi.org/10.1007/s11412-013-9172-5>
- Gweon, G., Jain, M., McDonough, J., Raj, B., & Rosé, C. P. (2013). Measuring prevalence of other-oriented transactive contributions using an automated measure of speech style accommodation. *International Journal of Computer Supported Collaborative Learning*, 8(2), 245–265.

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

- Gweon, G., Jain, M., McDonogh, J., Raj, B., & Rosé, C. P. (2012). Predicting idea co-construction in speech data using insights from sociolinguistics. *Proceedings of the 10th International Conference of the Learning Sciences (ICLS '12)* 2–6 July 2012, Sydney, Australia.
- Hmelo-Silver, C. (2013). Multivocality as a Tool for Design-Based Research, . In D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, & N. Law (Eds.), *Productive multivocality in the analysis of group interactions*. New York: Springer. http://dx.doi.org/10.1007/978-1-4614-8960-3_30
- Howley, I., Mayfield, E., & Rosé, C. P. (2011). Missing something? Authority in collaborative learning. *Proceedings of the 11th International Conference on Computer-Supported Collaborative Learning (CSCL 2011)*, 4–8 July 2011, Hong Kong, China. International Society of the Learning Sciences.
- Howley, I., Adamson, D., Dyke, G., Mayfield, E., Beuth, J., & Rosé, C. P. (2012). Group composition and intelligent dialogue tutors for impacting students' self-efficacy. *Proceedings of the International Conference on Intelligent Tutoring Systems (ITS 2012)*, 14–18 June 2012, Chania, Greece (pp. 551–556). *Lecture Notes in Computer Science*, Volume 7315. Berlin: Springer-Verlag.
http://dx.doi.org/10.1007/978-3-642-30950-2_71
- Howley, I., Kumar, R., Mayfield, E., Dyke, G., & Rosé, C. P. (2013). Gaining insights from sociolinguistic style analysis for redesign of conversational agent based support for collaborative learning. In D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, & N. Law (Eds.), *Productive multivocality in the analysis of group interactions*. New York: Springer. http://dx.doi.org/10.1007/978-1-4614-8960-3_26
- Howley, I., Mayfield, E., & Rosé, C. P. (2013). Linguistic analysis methods for studying small groups. In C. Hmelo-Silver, A. O'Donnell, C. Chan, & C. Chin (Eds.), *International handbook of collaborative learning*. Milton Park, UK: Taylor & Francis.
- Howley, I., Mayfield, E., Rosé, C. P., & Strijbos, J. W. (2013). A multivocal process analysis of social positioning in study groups. In D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, & N. Law (Eds.), *Productive multivocality in the analysis of group interactions*. New York: Springer.
http://dx.doi.org/10.1007/978-1-4614-8960-3_11
- Jain, M., McDonogh, J., Gweon, G., Raj, B., & Rosé, C. P. (2012). An unsupervised dynamic Bayesian network approach to measuring speech style accommodation. *Proceedings of the 13th Conference of the European Association for Computational Linguistics (EACL 2012)*, 23–27 April 2012, Avignon, France (pp. 787–797).
- Joksimović, S., Gašević, D., Kovanović, V., Adesope, O., & Hatala, M. (2014). Psychological characteristics in cognitive presence of communities of inquiry: A linguistic analysis of online discussions. *The Internet and Higher Education*, 22, 1–10. <http://dx.doi.org/10.1016/j.iheduc.2014.03.001>
- Joshi, M., Dredze, M., Cohen, W., & Rosé, C. P. (2012). Multi-domain learning: When do domains matter? *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2012)*, 12–14 July 2012, Jeju Island, Korea (pp. 1302–1312).
- Joshi, M., & Rosé, C. P. (2007). Using transactivity in conversation summarization in educational dialog. *Proceedings of the SLaTE Workshop on Speech and Language Technology in Education (SLaTE2007)*, 1–3 October 2007 Farmington, PA, USA.

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262. <http://dx.doi.org/10.18608/jla.2016.33.12>

- Koedinger, K. R., & Alevan, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3), 239–264. <http://dx.doi.org/10.1007/s10648-007-9049-0>
- Koedinger, K., Brunskill, E., D’Mello, S., Pardos, Z., & Rosé, C. P. (2015). Data mining and education. *WIRES Cognitive Science*, 6(4), 333–353. <http://dx.doi.org/10.1002/wcs.1350>
- Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge–learning–instruction framework: Bridging the science–practice chasm to enhance robust student learning. *Cognitive Science*, 36(5), 757–798. <http://dx.doi.org/10.1111/j.1551-6709.2012.01245.x>
- Kumar, R., Beuth, J., & Rosé, C. P. (2011). Conversational strategies that support idea generation productivity in groups. *Proceedings of the 9th International Conference on Computer-Supported Collaborative Learning (CSCL 2011), Volume 1: Long Papers*, 4–8 July 2011, Hong Kong, China (pp. 398–405). International Society of the Learning Sciences.
- Martin, J. R., & Rose, D. (2003). *Working with discourse: Meaning beyond the clause*. New York: Continuum.
- Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Basingstoke, UK: Palgrave Macmillan. <http://dx.doi.org/10.1057/9780230511910>
- Mayfield, E., Adamson, D., & Rosé, C. P. (2012). Hierarchical conversation structure prediction in multi-party chat. *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2012)*, 5–6 July 2012, Seoul National University, Seoul, South Korea (pp. 60–69). Association for Computer Linguistics.
- Mayfield, E., Adamson, D., & Rosé, C. P. (2013). Recognizing rare social phenomena in conversation: Empowerment detection in support group chatrooms. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (pp 104–113). Retrieved from <http://www.aclweb.org/anthology/P/P13/P13-1011.pdf>
- Mayfield, E., Garbus, M., Adamson, D., & Rosé, C. P. (2011). Data Driven Interaction Patterns: Authority and Information Sharing in Dialogue, *Proceedings of the AAAI Symposium on Building Representations of Common Ground with Intelligent Agents*.
- Mayfield, E., Laws, B., Wilson, I., & Rosé, C. P. (2014). Automating annotation of information flow for analysis of clinical conversation. *Journal of the American Medical Informatics Association*, 21(1), e122–e128. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3957397/>
- Mayfield, E., & Rosé, C. P. (2011). Recognizing authority in dialogue with an integer linear programming constrained model. In Y. Matsumoto & R. Mihalcea (Eds.), *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HTL ’11)*, 19–24 June 2011, Portland, OR, USA (pp. 1018–1026). Association for Computational Linguistics.
- Mayfield, E., & Rosé, C. P. (2013). LightSIDE: Open source machine learning for text accessible to non-experts. In M. D. Shermis & J. Burstein (Eds.), *Handbook of automated essay grading*, London: Routledge.

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262.
<http://dx.doi.org/10.18608/jla.2016.33.12>

- Mayfield, E., Wen, M., Rosé, C. P., & Golant, M. (2012). Discovering habits of effective online support groups. *Proceedings of the 17th ACM International Conference on Supporting Group Work (GROUP '12)*, 27–31 October 2012, Sanibel Island, FL, USA (pp. 263–272). New York: ACM.
<https://dx.doi.org/10.1145/2389176.2389216>
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, MA: Cambridge University Press.
- McLaren, B., Scheuer, O., De Laat, M., Hever, R., de Groot, R., & Rosé, C. P. (2007). Using machine learning techniques to analyze and support mediation of student e-discussions. *Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED 2007)*, 9–13 July 2007, Los Angeles, CA, USA (pp. 331–338). IOS Press. Retrieved from
<http://repository.cmu.edu/cgi/viewcontent.cgi?article=1134&context=hcii>
- Mu, J., Stegmann, K., Mayfield, E., Rosé, C. P., & Fischer, F. (2012). The ACODEA framework: Developing segmentation and classification schemes for fully automatic analysis of online discussions. *International Journal of Computer Supported Collaborative Learning*, 7(2), 285–305.
<http://dx.doi.org/10.1007/s11412-012-9147-y>
- Oshima, J., Matsuzawa, Y., Oshima, R., & Nihara, Y. (2013). Application of social network analysis to collaborative problem solving discourse: An attempt to capture dynamics of collective knowledge advancement. In D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, & N. Law (Eds.), *Productive multivocality in the analysis of group interactions*. New York: Springer.
http://dx.doi.org/10.1007/978-1-4614-8960-3_12
- PISA. (2015). Draft collaborative problem solving framework, OECD. Retrieved from
<https://www.oecd.org/pisa/pisaproducts/pisa2015draftframeworks.htm>
- Resnick, L., Asterhan, C., & Clarke, S. (Eds.). (2015). *Socializing intelligence through academic talk and dialogue*, Washington, DC: American Educational Research Association.
- Rosé, C. P. (2012). Assessing socio-emotional learning. In R. Luckin, J. Underwood, N. Winters, P. Goodyear, B. Grabowski, & S. Puntambekar (Eds.), *Handbook of educational technology*. Milton Park, UK: Taylor & Francis.
- Rosé, C. P. (2013). A multivocal analysis of the emergence of leadership in chemistry study groups. In D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, & N. Law (Eds.), *Productive multivocality in the analysis of group interactions*. New York: Springer. http://dx.doi.org/10.1007/978-1-4614-8960-3_13
- Rosé, C. P. (in press-a). Discourse analytics. *Handbook of data mining and learning analytics*. Hoboken, NJ: Wiley.
- Rosé, C. P. (in press-b). Learning analytics in the learning sciences. In F. Fischer, C. Hmelo-Silver, S. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences*. Milton Park, UK: Taylor & Francis.
- Rosé, C. P., Howley, I., Wen, M., Yang, D., & Ferschke, O. (in press). Assessment of discussion in learning contexts. In A. von Davier, M. Zhu, & P. Kyllonon (Eds.), *Innovative assessment of collaboration*. New York: Springer.

(2016). Towards careful practices for automated linguistic analysis of group learning. *Journal of Learning Analytics*, 3(3), 239–262. <http://dx.doi.org/10.18608/jla.2016.33.12>

- Rosé, C. P., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer Supported Collaborative Learning*, 3(3), 237–271. <http://dx.doi.org/10.1007/s11412-007-9034-0>
- Sawyer, K., Frey, R., & Brown, P. (2013). Peer led team learning in general chemistry. In D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, & N. Law (Eds.), *Productive multivocality in the analysis of group interactions*. New York: Springer. http://dx.doi.org/10.1007/978-1-4614-8960-3_9
- Siemens, G., & Baker, R. S. J. d. (2012, April). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12)*, 29 April–2 May 2012, Vancouver, BC, Canada (pp. 252–254). New York: ACM. <https://dx.doi.org/10.1145/2330601.2330661>
- Stahl, G. (2013). Interaction analysis of a biology chat. In D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, & N. Law (Eds.), *Productive multivocality in the analysis of group interactions*. New York: Springer. http://dx.doi.org/10.1007/978-1-4614-8960-3_28
- Strijbos, J. W. (2011). Assessment of (computer-supported) collaborative learning. *IEEE Transactions on Learning Technologies*, 4(1), 59–73. <http://dx.doi.org/10.1109/TLT.2010.37>
- Suthers, D. (2006). Technology affordances for inter-subjective meaning making: A research agenda for CSCL. *International Journal of Computer Supported Collaborative Learning*, 1, 315–337. <http://dx.doi.org/10.1007/s11412-006-9660-y>
- Suthers, D., Lund, K., Rosé, C. P., Teplovs, C., & Law, N. (2013). *Productive multivocality in the analysis of group interactions*. New York: Springer. <http://dx.doi.org/10.1007/978-1-4614-8960-3>
- Teasley, S. D. (1997). Talking about reasoning: How important is the peer in peer collaboration? In L. B. Resnick, R. Säljö, C. Pontecorvo, & B. Burge (Eds.), *Discourse, tools and reasoning: Essays on situated cognition* (pp. 361–384). Berlin: Springer. http://dx.doi.org/10.1007/978-3-662-03362-3_16
- van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modeling cognitive processes* (Chapter 7). New York: Academic Press.
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, 46(1), 71–95.