

Open Research Online

The Open University's repository of research publications
and other research outputs

Dialogue as Data in Learning Analytics for Productive Educational Dialogue

Journal Item

How to cite:

Knight, Simon and Littleton, Karen (2015). Dialogue as Data in Learning Analytics for Productive Educational Dialogue. *Journal of Learning Analytics*, 2(3) pp. 111–143.

For guidance on citations see [FAQs](#).

© 2015 The Authors

Version: Version of Record

Link(s) to article on publisher's website:

<https://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/4532/5181>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Dialogue as Data in Learning Analytics for Productive Educational Dialogue

Simon Knight

Connected Intelligence Centre, University of Technology, Sydney, Australia
sjgknight@gmail.com

Karen Littleton

Centre for Research in Education and Educational Technology, Open University, UK

ABSTRACT: This paper provides a novel, conceptually driven stance on the state of the contemporary analytic challenges faced in the treatment of dialogue as a form of data across on- and offline sites of learning. In prior research, preliminary steps have been taken to detect occurrences of such dialogue using automated analysis techniques. Such advances have the potential to foster effective dialogue using learning analytic techniques that scaffold, give feedback on, and provide pedagogic contexts promoting such dialogue. However, the translation of much prior learning science research to online contexts is complex, requiring the operationalization of constructs theorized in different contexts (often face-to-face), and based on different datasets and structures (often spoken dialogue). In this paper, we explore what could constitute the effective analysis of productive online dialogues, arguing that it requires consideration of three key facets of the dialogue: features indicative of productive dialogue; the unit of segmentation; and the interplay of features and segmentation with the temporal underpinning of learning contexts. The paper thus foregrounds key considerations regarding the analysis of dialogue data in emerging learning analytics environments, both for learning-science and for computationally oriented researchers.

Keywords: Machine learning, natural language processing, discourse-centric learning analytics, exploratory dialogue, accountable talk, collaboration, dialogue, sociocultural theory, learning analytics

1 INTRODUCTION

Effective dialogue¹ — whether spoken or written communication between dyads and larger groups — has strong associations with learning outcomes in a variety of contexts (see, for example, the collection edited by Littleton & Howe, 2010). However, the formalized identification of such dialogue is challenging and complex for both manual and computer-supported analytic methods (Mercer, 2010). Given that effective dialogue is implicated in securing positive educational outcomes, there is a need to consider how best to support its emergence in on- and offline contexts, and provide means to research its association with other facets of learning, such as metacognition and self-regulated learning.

The emerging prevalence of often large-scale learning environments poses challenges regarding how we are to foster productive educational dialogue in these environments. New computational techniques may afford opportunities to identify productive dialogue within, for instance, computer-mediated chat or verbal interaction data generated in the context of these environments. Such techniques may also be harnessed for the provision of formative or summative assessment,

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143.
<http://dx.doi.org/10.18608/jla.2015.23.7>

pedagogic feedback, and for researching relationships between dialogue and educational outcomes. Existing research concerning the role of dialogue in learning contexts has engaged in detailed consideration regarding the representation of dialogue data and its construction in research and learning contexts (Mercer, 2010; Mercer, Littleton, & Wegerif, 2004). These issues — pertinent to our research aims, and the development of representational tools to resource the analysis of dialogue — require consideration in the adoption of this research for development of analytic techniques. Such active construction and interpretation of these representations involves consideration of how dialogues are divided or segmented, the kinds of language of interest, features of this language, and how we operationalize the relationship between dialogue and learning over time. For instance, representations that exemplify coding-and-counting strategies are likely to yield rather different characterizations than those representations concerning growth over time (Suthers, 2006). Yet despite these differences, literature from a core related field in the area (CSCL) has historically been criticized for neglecting to make units of analysis, and the associated arguments for their selection, explicit (Strijbos, Martens, Prins, & Jochems, 2006). CSCL tools (including chat systems) and developing MOOC platforms afford rich potential for the support of productive learning dialogue. There is, however, also potential for relatively crude implementations of dialogue research in environments where simplistic automated analyses of language data are available at the click of a button. It is thus imperative that the lessons of research in offline contexts should be translated and made relevant for online analytic contexts.

The intent of this paper is not to present empirical “results,” but rather to provide a bridge from established work in the learning sciences to foreground some of the theoretical, methodological, and educational implications of computer-supported analysis techniques for dialogue data. In doing so, the paper provides a conceptually driven discussion of the contemporary analytic challenges in the treatment of dialogue as a form of data. We make explicit some of the considerations required in making decisions regarding appropriate units of analysis, arguing that the rich conceptual and methodological accounts arising from the extant prior literature, should underpin such decisions, and the subsequent analytic techniques that build on them. In particular, in considering the representation of dialogue data, we focus on specific facets of the representation of dialogue as data noting their importance for computational and manual analysis. We thus orient directly to the challenge identified by Rosé and Tovaes when they argue that, “more attention should be given to the problem of representing data appropriately” (2015, p. 6). The paper builds on “middle space” work in learning analytics (Suthers & Verbert, 2013) in bringing together conceptual considerations in the learning and computer sciences using a particular dialogue-construct to exemplify our claims. To some readers in the areas of education and computational linguistics, our concerns may seem familiar. However, in bringing together accounts of these concerns, we aim to foreground productive boundary objects that cut across disciplines: “physical or conceptual entities that each tradition interprets in its own way, but that provide common referents or points of articulation to ground conversations” (Suthers & Verbert, 2013, p. 2).

Our claim is that learning analytics researchers engaged in the analysis of productive educational dialogue must necessarily consider the interactions between feature selection, segmentation, and temporality — interactions that are often neglected. For example, in the learning sciences various

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143.
<http://dx.doi.org/10.18608/jla.2015.23.7>

researchers have highlighted the underexplored nature of the temporal analysis of learning (Littleton, 1999; Mercer, 2008; Mercer & Littleton, 2007), including contexts with relatively easy access to log files such as CSCL (Reimann, 2009), where research has tended to opt for “coding and counting” strategies (Suthers, 2006), or — largely — not focused on trace data at all (Gress, Fior, Hadwin, & Winne, 2010). This indicates a gap in established work: bringing the salient educational literature to bear on developing analytic techniques and language-based datasets within online learning contexts.

We highlight three particular key considerations to foreground the complex relationship between

1. Features of dialogue, taken to be salient indicators of some class of dialogue
2. Units of analysis over which those features are labelled
3. Temporal sequencing of such segments, and the features within them, through the interaction of which, productive educational dialogue emerges

This temporal interaction is particularly salient in the context of particular types of productive dialogue in which knowledge is built through the dialogue between multiple parties, over time, and through shifting, evolving, and transactive use of language. However, the points made here are of wider application given that, wherever education is taking place, temporality is at play in notions of change from one state (not knowing) to another (having learnt something) (Edwards & Mercer, 1987). As such, we engage with a theorized, social psychological perspective on computer-supported analysis of productive dialogue. This engagement is thus far in its infancy, leading to computational models that “miss the deep, underlying structure in the data that would enable the models to generalize effectively” (Gweon, Jain, McDonough, Raj, & Rosé, 2013, p. 246) and are highly context-specific (Mu, Stegmann, Mayfield, Rosé, & Fischer, 2012). The issues we will discuss in this paper are exemplified through a consideration of the sequence of dialogue presented in Table 1.

Table 1: A typical “Initiation Response Feedback/Evaluation” (IRF/IRE) sequence

| Line | Speaker | Utterance |
|------|---------|---|
| 1 | 1 | What do you think is the cause of x? |
| 2 | 2 | I think x is caused by (gives causes). |
| 3 | 1 | Okay, great, so those causes are (summarizes causes). |

In this table, we see a sequence commonly observed in classroom contexts, which follows what is termed an “Initiation, Response, Feedback” (IRF) or “Evaluation” (IRE) pattern (Sinclair & Coulthard, 1975). Line one shows the presence of the grammatical and linguistic features of a question — inviting a response. Using these features, a coding system might label this line as a question, or initiation. Consider now if lines 1 and 2 were taken together in a single segment. If this were the case, then their shared features might indicate an initiation-response pairing. Now consider if additional information were added — if, for example, a line 0 were added, consisting of another question (perhaps “Why does x happen?”), to which line 1 responds (with a further question). Over the span of the extract we would see two questions — although it would be incorrect to see them as two separate initiations — and two statements, one of which might be seen as a response, and another an evaluation. Note that identifying features in the individual utterances across this segment, of itself, does not facilitate our understanding of the excerpt — consideration of the

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143.
<http://dx.doi.org/10.18608/jla.2015.23.7>

classes of dialogue present in each line as indicated by the features within that unit of analysis (or segment) offers incomplete information. Nor does the labelling of individual utterances, using indexical features from the set of utterances, offer insight into the general structure of the sequence and its progression. Important features of the dialogue are represented in both the sequence of utterances and in the accretion of features over the sequence — for example, the uptake of terms indicated by “x” in this example. Consider now if the three utterances had been made by the same speaker; decisions regarding whether a single “class” of dialogue would be applied over the whole sequence, or on an utterance level should be made based on an understanding of the temporal nature of the dialogue. Similarly, the presence (or absence) of backchannelling (“mm,” “yes” and so on) from other participants can play a role in dialogue, but commensurately does not necessarily *break* utterances into smaller segments over which individual classes of dialogue would be applied. Indeed, consider the addition of a fourth utterance: “So what else do we know about those causes? What else do they make happen?” Note the importance of the speaker role (1st, 2nd, or a 3rd party) in assessing such an utterance; this fourth utterance illustrates an important point: the segmentation, codification, and counting of “question and answer” exchanges may obfuscate some richer interaction. In this case, in counting exchanges such as these kinds of “spiral IRF” (Rojas-Drummond, Mercer, & Dabrowski, 2001), exchanges counting “question answer” exchanges through a narrow lens — or utterance-based-segment level — would exclude insight into such spiral IRF exchanges, which build upon each other over time and across larger segments. Thus, as we discuss further below, few would argue that the kinds of productive educational dialogue we are interested in can be identified at an utterance level. However, despite this, perhaps because of the complexity of the analytic challenge being faced, various recent tools (see for example, Clarke, Chen, & Resnick, 2014; Ferguson, Wei, He, & Buckingham Shum, 2013) in fact do have a tendency to operationalize automated-coding schemes at the utterance level in order to offer proxies for productive educational dialogue.

Stahl (2013) too notes the significance of work that explores “transactivity” — a notion of exchange between participants — in the context of interest in multiple levels of analysis, from individuals to groups. Stahl notes that transactivity as originally conceptualized has tended to be explored at the analytic level of individuals who have (differing) partial information regarding a problem that requires pooled information to solve it, rather than at the group level. This focus on individuals in collaborative contexts, in contrast to collaborative units, is common to much groupwork research. For example, in exploring transactivity, Azmitia and Montgomery’s (1993) interest was in dialogue used to build upon a partner’s explicit reasoning statements, rather than on the language used to co-construct knowledge. In the examples above, this would be indicated by one participant using the terms introduced by another. However, fundamentally, understanding co-construction depends upon understanding the ways in which the context of co-constructive episodes and the context that they create, are related.

Thus, in educational contexts we are interested in how language dynamically resources future learning and constitutes a collective tool for reasoning, and not only its significance for the current learning of individuals in isolation (Knight & Littleton, 2015). Furthermore, it may well be possible to model some meaningful indicators for particular types of dialogue. For instance, if we are able to

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

operationalize some stable and general categories and patterns in dialogue use, then we can model such patterns for an automated approach; careful selection (and omission) of features is thus important (Rosé & Tovares, 2015).

So, as we will discuss below, there are a variety of decisions to make regarding operationalization for analytic techniques. For both automated and manual analytic techniques, understanding the significance of any given utterance involves understanding the context in which it is made — including the ways in which utterances are resourced by, and resource further utterances. The importance of context to dialogue has been studied in various computer-supported contexts (see e.g., Herrmann and Kienle 2008), but as the example in Table 1 exemplifies, the role of dialogue as context-creating, for example through indexical ties, is also key, and has received relatively less attention.

In this paper, we highlight the importance of a class of dialogue known as “exploratory” or “accountable,” noted for its relationship with improved learning outcomes. Research aiming to build on the rich empirical and theoretic grounding for accountable and exploratory dialogue, needs a conceptual understanding of how such approaches might be “translated” into computer-supported analytic contexts. Thus, although we focus on a particular type of dialogue, the conceptual approach, and many of its lessons, will be generalizable to other work in developing computational approaches to prior educational research.

In the following section, we discuss those kinds of dialogue known as exploratory or accountable, which are strongly associated with positive educational outcomes. We then discuss the issues of feature selection and segmentation before (in section 3.1 Operationalizing our Feature and Segmentation Level Representations) noting the complex picture of interaction between those two facets of dialogue data. The focus of the discussion in this paper is on the selection of features at the appropriate segmentation level and the application of techniques to classify based on such analysis. We particularly highlight these steps here because they represent the core challenges, applicable across all methods — manual or computational — to formalize approaches to coding (or classifying) qualitative data.

This paper does not address the technical and pedagogic mechanisms through which support of particularly productive forms of dialogue might be instantiated. Instead, the paper highlights key considerations in the application of computational techniques to our understanding of productive educational dialogue, focusing largely on the body of research exploring productive dialogue in classroom, or free-chat based environments. In particular, as noted above, our focus is on the class of dialogue that broadly clusters around notions of transactivity, accountable talk, and exploratory dialogue — dialogue in which participants are receptive to other’s ideas, engage with each other, and build a shared understanding together through their dialogue. This class of dialogue is particularly important, in part because it has an empirically evidenced association with improved learning outcomes; but moreover, because it has a strongly *theorized* association with learning, and with the use of language to create and share meaning. The claim, rooted in sociocultural theory, is that, if our object of enquiry is the dialogue itself — both as a representation of, and tool for

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

learning — then we ought to be interested in the ways that dialogue is used to create “common knowledge”; a shared understanding built up over time by participants in a dialogue, and which is a fundamental part of learning (Edwards & Mercer, 1987; Littleton & Mercer, 2013). Thus, such dialogue is — in and of itself — educationally valuable as a metacognitive tool for both individual learning and a co-constructive tool for collaborative knowledge building (Edwards & Mercer, 1987; Littleton & Mercer, 2013). In the next section, we discuss productive educational dialogue in more depth, using this discussion to foreground the salience of “context” in understanding such dialogue and in particular for *computational approaches* to understanding such dialogue. The remaining sections of the paper discuss specific challenges regarding feature selection, segmentation, and temporality in the representation of productive educational dialogue.

2 WHAT IS PRODUCTIVE EDUCATIONAL DIALOGUE?

Wherever education is taking place, commonality — a shared perspective — is key, and dialogue is the main tool used to negotiate and create such a perspective (Edwards & Mercer, 1987). This body of shared contextual knowledge, built up through dialogue and joint action and forming the basis for further communication, has been termed “common knowledge” (Edwards & Mercer, 1987). Thus, common knowledge forms a key constitutive facet of context for speakers in a dialogue, as well as being a fundamental aspect of education in which a mutuality of understanding is crucial. Furthermore, the strong consensus among researchers is that, in a variety of contexts, productive dialogue is associated with learning (see the collection edited by Littleton & Howe, 2010) and that “Engaging children in extended talk which encourages them to ‘interthink’ and explain themselves... stimulates both their subject learning, and general reasoning skills (Mercer, Dawes, Wegerif, & Sams, 2004; Mercer & Sams, 2006; Mercer, Wegerif, & Dawes, 1999; Rojas-Drummond, Littleton, Hernández, & Zúñiga, 2010), as well as their social and language skills (Wegerif, Littleton, Dawes, Mercer, & Rowe, 2004)” (Knight, 2013, p. 1). Mercer and colleagues have extensively researched such dialogue, and developed an intervention strategy called “Thinking Together” designed to teach children to engage in constructive dialogue in classroom contexts through the teaching of particular styles of dialogue, and the use of pedagogic strategies such as “ground rules” (shared group rules for collaborative dialogue) for talk written to encourage productive groupwork.² They have highlighted a particular form of productive dialogue, which, adapting the term from Douglas Barnes’ (Barnes & Todd, 1977) original broadly individualistic description, they have termed “exploratory.” They contrast this with two other types of typically less productive dialogue — disputational and cumulative, as in

Table 2.

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143.
<http://dx.doi.org/10.18608/jla.2015.23.7>

Table 2: Mercer and Colleagues' Typology of Talk*

| Type of Talk | Characteristics | Analytic features |
|---------------|---|---|
| Disputational | Groups tend to disagree and engage in individually based decision making, with little attempt to engage with each other's ideas or pool resources. | Dialogue characterized by assertions and challenges, with short exchanges. |
| Cumulative | Knowledge is exchanged but not engaged with, with little evaluation or elaboration and a tendency to accumulation. | Dialogue characterized by confirmation of assertions, and repetition. |
| Exploratory | Groups engage with ideas critically, but constructively, with suggestions for improvements made along with challenges to ideas. Participation is from the whole group, and opinions are actively solicited. "Compared with the other two types, in exploratory talk knowledge is made more publicly accountable and reasoning is more visible in the talk." | Dialogue characterized by terms and phrases associated with engagement and explanation — for example: "I think," "because/'cause," "if," "for example," "also." |

*Adapted from Mercer and Littleton 2007, pp. 58–59.

A similar characterization of productive dialogue across a range of ages — labelled "Accountable Talk" — has emerged from the work of other researchers (Michaels, O'Connor, Hall, & Resnick, 2002; Resnick, 2001). That characterization describes Accountable Talk as encompassing three broad dimensions:

1. Accountability to the learning community: participants listen to and build their contributions in response to those of others
2. Accountability to accepted standards of reasoning: talk that emphasizes logical connections and the drawing of reasonable conclusions
3. Accountability to knowledge: talk based explicitly on facts, written texts, or other public information (Michaels, O'Connor, & Resnick, 2008, p. 283)

As with the typology of talk developed by Mercer and colleagues, the emphasis of Accountable Talk is not on learning particular subject or topic knowledge and language, but rather on learning to engage with others' ideas, and in so doing use the skills of explanation and reasoning, learning to use language as a tool for thinking and — in the terms of Littleton and Mercer (2013) — *interthinking*. As indicated in the introduction and in Table 1, understanding such dialogue involves understanding the ways the dialogue is both contextualized and context creating, as we now discuss.

2.1 The Challenge of Context

Educational researchers within the sociocultural tradition highlight the importance of dialogue not only for the representation of context but also for its creation. This distinction highlights the need to understand that context should not only be assumed from the state of the dialogue at any particular point (assuming dialogue represents context), but that we should also explore the ways in which the context emerges and changes over time as a feature of the dialogue (assuming dialogue involves the co-construction of context). Let us return to the short example given in Table 1. Even in such a short excerpt, we see the uptake of terms between speakers, building a shared understanding. We can also see that the dialogue does more than simply provide a context (“we are talking about ‘x’ now, therefore my dialogue should be in ‘x’ mode”), rather the dialogue constitutes that context in the intersubjective creation of what Kärkkäinen (2006) has termed “epistemic stances.” Again, note that such context cannot be captured by utterance-level analysis of dialogue turns, because such an analysis cannot capture the range of features — including uptake of terms — indicative of this co-construction of context over time. Thus, our first claim from the established literature is that a unit of analysis — or level of segmentation — focusing on individuals (and their individual utterances) is unlikely to capture the full potential of learning dialogue; dialogue for learning is a co-constructive enterprise.

Recently, Littleton and Mercer (2013) have discussed this complexity of common knowledge as context — noting that common knowledge is both historical and dynamic in nature. Historical or “background” common knowledge involves the kind of language use that depends on a shared common understanding being taken for granted based on some shared community. Dynamic common knowledge, in contrast, refers to the kind of common knowledge built up through dialogue and associated activities, for example repetition of keywords through conversation or the kind of “recapping” behaviours teachers often engage in at the start and end of lessons (Littleton & Mercer, 2013). To give an example, in a classroom context the “same” question (in syntactic and semantic terms) might be asked at the beginning and end of a lesson, while serving different (pragmatic) functions: in the first instance, to gauge baseline understanding and provide a reference point for the second posing of the question, which is to see how the question may now be reinterpreted. The pragmatic level of description is therefore important: syntactic or semantic levels of description can be blind to understanding what is being done in interaction, where terms or phrases are taken to have fixed meaning over time, as we began to illustrate through the short exchange in Table 1. Thus, our second claim is that static accounts of learning language encoded in fixed lexicons will fail to capture the dynamic ways in which dialogue features emerge in interaction, and change over time, in interaction (as above).

Partly due to this consideration, sociocultural researchers advocate using sociocultural discourse analysis, which emphasizes both qualitative and quantitative methods, using approaches in which — in contrast to some other qualitative methods — the quantitative data is taken to aid the understanding of the qualitative, as opposed to the converse (Mercer, 2004; Mercer, Littleton, et al., 2004). Such researchers often include excerpts of dialogue, concordance analysis, and other

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143.
<http://dx.doi.org/10.18608/jla.2015.23.7>

contextual markers such as cohesive ties in their reporting. Such techniques are drawn from “systemic functional linguistics” (Halliday, Hasan, & Christie, 1989), which assumes that types of text have contexts because they are members of a particular genre, and these contexts are revealed through the way such texts are written. Thus, context is imbued into texts at the time of writing, for example, using contextualizing key words (e.g., describing a book as a “book” compared to a “course book” raises different expectations), headings (e.g., the traditional article headings from abstract to conclusion), and positioning markers (e.g., positioning with respect to existing literature, including citations and phrases such as “we agree with”). In sociocultural discourse analysis, this assumption is adapted from that of “texts” to the co-construction of context through dialogue in which “‘context’ is created anew in every interaction between a speaker and listener or writer and reader. From this perspective, we must take account of listeners and readers as well as speakers and writers, who [co]create meanings together” (Mercer, 2000, p. 21). It is thus that sociocultural researchers may seek to understand the temporal aspects of context, as involving continuity across dialogue, by looking for repetition of words, synonyms, and ways of approaching problems to understand how “speakers can jointly, co-operatively create cohesion in... their speech” (Mercer, 2000, p. 62). Such analysis has commonly been conducted using concordance software, which facilitates the exploration of “Key Words In Context” (KWIC) by displaying words searched for in their original context (typically, showing a sub-portion or whole sentence in which the keyword is located). Through these means, locally salient features can be identified through the repetition of key words, and their changing meaning over time explored — and used to segment data — as terms are negotiated and renegotiated in their use. Thus, our third claim — combining the first two — highlights that dialogue for learning can (only) be observed in interaction, and over time, and that there are a set of methods to establish these relationships.

Our final claim provides an evident challenge: The translation of the claims above to practical steps in automating, or semi-automating analysis. There is clear, and very justifiable, interest in the application of existing educational research to online contexts, particularly where analysis and learning support may be automated. However, it is appropriate to note here that more research is needed to understand the application of such research to online contexts. For example, relatively little research has been conducted investigating the use of exploratory dialogue in online contexts (see for example, Littleton & Whitelock, 2005), with only a few studies of such asynchronous dialogue (see for example, Ferguson, Whitelock, & Littleton, 2010), and none that we are aware of in the context of large multi-party and multi-modal chat systems. Despite this, there are clear differences between face-to-face and online communication, and many online contexts provide different types of opportunities for communication. Thus, there are challenges inherent in detection of exploratory dialogue across the features that we would expect to see both on- and offline. Yet this is a complex issue — the ways in which data drives theory, and vice-versa, and the translation of research from one context to another is a challenge, but not one that we should gloss.

2.2 Computational Approaches for Analysis of Productive Educational Dialogue

One means to address the complexities of context in dialogue data is to “design in” the types of language we wish to analyze; providing opportunities for their display and their capture in ways that

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

are easy to process by machines; for example, through threading, tagging, and “linking” techniques in CSCL environments. Certainly, the element of “design” in analysis of dialogue is important; if we think that some dimension of activity is important, we should be sure to provide opportunity for the occurrence of that activity in our learning contexts. Furthermore — as shall be highlighted below — the structure of captured dialogue, both with respect to its social structure, and its data representation, is important for the deployment of analytic tools. For example, design strategies have strong potential for the development of meaningful, and enjoyable, learning activities centred on language use (Gee, 2008; Shaffer, 2008). Indeed, much work has been conducted on developing environments that support, and make available for researchers, particular types of dialogue.

While design may reduce computational difficulties (for example, by introducing threading to structure discussions), the points made in this paper with respect to the importance of context are still fundamental to understanding the dynamic features of dialogue through which learning is co-constructed. CSCL environments may be seen as complementary to such dialogue, in particular where they embody some of the systems through which exploratory and accountable dialogue are more likely to occur — the “ground rules” or guidance for production of each. It is because of these complexities that systems have been developed specifically to support particular types of formalized argumentation schema (Clark, Sampson, Weinberger, & Erkens, 2007; Weinberger, Ertl, Fischer, & Mandl, 2005; Weinberger & Fischer, 2006). However, a core consideration in such platforms is the ways in which the platform structures, or scripts, dialogue, rather than the analysis of unstructured dialogue.

Within the analysis of such unstructured dialogue, computational linguistic approaches have had some success in applying educational research to the identification of a variety of discourse markers indicative of productive dialogue (Rosé et al., 2011). For example, in analysis of “exploratory dialogue,” manual discourse analysis begins with the exploration of terms such as *for example, I think, because/cause, if, also* (Mercer & Littleton, 2007). Such markers for this sort of dialogue can be readily identified automatically in computer-mediated-communication (CMC) contexts (Ferguson & Buckingham Shum, 2011; Ferguson et al., 2013). Yet, such utterance-based approaches are relatively limited in their scope and application.

Beyond such keyword feature based methods, researchers have, for example, explored transactivity (sometimes called intertextuality) in the classification of utterances from the following: small groups (Sionti, Ai, Rosé, & Resnick, 2011; Stahl & Rosé, 2012); dyads (Gweon, Jun, Lee, Finger, & Rosé, 2011); whole-class discussions (Ai, Sionti, Wang, & Rosé, 2010); and in relation to the summarization of group discussions (Joshi & Rosé, 2007). This property of dialogue is closely associated with that of “cohesive ties” described above in which the “uptake” of terms from one speaker by another is indicative of transactivity, the definitions of which are summarized as sharing two aspects:

Namely, the requirement for reasoning to be explicitly displayed in some form, and the preference for connections to be made between the perspective of one student and that of another. Beyond that, many authors appear to classify utterances in a graded fashion, in other words, as more or less transactive, depending on two factors; the degree to which an utterance involves work on reasoning, and the degree to which an

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143.
<http://dx.doi.org/10.18608/jla.2015.23.7>

utterance involves one person operating on/thinking with the reasoning of another.
 (Sionti et al., 2011, p. 6)

An allied construct implicated in the exploratory and accountable dialogue research is “heteroglossia.” Heteroglossia (Bakhtin, 1986) is related to the multivocality of perspective, the characteristic of a text as displaying, and being open to, multiple views — a significant element of dialogic education (Wegerif, 2011). Building on Martin and White’s (2005) description of dialogic expansion (in which dialogue is positioned to allow that alternative positions are available) and dialogic contraction (in which the scope of permitted perspectives is restricted), heteroglossia has recently been operationalized in the computational linguistics context as

the extent to which a speaker shows openness to the existence of other perspectives apart from the one that is reflected in the propositional content of the assertion being made... Within our heteroglossia analysis, assertions framed in such a way as to acknowledge that others may or may not agree, are identified as heteroglossic. We describe it as identifying wording choices that do or don’t treat other perspectives than what is expressed in the propositional content of the assertion as open for consideration within the continuing discussion. (Rosé & Tovaes, 2015, pp. 10–11)

To exemplify the challenges of context to operationalizing such a construct, it is worth turning to a specific example. In the context of a science classroom task, this has been operationalized by Rosé and Tovaes at the coding level for a four-part utterance level coding scheme:

- **Heteroglossic-Expand (HE)** phrases tend to make allowances for alternative views and opinions (such as *“She claimed that glucose will move through the semi-permeable membrane.”*)
 - **Heteroglossic-Contract (HC)** phrases attempt to thwart other positions (such as *“The experiment demonstrated that glucose will move through the semi-permeable membrane.”*)
 - **Monoglossic (M)** phrases make no mention of other views and viewpoints (such as *“Glucose will move through the semi-permeable membrane.”*)
 - **Non-Assertion (NA)** phrases do not assert any propositional content. This includes questions, such as *“Will glucose move through a semi-permeable membrane?”* or fixed expressions lacking propositional content such as *“Okay.”*
- (Rosé & Tovaes, 2015, pp. 10–11)

Thus, within a set of utterances from a small group, each utterance may be coded at one of the levels described based upon its semantic and syntactic content — the words used, and the ways in which individual sentences are structured to relate to other utterances. In this example, it is worth considering the challenge of identifying particular types of dialogue in isolation from a temporal analysis, through which one might be more able to identify how, for example, phrases allowing the possibility of alternative perspectives (HE) are in fact related to the airing of such perspectives (M),

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

their challenging (HC), or questioning (NA) in more or less constructive ways. For example, the scheme cannot capture the ways in which repetition of claims, contextualized by simple re-statement, unconstructive criticism, or engaged evaluation, differ insofar as they are classified at an utterance level, nor with regard to their indexical relations. Nor is there a means to code larger sections of dialogue (exchanges of multiple utterances between speakers), whether they are constructively conducted or not.

Building further on Bahktinian ideas, a body of work has emerged using Natural Language Processing (NLP) techniques to explore notions of voice, ventriloquism, and echoes — the ways ideas are expressed, re-emitted by others, and intertwined into collective voice. This work selects identifying characteristics in an individual's dialogue, endeavouring to identify where these features later emerge in a collaborator's dialogue, and are changed and evolved over time. This analysis has thus explored the relationship of these constructs to collaborative inter-animation and polyphonic dialogue, which is made up of both coherence and divergence (see, for example, Chiru, Rebedea, & Trausan-Matu, 2013; Rebedea, Chiru, & Gutu, 2014; Trausan-Matu, Dascalu, & Dessus, 2012). In this work, the kind of productive dialogue we have outlined in this paper is described in terms of coherence — the ways in which speakers share and build common knowledge — and divergence — the ways in which speakers critique and present new ideas. The lexical markers of these constructs are described in terms of linguistic coherence, the uptake and convergence of specific terms by multiple speakers in a dialogue, and the introduction of novel voices into a dialogue, for example, as marked by shifts in a dialogue's terms.

Our intention here is not to critique the important concepts presented, nor their operationalization. Rather, it is to highlight the difficulties with coding (or classifying) particular sorts of productive dialogue — particularly where utterances are taken in isolation from the rest of a text or transcript, and to highlight some of the exemplary state-of-the-art work in this area that has sought to address the issues raised in this paper. Of course, grammatical features offer insight into indexical relations among utterances — and this is a key issue to which we return shortly. Crucially though, we highlight the complexities of both coding particular types of talk at the utterance level and using features from single utterances in order to carry out such coding. That is, we highlight the challenge of labelling individual utterances as exploratory (or otherwise) — because their productive force is not seen in individual utterances, but in segments of interactional dialogue — and of using features from individual utterances to conduct such labelling, because of the loss of information regarding the indexical features of productive dialogue in so doing.

2.3 Challenges to Developing Computational and Manual Analytic Approaches

In the introduction, we highlighted (in Table 1) a short exchange, indicating the ways in which even within a small number of utterances the meaning of each expression is related to the wider context. It is thus that analysis of utterance level features (question marks or key phrases, for example) provides only a partial lens onto the meaning of such utterances. As we describe above, the operationalization of theorized approaches to productive educational discourse is a challenge for computational techniques. Consider the following excerpt (Table 3) and brief description. Here we

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143.
<http://dx.doi.org/10.18608/jla.2015.23.7>

provide a longer example of the kind of classification problem to be addressed, in this instance by illustrating a transition from an episode of cumulative to exploratory dialogue. The excerpt below³ comes from recent research on a group of four undergraduate students who were creating a multimodal theatrical performance. In the excerpt given, the segment following line 16 was originally coded as “exploratory” in nature, marking a point at which progress begins to be made, and participants interact through their dialogue.

Table 3: Example of Exploratory Dialogue

| | Speaker | Utterance |
|----|---------|--|
| 1 | C1 | What do you think to the idea of doing the whole sound from recordings? |
| 2 | C2 | Yeh |
| 3 | C1 | dance recordings? |
| 4 | C2 | definitely |
| 5 | C1 | (build the entire thing) its just |
| 6 | C2 | But obviously manipulating |
| 7 | C1 | Yeh |
| 8 | C2 | To make some big sounds out of it ... about it. |
| 9 | C1 | Yeh |
| 10 | C1 | Mm |
| 11 | C2 | So we have enough, so long as we get enough texture and enough variety timbre it'll be alright. Like a if ee ee you know if the swing creaks its perfect, and something like that. You know? |
| 12 | C1 | Mm |
| 13 | C2 | The sound of footsteps or shuffling or arm movements, breathing |
| 14 | C1 | Exactly and I'm thinking |
| 15 | C2 | that's pretty much enough instruments |
| 16 | C1 | That was exactly what I was saying |
| 17 | TV | We've just said like cuz we cuz we may as well make the swing |
| 18 | C2 | Yeh |
| 19 | TV | And then I was like saying (as if we sort of) get like a chair and then we make that the swing |
| 20 | C1 | The chair from the boots? |
| 21 | C2 | Oh so they can start the feet go in then it can start swinging straight away? s'that what you're saying? |
| 22 | TV | could be |
| 23 | TD | She was just saying she was just meaning like make the, make the swing different by making it into like could use like an old wooden chair instead of using like |
| 24 | C2 | ok |
| 25 | C1 | yeh |
| 26 | C2 | just a bit of wood |
| 27 | TV | but then |
| 28 | C2 | I see |
| 29 | TV | We could use that, from the start, d'you know the circle of shoes that can always stay round the bottom of the swing |

(Dobson, 2012, p. 274)

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143.
<http://dx.doi.org/10.18608/jla.2015.23.7>

This excerpt, in a deeper way than the example given in Table 1, offers a reference point for a number of our claims. The excerpt illustrates that segmentation by topic is not an effective method for segmentation of exploratory dialogue; in this excerpt the topic remains broadly the same throughout, although it is possible different granularities of segmentation could be drawn out. Even if finer grained topics are identified, it is not these that mark shifts in productive educational dialogue or exploratory dialogue; nor indeed is it the use of exploratory terms such as “because,” which appear throughout the extract. In this case, we can see that what marks a change in the type of dialogue is a shift from dominance by one speaker, to the presence of multiple voices. Furthermore, these voices “take up” each other’s language — they are transactive (and are, therefore, grouped by topic — reminding us that this is still an important feature); we can thus identify two distinct *segments* (lines 1–15 and 16–29) within which we can operationalize *features* that map broadly to the typology of talk in

Table 2).

In the following sections, then, we discuss these steps, with particular reference to feature selection, segmentation, and operationalization in the context of a particular theoretical stance. In each section, we give indications of the problem being addressed, before providing some exemplifications for how this problem has been tackled, and some possibilities for moving forwards. In the final subsection here, we highlight one particular area where issues of feature selection and segmentation come together — that of “temporality” in discourse data.

In the closing sections (3.1 Operationalizing our Feature and Segmentation Level Representations and 3.1.1 Temporality: A key facet of representing data for productive *dialogue*), we discuss some specific issues around the complexity of tracking exploratory dialogue, in each case here presenting the issue around “context sensitivity,” alongside discussing the adequacy of various computational approaches for meeting these challenges. While the discussion is focused on exploratory dialogue, we suggest the points made apply more widely than that. We also note some parallels with manual coding schemes, and suggest there may be useful dialogue to be had between those working on manual coding schemes and those working in automation.

3 REPRESENTING DATA FOR PRODUCTIVE DIALOGUE

It is apparent that there are methodological challenges raised by a conceptual understanding of productive educational dialogue. Moves to develop computer-supported techniques for analysis of such dialogue can be enriched by established research in education and related fields. However,

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143.
<http://dx.doi.org/10.18608/jla.2015.23.7>

such work may need “translating” for these new approaches: the point here is to consider the suitability of various approaches for our educational context (rather than for other machine learning tasks). In this section, we outline some of the specific challenges for such computational approaches.

Perhaps the most basic means through which particular classes of dialogue may be identified is through bag-of-words, or cue-phrase approaches. Using such approaches, relatively simple automated tools can count the number of occurrences of particular terms or phrases across a corpus. Indeed, superficially, the use of concordance analysis and KWIC (mentioned above in section 2.1 The Challenge of Context) uses precisely this technique in manual analysis. However, in such manual analysis, quantitative results are taken to inform qualitative analysis (as opposed to vice-versa), and automated phrase detection is used to highlight the context in which the target phrases are used — that is, to ease the process of qualitative analysis. Although there are surface similarities, we should be cautious about glossing contrasting purposes.

A particular challenge for many classification approaches is the selection of *features* indicative of a particular class of dialogue. Except in cases in which the dialogue classes are highly static, varying little over instance or context, the difficulties of selecting features to classify dialogue for both precision (instances given a class are generally of that class) and recall (of the complete set of an instance class, most are correctly classified) is a challenge. Furthermore, the ways in which features are encoded must take account of features of the text — the ways texts are structured, multiple-parties in dialogues are represented, tense and other semantic-pragmatic features are indicated, and so on.

As such, any classifier that does not function as an online classifier (i.e., updating at each new message to build in further information on local context) is unlikely to be able to build an adequate model of the fine (e.g., exchange level) and coarse (e.g., session level) salient features that constitute and are constituted in dialogue contexts. This concern is particularly salient where a classifier might be trained (or designed) on a set of features from one context, and then deployed in another.

To give an example, the k-nearest neighbours (knn) approach described by Wei et al., (2013) obtains new features by using a set of human-coded instances, extracting new features from instances close to those classified as “exploratory” in terms of topical information (i.e., not in terms of temporality), with those features that improve the classification accuracy being added to the original feature set of keyterms. An advantage of this model is that — within the constraints of the session topic — it can build in local topical salience while also utilizing our pre-existing knowledge regarding the sorts of cue-phrases we expect to see in exploratory dialogue instances. However, while such approaches are highly effective for the observation of highly stable term-frequencies and their mapping to classes, they are not effective where topics change frequently, terms are imbued with new meaning over the course of a dialogue, or terms might be related to differing classes of dialogue depending on other contextual features.

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

In addition, in focusing on the utterance level, surface level features — such as who is speaking, and the specific terms they use — are the only source of information. Such an approach foregrounds these surface features at the cost of more contextually salient features — such as whether a single individual is speaking over a given period — and these locally defined (at the level of the exchange, not the transcript) features remain hidden to the classifier. Such surface features do not provide the sort of contextual cues that ‘tie together utterances into interactive exchanges for co-construction; for such ties, we must turn to concepts such as transactivity and cohesive ties.

While some interactional features may be encoded using grammatical features such as pronoun use (for example, “‘your’/‘my’/‘our’ idea”) to explore the development of ideas (see, for example, Thompson, Kennedy-Clark, Kelly, & Wheeler, 2013), or through social network analysis, which explores who is interacting with whom (see for example, Rosen, Miagkikh, & Suthers, 2011), such analysis in the context of feature selection can only add features at the analysis level (for example, the utterance level, or as we discuss below, the broader segment) thus its utility in describing the co-constructive features of the dialogue is limited. A fundamental consideration here is the ways in which context is both built up through, and represented in the dialogue. Given this theoretical association with the kind of productive dialogue of interest here, analysis should consider such dynamic, and background common knowledge; this relates to the ways in which feature selection provides means to represent interaction, and exchange, and the ways in which these are encoded.

A number of the concerns here are issues of *segmentation*, the text from which features are labelled or extracted; for example, with the uni-gram or bi-gram more suitable for topic analysis than the detection of stylistic differences. Thus, features are used to map text to particular classifications or attributes (Rosé & Tovares, 2015). However, while such classifications might be attributed based on a range of features within a particular span of text, it is important to consider how the segments across which features are detected impacts on the understanding of such classification.

Indeed, the strength of the typology of talk described by Mercer and colleagues — of which exploratory dialogue is the most educationally productive — is not in its basis as a coding scheme. Rather, it is as a “useful frame of reference for making sense of the variety of talk... it helps an analyst perceive the extent to which participants in a joint activity are at any stage behaving collaboratively or competitively and whether they are engaging in critical reflection or in the mutual acceptance of ideas” (Mercer & Littleton, 2007, p. 55).

When understood like this, particularly with reference to common knowledge and the cohesive ties discussed above, the notion of categorizing individual level utterances makes little sense, particularly where local features of temporally and topically related utterances are not features in this classification. Moreover, when researchers engage in such analysis manually, their purpose is not to pick out individual exploratory contributions, but rather to look for the co-construction of meaning, over time, through the dialogue. Exploratory dialogue is thus fundamentally about interthinking, and co-reasoning. While certainly computational approaches should not necessarily mirror the same rules or procedures as human analysts, these are not simply procedural issues — they speak to the sort of analysis that can be effectively conducted (Rosé et al., 2008).

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143.
<http://dx.doi.org/10.18608/jla.2015.23.7>

Failures in performance of approaches dealing with only decontextualized text-segments may be due to a problematic premise underlying such an approach to text classification — one that assumes simple representation of utterance level data can represent the richness of interaction to an adequate extent. Again it is useful here to note that *performance* requires an operationalization — we assess the performance of classifiers in terms of how well they can identify features of interest in the text. Of course, such operationalization can be contested on the grounds of limitations of the feature space (for example, limiting features to keywords and not encoding interactive elements as features). That is not to write off the utility (and indeed, contribution) of simple approaches — which might, for example, usefully provide an overview of surface features of interest, such as indicating who is speaking when in order to understand linguistic dominance in a given dialogue.

This problem is not unique to computer-automation contexts. The growth of quantitative analysis as an approach to address issues with qualitative analysis has in cases suffered from similar concerns (Strijbos et al., 2006). Indeed, as Strijbos et al. point out, “A review of CSCL conference proceedings revealed a general vagueness in definitions of units of analysis. In general, arguments for choosing a unit were lacking and decisions made while developing the content analysis procedures were not made explicit” (p. 31). This raises concerns that much CSCL work fails to define units of analysis, nor to report the reliability for that level of segmentation. Here too they note that “the applicability of a unit that is smaller than a message is affected by: (a) the object of the study, (b) the nature of communication, (c) the collaboration setting, and (d) the technological communication tool used” (p. 35), a concern of some import given that the boundaries to which we apply our categories or codes rely on our segmentation method.

3.1 Operationalizing our Feature and Segmentation Level Representations

This issue of operationalization is a familiar topic, for example Rosé and Tovaes⁴ (2015) highlight the impact of different levels of granularity in segmentation for the training and classification of various forms of natural language data. Some dialogues include particular encoded markers — inputs in CSCL environments, repetition of key terms indicating topic shifts, moderator interventions, etc. — demarking broad segments of dialogue. However, natural language segmentation decisions are complex, with the smallest meaningful unit the uni-gram (typically a word), while other techniques may use whole paragraphs (e.g., a number of dialogue turns), or even whole documents, thus, for example simply using punctuation for segmentation is often inadequate (Rosé et al., 2008). This highlights a further aspect of the machine readable representation of data — segmentation even at the “feature level” is crucial, that is, which components of a dialogue turn are encoded for machine reading — cue-terms at the uni-gram (word), bi-gram (two words), tri-gram (three word), etc. level; other features such as syntactical elements indicative of particular orientations, questioning, etc.; turn-taking markers (e.g., marking when the speaker transitions). This is crucial because, “the accuracy of segmentation might substantially alter the results of the categorical analysis because it can have a dramatic effect on the feature based representation that the text classification algorithms base their decisions on...” (Rosé et al., 2008, p. 264).

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143.
<http://dx.doi.org/10.18608/jla.2015.23.7>

Thus, the interaction between segmentation and feature-level representation is important for understanding how features co-occur (within and between segments), and how multiple “conflicting” features are to be understood in segments — a more complex problem in longer segments. Segmentation is also important when working to understand the discursive properties of segments in which features represent properties of dialogue such as turn taking. There are at least three means through which context and features might be seen to interact in segmentation:

1. The first method is to add contextual markers to contributions that indicate some feature of their origin, but do not connect them to other documents in any deeper way. For example, adding a feature to indicate whether the speaker of an utterance is the same as the previous speaker or not, while coding at the whole transcript level. Methods to apply dialogue acts (Austin, 1975) as labels — labels indicating the type of “move” being made — (see for example, Erkens & Janssen, 2008; Král & Cerisara, 2012; Stolcke et al., 2000) — would also fall into this category. This is arguably not dissimilar to taking individual quotations from a transcript and providing some contextual information (although it is unusual for this to be formalized in the way feature selection is).
2. A second method would be to segment a transcript — using some preselected procedure; for example topic analysis to group utterances — and then code the whole segment as a collection of utterances, in a coarser way. We have come to think of this approach as the “bigger bag” method, indicating that the approach looks across a larger scope (the segment) but does not represent internal structure within that segment. While the typology of talk described by Mercer and colleagues is not a coding scheme, the sort of analysis some researchers working in this tradition conduct could be argued to be similar to this approach — drawing out connected utterances by topic and surface features indicating exchange in order to use them as examples of “exploratory,” “disputational,” or “cumulative” dialogue at a coarser level of analysis.
3. A final approach is to segment (again, by topic for example, adding a feature for this aspect) and then add features to the segment at an utterance level as well — building in both local and contextual features to the coding of individual contributions. That is, add global features, and then within each segment look for locally salient features; if the above approach is a “bigger bag” method, this approach further adds an interest in the locations of terms within the bag — specifically, their temporal sequence. One might imagine an example in which utterance level classes are applied within a segment based on features of the utterances, with the sort (and order) of utterances within that segment dictating the coarse-grain class for that segment.

The situation is further complicated in so far as some features will be context dependent in the sense of *creating* the context for their salience; for example, topical features define the context through which particular key terms (topical terms) become “features” in the machine learning sense. Again, decisions regarding the means through which to segment have impacts on the ways in which topics and other such features may be identified. In addition, given the way classifiers are trained, and the subsequent analysis of their success, we should be constantly mindful of our segmentation

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143.
<http://dx.doi.org/10.18608/jla.2015.23.7>

process in assessing the success of classifiers. How well a classifier has worked will depend on what it is to be used *for* — success does not exist in a vacuum.

It is to these particular issues, the ways in which feature selection, and segmentation play out in our operationalization, and the choices we make to which we turn now. In section 3.1.1 Temporality: A key facet of representing data for productive *dialogue*) we then note another space in which this feature-segmentation issue is foregrounded — the temporal nature, and analysis of dialogue data.

Of course, we may often be interested in more than simply what is said, but to whom it is said. Specifically, we have an interest in whether or not dialogue is dominated by (or indeed, solely from) single voices, or whether there is interaction between participants. This second sense may be taken at a fairly basic level — simply, whether or not features can be detected indicative of question and answer pairings, for example — or be related to the concern of “context” above, and the notion of common knowledge introduced earlier, that contributions may invoke the voice of others indirectly or through reference to their messages, and this may occur at any point in a dialogue.

The difficulty for computational approaches lies with “tying” these contributions together given their temporal separation. This sort of occurrence is common in learning contexts in which teachers (or moderators) may refer to earlier points that they believe it may be useful to discuss further. Yet, associating such contributions automatically is a computational challenge, raising the concerns described above with respect to the importance of “context.” It is precisely to this challenge that the work referred to earlier (see, for examples, Chiru et al., 2013; Rebedea et al., 2014; Trausan-Matu et al., 2012) orients, tying contributions over temporal stretches together through term-based analysis of “voice.”

A key consideration, again, is the interaction between segmentation and feature selection. To give a key example, we invite the reader to consider again the excerpt in Table 3, and consider any single utterance. It is apparent that, at that level, the presence of single or multiple instances of any feature (a key term for example), is unlikely to indicate a particular class. One means to address this issue is to look for co-occurrence of terms across a particular span (from the sub-utterance up to a whole transcript). For example, building on Gee’s (1986, 2003, 2010) sociolinguistic work, Shaffer and colleagues (Rupp, Gushta, Mislevy, & Shaffer, 2010; Shaffer, 2006) create segments, or, in Gee’s terms, “stanzas,” within which patterns of co-occurrence of terms can be identified. In those cases, though, where analysis is automated (in “Epistemic Games”), the dialogue is structured to facilitate identification of topically related talk in order to simplify the segmentation process. Other means through which segments may be identified include analysis of dialogue acts, (see, Erkens & Janssen, 2008; Král & Cerisara, 2012; Stolcke et al., 2000) through grammatical features indicating an exchange and shift to a new exchange, and more broadly than that analysis of “break points” in an exchange or topic indicative of new sections or types of dialogue (see for example, Chiu, 2008). In each case, the use of tokenizing for grammatical features (for example, through the use of the Part of Speech Taggers), and lexical tools for identification of topics, are important — both of these are feature selection tools, and again we see the interplay of features and segmentation — feature

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

selection can be used to provide the means through which to segment, in order to identify regularities in features across segments.

However, an issue remains. Specifically, as we have noted, the ways in which a single term develops meaning over time, raises a challenge for the kinds of segmentation alluded to above. To understand such topical shifts, and the ways in which discourse shifts, and evolves, the temporal component of the discourse data must be considered.

3.1.1 Temporality: A key facet of representing data for productive dialogue

A core component of the representation and segmentation of data — particularly regarding the co-construction of common knowledge — is the temporal nature of that data. Yet, this element of dialogue, and learning more generally, has typically been underexplored in both applied and research contexts (Littleton, 1999; Mercer, 2008; Mercer & Littleton, 2007), despite the fact it is a key element of context, and productive educational dialogue (as we discuss in section

2.1 The Challenge of Context above). This is a complex issue; temporality involves consideration of duration, sequence, pace, and salience of target events (Wise, Zhao, Hausknecht, & Chiu, 2013). Yet, despite the relative ease of access CSCL researchers have to process data (through log files for example), relatively little research has made use of this temporal information (Reimann, 2009), with most research opting for a “coding and count” strategy (Suthers, 2006). As Kapur (2011) notes, this kind of aggregation of events glosses temporal variability such that two dialogues might appear to be identical in the *count* of a certain class of dialogue, while in fact,

For one group, it could be that explanations were followed by more explanations, and likewise for critique. For the other group, it could be the case that explanations followed critiques that in turn led to more explanations and critique. In other words, for the first group, the learning mechanisms invoked by explanations and critique could be independent of each other whereas for the second group, they could be co-evolving and dependent. (Kapur, 2011, pp. 41–42)

Issues are more complex yet. A recent introduction to a special issue on self-regulated learning noted two types of considerations in temporality analysis: Those that explore the continuous flow of events, their positioning, rates, and duration; and those that analyze the arrangements of events within sequences, exploring the organization of multiple events over time (Molenaar & Järvelä, 2014). We discuss some of these approaches further in the section below, but first we make some general remarks considering the importance of this level of representation.

One of the articles in that collection (Winne, 2014) notes several questions regarding a view of self-regulated learning as unfolding over time, which we paraphrase here:

1. **What demarks events or time-spans?** Two common options are to use standardized approaches, such as every x seconds or utterances, and to use naturally occurring breaks (such as topic shifts, or markers of particular pre-defined classes of activity).

2. **How flexible should event spans be?** On the one hand, researchers might take very tightly defined event spans (every x seconds, every x number of events), while alternatively they might look for spans around the occurrence of particular target events.
3. **How are event patterns considered?** The patterning of events within and between samples of event-sequences via both data-mining and manual-analysis methods post-hoc or bottom up, or by conceptual analysis a-priori or top down.
4. **What are parameters of patterns?** This question relates to the delineating of a pattern, which might be segmented via two key means: 1) by how long it lasts, or how many “events” (components, features, etc.) occur within it; 2) by how “pure” it is, whether or not it is broken by other events (for example off-task talk in the middle of productive dialogue), or contains sub-patterns (for example, some small sections of disagreement or cumulative talk in an exploratory episode).
5. **What is the relationship between the pattern or feature set and other relevant data?** There are obvious implications here around the contexts for learning (for example, whether certain patterns are more likely to occur in particular contexts such as classroom setups). More broadly (and as Winne notes), we should also be interested in whether the duration or frequency of the pattern is related to learning outcome data.

Psychological constructs (such as self-regulated learning) can and should be thought of in light of such considerations (Winne, 2014). Indeed, recent work exploring the temporal patterns distinguishing “productive” knowledge building threads indicates the importance of such approaches (Chen & Resendes, 2014). In this work, a lag-sequential analysis (which we discuss further below) indicated that “productive inquiry threads involved significantly more transitions among questioning, theorizing, obtaining information, and working with information; in contrast, responding to questions and theories by merely giving opinions was not sufficient to achieve knowledge progress” (Chen & Resendes, 2014, p. 1). Clearly in relation to the building of common knowledge, such considerations also play a role. Earlier (

2.1 The Challenge of Context) we noted a distinction between background and dynamic common knowledge. While machine learning techniques certainly cannot hope to address many of the highly narrative senses in which temporality plays a role in dialogue, as the section above indicates, certainly there are ways in which temporality can be operationalized and represented to give insight into facets of its importance in dialogue.

3.1.2 Computational approaches to modelling temporality

Of specific interest in our case — and of analytic potential — is the analysis of sequences, and building of ideas. While duration (another core component of temporality) is no doubt important in learning contexts, it is not fundamental to our analysis of productive educational dialogue. Approaches through which sequence itself can be analyzed offer the opportunity to explore the evolving nature of the variables (or features) involved in a process, rather than treating features as fixed entities that vary only in their value (i.e., their occurrence count) (Reimann, 2009, p. 246). Given our argument that discourse both represents, and is constitutive of co-construction and interthinking, such an approach may offer important insight. In this section, we outline some possible practical means through which we might take steps into the automated analysis of

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

sequence data around exploratory dialogue, in addition to the exemplifications described in 2.2 Computational Approaches for Analysis of Productive Educational Dialogue.

We are aware of very recent attempts to model the relationship between dialogue sequences and learning outcomes. For example, Chen and Resendes' (2014) use of "lag sequential analysis" (LSA) (Faraone & Dorfman, 1987; Putnam, 1983) for the analysis of pre-classified knowledge building sequences; Kuvalja, Verma, and Whitebread's (2014) analysis of self-regulated learning sequences using T-patterns (Magnusson, 1996, 2000); Kinnebrew, Segedy, and Biswas' (2014) analysis of sub-patterns within sequences to explore groupings of these patterns (a possible way to detect behaviour types). We note also that Markov models⁵ are used in temporal contexts (see recently, Chiu & Fujita, 2014, although many such studies exist). However, as Reimann notes,

Like variable-based modeling, Markov models entail the assumption that history does not matter: The entire influence of the past occurs through its determination of the immediate present, which in turn serves (via the process) as the complete determinant of the immediate future (Abbott, 1990, p. 378). Histories are a kind of "surface reality" (Abbott) that are generated by deeper, underlying probabilistic processes that find expression in the value of variables or the conditional probabilities of event transitions. (Reimann, 2009, p. 246)

Certainly, such approaches have potential for important analytic insights in the learning sciences. However, while they allow for interesting analysis of short recurring sequences, they are not suitable for the kinds of temporal analysis we are interested in. Although T-pattern analysis can be used to explore longer, more temporally separated sequences, LSA, Markov models, and (insofar as it is temporal) hierarchical analysis are best suited to short recurring sequences. In all cases, though, the analysis is concerned with patterns that repeat with little variance. These analyses might be useful for detection of classes to be treated as features, for example with regard to segmentation of particular features of a dialogue HMM may provide useful tools for the identification of topics in discourse data, and the segmentation of that data by the detected topics (even in unstructured discourse) (Purver, Griffiths, Körding, & Tenenbaum, 2006). Indeed, it is important to note that in each of these cases the segmentation-feature relationship remains key. For example, in Chen and Resendes (2014) work, three features were used (for which they have good reliability):

1. The contribution type (questioning, theorizing, obtaining information, working with information, synthesis and analogies, supporting discussion) — which act as the feature types
2. Topic groupings — which offer a segmentation method
3. Whether the thread is labelled productive or improvable — which offer an outcome

However, none of these approaches facilitates the analysis of evolving dialogue, the kinds of iteratively changing use of terms crucial to the kinds of co-constructive language use in which we are interested.

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

Three recent developments deserve attention here. The first, Dyke, Kumar, Ai, and Rosé (2012), notes that, while earlier work (Chiu & Khoo, 2005; Jeong, 2005; Reimann, Frerejean, & Thompson, 2009) offers important insights in to sequences of events, because they are interested in relatively fine-grain regularities across time, they are poor at accounting for change over time, mid-range granularities and dependencies, and the multi-time-scale nature of many social processes involved in learning. They thus propose “interactive time-series visualizations, calculated from sliding window analyses [which] can make a methodological contribution to understanding the irregularities of the unfolding of collaborative processes over time” (Dyke et al., 2012, p. 2).

The model they propose takes a given feature (to use their example, the distribution of turns among participants) and rather than creating a summary of that feature over the whole dialogue, produces a value for time-slices over the period, indicating how it changes over time. Thus, rather than seeing that all participants had an equal number of turns, we might see that throughout a given dialogue, the turn-taking pattern changed. At a small window size (1 turn), this would be unproductive (because a single participant has 100% of the turns during a single turn); however, over larger segments insight may be given, with potential to produce “smoothed” visualization to support making sense of the data.

While certainly this gives more insight into longer term progression of a dialogue, facilitates a transition from micro to macro levels of analysis, and can theoretically account for the interaction of features or indicators (e.g., by plotting both against each other), the cumulative nature of dialogue is still excluded from the analytic frame. That is, while sequence may be productively explored, the accretion of ideas and the ways in which they change over time cannot. In contrast, in Introne and Drescher’s (2013) work, they note that

Unlike a document, which offers a snapshot of a relatively stable distribution of topics, a conversation is an unfolding process in which the definition of a topic is continually renegotiated. The lack of a relatively well-structured, intentionally designed document limits the utility of the machine-learning approaches that drive modern topic-tracking algorithms. (Introne & Drescher, 2013, p. 341)

In that paper they thus take as their “unit of analysis a sequence of replies, seek[ing] to understand how clusters of words in these reply sequences change, merge, and split” (p. 341) with a particular interest in modelling the statistical properties of the co-occurrence of words over time, as opposed to modelling probabilities based on dictionary entries or other corpora.

A final approach, which Mayfield and Rosé (2011) describe, involves developing a classifier based on the segmentation of sets of utterances classified according to exchange rules established in previous empirical literature. The general case of such an approach might involve a segment of a transcript (such as in Table 3) being identified — by both structural and topical features. This segment could then either be broken down into further segments depending on rules, such as the quick succession of two questions, or classified as a binary class if a particular set of features occurred in a predefined order; that is, the temporal nature of the dialogue — the potential sequential structures — can be formalized based on prior research. Such an approach is interesting precisely because it allows

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

researchers to formalize the results of a body of work into the classifier. Naïve classifiers are particularly useful for problems where solutions and constraints are not well known; the educational literature provides a rich body of empirical research and theoretically grounded work regarding the types of dialogue we wish to encourage in learning contexts.⁶

Of course, this likely only allows the selection of features around one element of “context” — the other lasting over longer exchanges such as whole lessons (a larger size grain of *dynamic* common knowledge), or indeed built up over years of interaction (historical common knowledge). To some extent, reference to common knowledge could be encoded, for example, through topic modelling that can label reference to such knowledge, or task-related context such as terms taken from given task descriptions and so on. Importantly, because the kind of dialogue we are interested in involves a dynamic feature set, built up through the dialogue itself, pre-defined feature sets (such as cue-phrases indicative of exploratory dialogue, like “because,” “so,” “I think,” etc.) are insufficient to identify many types of co-constructive dialogue, even if they are applied over appropriate segment sizes. However, the approach described by Mayfield and Rosé (2011) is entirely content agnostic, and generally, this is an advantage for a classifier in dealing with dialogue where the *content* of the utterances is of less interest than their *form* and interrelationships. Moreover, the potential to encode “rules” arising from prior empirical understanding of exploratory and accountable dialogue into a classifier is of great interest offering potential to build on existing learning sciences research.

4 CONCLUSION

We have foregrounded the ways in which three considerations of dialogue as data — its features, segmentation, and temporal nature — come together, and are crucial to understanding productive educational dialogue. In doing so, we have engaged with well-theorized, social psychological perspectives on dialogue as a critical facet of learning. By relating this theory to concrete examples of both discourse excerpts (Table 1 and Table 3), and analytic techniques used — both manual and computational — we have sought to address the concern raised in the introduction that computer-supported analysis of productive dialogue is thus far in its infancy, with computational models that “miss the deep, underlying structure in the data that would enable the models to generalize effectively” (Gweon et al., 2013, p. 246), and are highly context-specific (Mu et al., 2012). Fundamentally, we have argued that, whether implicitly or explicitly, analysis of productive educational dialogue must include consideration (or assumptions) around the temporal nature of that dialogue and its relationship to observable features and their scope over segments. As such, in both manual and computational analyses researchers should be explicit about their grounding assumptions in terms of observed “raw” data, application of operationalized coding schemes to features, and analysis or exposition of the processed data.

Having considered the issues raised by analytic techniques for productive educational dialogue, it is worth now reflecting on the sorts of data being considered here, and available to us more generally, and the implications of those data as *contexts* for analysis in their own right. Here our interest is largely productive educational dialogue such as “exploratory dialogue,” and a further point regarding data as context should be made: in many unstructured, unsupported environments without further

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

pedagogical support making clear expectations around the desirability of exploratory dialogue, we would *expect* it to have a fairly low incidence. Productive dialogue is something that often needs supporting, and requires specific teaching of skills.

Furthermore, there is a wider concern regarding the messiness of “chat” data, as opposed to data in more structured environments. Identifying “threads” in such environments is complex — multiple conversations may interweave, with frequent single-post contributions that seem not to fit into any wider exchange, and (as we saw above) sometimes long stretches between “connecting” contributions. The aptly titled “R-u-typing-2-me?” (Fuks, Pimentel, & de Lucena, 2006) discusses the development of a tool to address this “chat confusion,” highlighting difficulties relating to the high number of messages posted, identifying speaker targets, topic identification, and demarking threads of interactions (or exchanges). Here, though, the aim was not to facilitate dialogue in such unstructured environments, but to provide a structure to it. Thus, it may be that identifying sections of dialogue to which classifiers can usefully be applied (above the single contribution level) may remain a large challenge in the context of such chat data — our expectations here should be realistic.

As has been touched on throughout the preceding discussion, many of the concerns we might have regarding the success of a classifier or natural language processing approach (or indeed, any other method) depend fundamentally on the purposes for which the analytic is being deployed, including the data environment in which it is used. Success does not exist within a vacuum, it can only be considered as “success *at* something.” The measures of precision, recall, and the harmonic mean (F_1)⁷ give some indication of the various ways in which we measure success. If our interest is in supporting productive dialogue in action, a lower tolerance for false positives (falsely identifying exploratory dialogue) might be acceptable, where it would not be for a formal assessment model. Relatively simple models can be imagined in which “likely” exploratory sections of a dialogue are visualized in order to support end-users in finding the most productive parts of a dialogue, or student self-reflection (see for example Ferguson et al., 2013, p. 8), or through which simpler indicators of productive dialogue (questions, keywords, styles of speech, etc.) can be identified. However, we have here raised some concerns regarding the ability of analytics that do not represent data at the appropriate level to offer insight into the quality of dialogue (because they do not represent contextual features of dialogue indicative of co-construction), or the nature of that dialogue more broadly (because they do not represent important contextual features of dialogue indicative of topical shifts and progression). Thus, while relatively simple analytics may provide a useful step to support for self-reflection (where humans can “plug the gaps” to some extent), or recommendation (where the job is just to narrow the scope of search from the whole transcript to subsections), their use for deeper analysis of contentful productive dialogue is problematic. This reflects the claims made in sociocultural research, in which excerpts of dialogue are often contextualized by quantitative data, rather than to aid the understanding of that quantitative data.

Dialogue is a crucial part of learning, and new developments afford opportunity for new analysis of this dynamic human interaction, including the potential for novel new formative and summative assessments. However, dialogue is a multi-faceted, co-constructed, and dynamic tool. The choice of

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143.
<http://dx.doi.org/10.18608/jla.2015.23.7>

tools we deploy, and their limitations and — perhaps unstated — views on the nature of language used for learning matter deeply. In this paper, we have indicated some of the fine-grain ways in which analytics should be contextualized — in light of existing educational research offering prior knowledge for our analytic techniques. In particular, we note the challenges around translating well theorized and empirically supported commitments in one context, to the operationalization of analytic techniques and the commensurate differences in data-source. We highlight the importance of feature selection, segmentation, and temporality in understanding discourse data, and encourage researchers to be explicit regarding their theorizing and practical considerations around these fundamental components of productive educational dialogue. Any approach wishing to tackle the complex nature of productive educational dialogue will necessarily need to consider these issues.

ACKNOWLEDGEMENTS

We are grateful to Rebecca Ferguson, Simon Buckingham Shum, Yulan He, and Zhongyu Wei for “bouncing ideas” around in this middle space between educational and machine-learning research. We also gratefully acknowledge useful conversations about this work with Elijah Mayfield and Carolyn Rosé, who also gave an earlier draft a critical reading. Thanks to Elizabeth Dobson for her assistance with sourcing, and for giving her permission to use a dialogue excerpt drawn from her research on the processes of creative collaboration. Thanks too to Bart Rienties and Judith Kleine Staarman who gave earlier drafts of the paper critical readings.

NOTES

¹ “Dialogue” and “discourse” are used interchangeably in this paper and indicate any communicative spoken or written linguistic-exchange data.

² See the Thinking Together materials hosted at the University of Cambridge <http://thinkingtogether.educ.cam.ac.uk/>

³ We gratefully acknowledge Elizabeth Dobson for her assistance in sourcing this “transition” excerpt and allowing us to use it. For further information see Dobson’s PhD (2012) and Dobson, Flewitt, Littleton, and Miell (2011). We should note that, although Dobson terms it “exploratory dialogue,” cumulative dialogue was characterized slightly differently to emphasize its useful contribution in various contexts; for the sake of simplicity, we refer to cumulative dialogue here. Similarly, we have simplified the original excerpts annotation for our purposes here.

⁴ See also Arguello and Rosé (2006) for a consideration regarding topic segmentation, for which they argue you first need a definition of “topic,” which they suggest should be

1. Reproducible by human annotators
2. Not reliant on domain-specific or task dependent knowledge
3. Grounded in generally accepted principles of discourse structure (in order that shifts in topic are recognizable from surface characteristics of the dialogue)

They thus seek to identify topic boundaries in this instance using a hidden Markov model to mark topics as states and topic shifts as state transition probabilities (Arguello & Rosé, 2006).

⁵ Such models build on the underlying assumption that given a state (an event, or feature indicative of some particularly salient facet of dialogue) we can determine a probability distribution for the subsequent state. For example, if an utterance is labelled as a “question,” we can determine the probability distribution that the following state will be “answer,” and model these distributions for

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143.
<http://dx.doi.org/10.18608/jla.2015.23.7>

classification purposes. Note again the important interaction here between features and segments in determining states (where states might apply to a segment level context rather than utterance).

⁶ We note a similar point by Gweon, Jain, McDonough, Raj, and Rosé, who note “theories from social and cognitive psychology can usefully inform the manner in which data is transformed prior to machine learning or the way the structure of a model is specified in order to render the process analysis learnable by state-of-the-art machine learning algorithms” (2013, p. 246). In this paper, a Dynamic Bayesian network is used to detect features of “accommodation,” looking for the salient features between temporally adjacent utterances. While such an approach may also be productive for exploratory talk (in contrast to the ILP model, which imposes some prior restrictions), in the first instance ILP is likely to be a best first step in particular as it allows for coding over longer exchanges than temporally adjacent segments. Furthermore, the general points regarding the need for segmentation and feature demarcation remain fundamental to such an approach, as does the role of theory in translating data for machine learning techniques.

REFERENCES

- Abbott, A. (1990). A primer on sequence methods. *Organization Science*, 1(4), 375–392.
<http://dx.doi.org/10.1287/orsc.1.4.375>
- 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* (Vol. 1) (pp. 976–983). Chicago, IL: International Society of the Learning Sciences.
- Arguello, J., & Rosé, C. P. (2006). Topic segmentation of dialogue. *Proceedings of the Analyzing Conversations in Text and Speech Workshop at HLT-NAACL 2006* (pp. 42–49). USA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=1564542>
- Austin, J. L. (1975). *How to do things with words*. Oxford, UK: Oxford University Press. (Original work published in 1955)
- Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social Development*, 2(3), 202–221. <http://dx.doi.org/10.1111/j.1467-9507.1993.tb00014.x>
- Bakhtin, M. M. (1986). *Speech genres and other late essays* (Vol. 8). Austin, TX: University of Texas Press.
- Barnes, D., & Todd, F. (1977). *Communication and learning in small groups*. London: Routledge & K. Paul.
- Chen, B., & Resendes, M. (2014). Uncovering what matters: Analyzing transitional relations among contribution types in knowledge-building discourse. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 226–230.
<http://dx.doi.org/10.1145/2567574.2567606>
- Chiru, C., Rebedea, T., & Trausan-Matu, S. (2013). NLP-based heuristics for assessing participants in CSCL chats. In D. Hernández-Leo, T. Ley, R. Klamka, & A. Harrer (Eds.), *Scaling up learning for sustained impact* (pp. 84–96). Berlin/Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-642-40814-4_8
- Chiu, M. M. (2008). Flowing toward correct contributions during group problem solving: A statistical discourse analysis. *Journal of the Learning Sciences*, 17(3), 415–463.
<http://doi.org/10.1080/10508400802224830>

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

- Chiu, M. M., & Fujita, N. (2014). Statistical discourse analysis of online discussions: Informal cognition, social metacognition and knowledge creation. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 217–225. <http://dx.doi.org/10.1145/2567574.2567580>
- Chiu, M. M., & Khoo, L. (2005). A new method for analyzing sequential processes: Dynamic multilevel analysis. *Small Group Research*, 36(5), 600–631. <http://dx.doi.org/10.1177/1046496405279309>
- Clark, D., Sampson, V., Weinberger, A., & Erkens, G. (2007). Evaluating the quality of dialogical argumentation in CSCL: Moving beyond an analysis of formal structure. *Proceedings of the 8th International Conference on Computer-Supported Collaborative Learning* (pp. 13–22). New Brunswick, NJ, USA: International Society of the Learning Sciences.
- Clarke, S., Chen, G., & Resnick, L. (2014). Classroom discourse analyzer: Leveraging discourse analytics for teacher learning. Presented at the 3rd Workshop on Intelligent Support for Learning in Groups at the 12th International Conference on Intelligent Tutoring Systems, 5 June 2014, Honolulu, Hawaii, USA.
- Dobson, E. (2012). *An investigation of the processes of interdisciplinary creative collaboration: The case of music technology students working within the performing arts* (Doctoral dissertation, The Open University, the UK). Retrieved from <http://eprints.hud.ac.uk/14689>
- Dobson, E., Flewitt, R., Littleton, K., & Miell, D. (2011). Studio based composers in collaboration: A socioculturally framed study. In *Proceedings of the International Computer Music Conference* (pp. 373–376).
- Dyke, G., Kumar, R., Ai, H., & Rosé, C. P. (2012). Challenging assumptions: Using sliding window visualizations to reveal time-based irregularities in CSCL processes. *Proceedings of the 10th International Conference of the Learning Sciences* (Vol. 1, Full Papers, pp. 363–370). Sydney, Australia: International Society of the Learning Sciences.
- Edwards, D., & Mercer, N. (1987). *Common knowledge: The development of understanding in the classroom*. London: Routledge.
- Erkens, G., & Janssen, J. (2008). Automatic coding of dialogue acts in collaboration protocols. *International Journal of Computer-Supported Collaborative Learning*, 3(4), 447–470. <http://dx.doi.org/10.1007/s11412-008-9052-6>
- Faraone, S. V., & Dorfman, D. D. (1987). Lag sequential analysis: Robust statistical methods. *Psychological Bulletin*, 101(2), 312–323. <http://dx.doi.org/10.1037/0033-2909.101.2.312>
- Ferguson, R., & Buckingham Shum, S. (2011). Learning analytics to identify exploratory dialogue within synchronous text chat. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (LAK '11), 99–103. <http://dx.doi.org/10.1145/2090116.2090130>
- Ferguson, R., Wei, Z., He, Y., & Buckingham Shum, S. (2013). An evaluation of learning analytics to identify exploratory dialogue in online discussions. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 85–93. <http://dx.doi.org/10.1145/2460296.2460313>
- Ferguson, R., Whitelock, D., & Littleton, K. (2010). Improvable objects and attached dialogue: New literacy practices employed by learners to build knowledge together in asynchronous settings. *Digital Culture & Education*, 2(1), 103–123.

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

- Fuks, H., Pimentel, M., & de Lucena, C. J. P. (2006). R-U-Typing-2-Me? Evolving a chat tool to increase understanding in learning activities. *International Journal of Computer-Supported Collaborative Learning*, 1(1), 117–142. <http://dx.doi.org/10.1007/s11412-006-6845-3>
- Gee, J. P. (1986). Units in the production of narrative discourse. *Discourse Processes*, 9(4), 391–422. <http://dx.doi.org/10.1080/01638538609544650>
- Gee, J. P. (2003). *What video games have to teach us about learning and literacy* (2nd ed.). New York: Palgrave Macmillan.
- Gee, J. P. (2008). Learning and games. In K. Salen (Ed.), *The ecology of games: Connecting youth, games, and learning* (Vol. 3, pp. 21–40). Cambridge MA: MIT Press. <http://mitpress2.mit.edu/books/chapters/0262294249chap2.pdf>
- Gee, J. P. (2010). *An Introduction to discourse analysis: Theory and method* (3rd ed.). Milton Park, UK: Abingdon/New York: Routledge.
- Gress, C. L. Z., Fior, M., Hadwin, A. F., & Winne, P. H. (2010). Measurement and assessment in computer-supported collaborative learning. *Computers in Human Behavior*, 26(5), 806–814. <http://dx.doi.org/10.1016/j.chb.2007.05.012>
- 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. <http://dx.doi.org/10.1007/s11412-013-9172-5>
- Gweon, G., Jun, S., Lee, J., Finger, S., & Rosé, C. P. (2011). A framework for assessment of student project groups on-line and off-line. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing Interactions in CSCL* (pp. 293–317). Berlin: Springer. http://dx.doi.org/10.1007/978-1-4419-7710-6_14
- Halliday, M. A. K., Hasan, R., & Christie, F. (1989). *Language, context, and text: Aspects of language in a social-semiotic perspective*. New York: Oxford University Press.
- Herrmann, T., & Kienle, A. (2008). Context-oriented communication and the design of computer-supported discursive learning. *International Journal of Computer-Supported Collaborative Learning*, 3(3), 273–299. <http://dx.doi.org/10.1007/s11412-008-9045-5>
- Introne, J. E., & Drescher, M. (2013). Analyzing the flow of knowledge in computer mediated teams. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 341–356. <http://dx.doi.org/10.1145/2441776.2441816>
- Jeong, A. (2005). A guide to analyzing message–response sequences and group interaction patterns in computer-mediated communication. *Distance Education*, 26(3), 367–383. <http://dx.doi.org/10.1080/01587910500291470>
- Joshi, M., & Rosé, C. P. (2007). Using transactivity in conversation for summarization of educational dialogue. In *SLaTE Workshop on Speech and Language Technology in Education*, 1–3 October 2007 (pp.53-56). Farmington, PA, USA:ISCA Archive. Retrieved from http://www.isca-speech.org/archive_open/archive_papers/slate_2007/sle7_053.pdf
- Kapur, M. (2011). Temporality matters: Advancing a method for analyzing problem-solving processes in a computer-supported collaborative environment. *International Journal of Computer-Supported Collaborative Learning*, 6(1), 39–56. <http://dx.doi.org/10.1007/s11412-011-9109-9>
- Kärkkäinen, E. (2006). Stance taking in conversation: From subjectivity to intersubjectivity. *Text & Talk*, 26(6), 699–731. <http://dx.doi.org/10.1515/TEXT.2006.029>

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

- Kinnebrew, J. S., Segedy, J. R., & Biswas, G. (2014). Analyzing the temporal evolution of students' behaviors in open-ended learning environments. *Metacognition and Learning*, 9(2), 187–215. <http://dx.doi.org/10.1007/s11409-014-9112-4>
- Knight, S. (2013). Creating a supportive environment for classroom dialogue. In S. Hennessey, P. Warwick, L. Brown, D. Rawlins, & C. Neale (Eds.), *Developing interactive teaching and learning using the IWB*. Maidenhead, UK: Open University Press.
- Knight, S., & Littleton, K. (2015). Discourse-centric learning analytics: Mapping the terrain. *Journal of Learning Analytics*, 2(1), 185–209. Retrieved from <https://epress-dev.lib.uts.edu.au/index.php/JLA/article/view/4043>
- Král, P., & Cerisara, C. (2012). Dialogue act recognition approaches. *Computing and Informatics*, 29(2), 227–250. <http://dx.doi.org/10.3115/1118121.1118134>
- Kuvalja, M., Verma, M., & Whitebread, D. (2014). Patterns of co-occurring non-verbal behaviour and self-directed speech: A comparison of three methodological approaches. *Metacognition and Learning*, 9(2), 87–111. <http://dx.doi.org/10.1007/s11409-013-9106-7>
- Littleton, K. (1999). Productivity through interaction. In K. Littleton & P. Light (Eds.), *Learning with computers: Analysing productive interaction* (pp. 179–194). London, UK: Routledge.
- Littleton, K., & Howe, C. (2010). *Educational dialogues: Understanding and promoting productive interaction*. Abingdon, UK: Routledge.
- Littleton, K., & Mercer, N. (2013). *Interthinking: Putting talk to work*. London: Routledge.
- Littleton, K., & Whitelock, D. (2005). The negotiation and co-construction of meaning and understanding within a postgraduate online learning community. *Learning, Media and Technology*, 30(2), 147–164. <http://dx.doi.org/10.1080/17439880500093612>
- Magnusson, M. S. (1996). Hidden real-time patterns in intra- and inter-individual behavior: Description and detection. *European Journal of Psychological Assessment*, 12(2), 112–123. <http://dx.doi.org/10.1027/1015-5759.12.2.112>
- Magnusson, M. S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments, & Computers*, 32(1), 93–110. <http://dx.doi.org/10.3758/BF03200792>
- Martin, J. R., & White, P. R. (2005). *The language of evaluation*. New York: Palgrave Macmillan.
- Mayfield, E., & Rosé, C. P. (2011). Recognizing authority in dialogue with an integer linear programming constrained model. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT '11)* (Vol. 1, pp. 1018–1026). Portland, OR: ACM.
- Mercer, N. (2000). *Words & minds: How we use language to think together*. London: Routledge.
- Mercer, N. (2004). Sociocultural discourse analysis: Analysing classroom talk as a social mode of thinking. *Journal of Applied Linguistics*, 1(2), 137–168. <http://dx.doi.org/10.1558/japl.v1i2.137>
- Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *Journal of the Learning Sciences*, 17(1), 33–59. <http://dx.doi.org/10.1080/10508400701793182>
- Mercer, N. (2010). The analysis of classroom talk: Methods and methodologies. *The British Journal of Educational Psychology*, 80(1), 1–14. <http://dx.doi.org/10.1348/000709909X479853>
- Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: Ways of helping children to use language to learn science. *British Educational Research Journal*, 30(3), 359–377. <http://dx.doi.org/10.1080/01411920410001689689>

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

- Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach* (new edition). London: Routledge.
- Mercer, N., Littleton, K., & Wegerif, R. (2004). Methods for studying the processes of interaction and collaborative activity in computer-based educational activities. *Technology, Pedagogy and Education*, 13(2), 193–209. <http://dx.doi.org/10.1080/14759390400200183>
- Mercer, N., & Sams, C. (2006). Teaching children how to use language to solve maths problems. *Language and Education*, 20(6), 507–528. <http://dx.doi.org/10.2167/le678.0>
- Mercer, N., Wegerif, R., & Dawes, L. (1999). Children's talk and the development of reasoning in the classroom. *British Educational Research Journal*, 25(1), 95–111. <http://dx.doi.org/10.1080/0141192990250107>
- Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in Philosophy and Education*, 27(4), 283–297. <http://dx.doi.org/10.1007/s11217-007-9071-1>
- Michaels, S., O'Connor, M. C., Hall, M. W., & Resnick, L. (2002). *Accountable talk: Classroom conversation that works*. Pittsburgh, PA: University of Pittsburgh.
- Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning*, 9(2), 75–85. <http://dx.doi.org/10.1007/s11409-014-9114-2>
- 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>
- Purver, M., Griffiths, T. L., Körding, K. P., & Tenenbaum, J. B. (2006). Unsupervised topic modelling for multi-party spoken discourse. *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, 17–24. <http://dx.doi.org/10.3115/1220175.1220178>
- Putnam, L. L. (1983). Small group work climates: A lag-sequential analysis of group interaction. *Small Group Research*, 14(4), 465–494. <http://dx.doi.org/10.1177/104649648301400405>
- Rebedea, T., Chiru, C.-G., & Gutu, G.-M. (2014). How useful are semantic links for the detection of implicit references in CSCL chats? *RoEduNet Conference 13th Edition: Networking in Education and Research Joint Event RENAM 8th Conference*, 1–6. <http://dx.doi.org/10.1109/RoEduNet-RENAM.2014.6955311>
- Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, 4(3), 239–257. <http://dx.doi.org/10.1007/s11412-009-9070-z>
- Reimann, P., Frerejean, J., & Thompson, K. (2009). Using process mining to identify models of group decision making in chat data. In A. Dimitracopoulou, C. O'Malley, D. Suthers, P. Reimann (Eds.), *Proceedings of the 9th International Conference on Computer-Supported Collaborative Learning* (Vol. 1, pp. 98–107). Rhodes, Greece: International Society of the Learning Sciences.
- Resnick, L. B. (2001). Making America smarter: The real goal of school reform. In A. L. Costa (Ed.), *Developing minds: A resource book for teaching thinking*, 3rd ed. (pp. 3–6). Alexandria, VA: Association for Supervision and Curriculum Development.

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

- Rojas-Drummond, S., Littleton, K., Hernández, F., & Zúñiga, M. (2010). Dialogical interactions among peers in collaborative writing contexts. In K. Littleton & C. Howe (Eds.), *Educational dialogues: Understanding and promoting productive interaction* (pp. 128–148). Abingdon, UK: Routledge.
- Rojas-Drummond, S., Mercer, N., & Dabrowski, E. (2001). Collaboration, scaffolding and the promotion of problem solving strategies in Mexican pre-schoolers. *European Journal of Psychology of Education*, 16(2), 179–196. <http://dx.doi.org/10.1007/BF03173024>
- Rosé, C. P., Dyke, G., Law, N., Lund, K., Suthers, D., & Teplovs, C. (2011). Leveraging researcher multivocality for insights on collaborative learning: White Paper. *Alpine Rendez-Vous 2011*. Retrieved from <http://hal.archives-ouvertes.fr/hal-00692027>
- Rosé, C. P., & Tovaes, A. (2015). What sociolinguistics and machine learning have to say to one another about interaction analysis. In L. Resnick, C. Asterhan, & S. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue*. Washington, DC: American Educational Research Association.
- 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>
- Rosen, D., Miagkikh, V., & Suthers, D. D. (2011). Social and semantic network analysis of chat logs. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge (LAK '11)*, 134–139. <http://dx.doi.org/10.1145/2090116.2090137>
- Rupp, A. A., Gushta, M., Mislevy, R. J., & Shaffer, D. W. (2010). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *The Journal of Technology, Learning and Assessment*, 8(4). Retrieved from <http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1623>
- Shaffer, D. W. (2006). Epistemic frames for epistemic games. *Computers & Education*, 46(3), 223–234. <http://dx.doi.org/10.1016/j.compedu.2005.11.003>
- Shaffer, D. W. (2008). *How computer games help children learn*. New York: Palgrave Macmillan.
- Sinclair, J. M. H., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford, UK: Oxford University Press.
- Sionti, M., Ai, H., Rosé, C. P., & Resnick, L. (2011). A framework for analyzing development of argumentation through classroom discussions. In N. Pinkwart & B. M. McLaren (Eds.), *Educational technologies for teaching argumentation skills* (pp. 28–55). UAE: Bentham Science Publishers. <http://dx.doi.org/10.2174/97816080501541120101>
- Stahl, G. (2013). Transactive discourse in CSCL. *International Journal of Computer-Supported Collaborative Learning*, 8(2), 145–147. <http://dx.doi.org/10.1007/s11412-013-9171-6>
- Stahl, G., & Rosé, C. P. (2012). Group cognition in online teams. In E. Salas, S. Fiore M., & M. Letsky P. (Eds.), *Theories of team cognition: Cross-disciplinary perspectives* (pp. 517–548). New York: Routledge/Taylor & Francis.
- Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., ... Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3), 339–373. <http://dx.doi.org/10.1162/089120100561737>

(2015). Dialogue as data in learning analytics for productive educational dialogue. *Journal of Learning Analytics*, 2(3), 111–143. <http://dx.doi.org/10.18608/jla.2015.23.7>

- Strijbos, J.-W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers & Education*, 46(1), 29–48. <http://dx.doi.org/10.1016/j.compedu.2005.04.002>
- Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning*, 1(3), 315–337. <http://dx.doi.org/10.1007/s11412-006-9660-y>
- Suthers, D. D., & Verbert, K. (2013). Learning analytics as a “middle space.” *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge (LAK '13)*, 1–4. <http://dx.doi.org/10.1145/2460296.2460298>
- Thompson, K., Kennedy-Clark, S., Kelly, N., & Wheeler, P. (2013). Using automated and fine-grained analysis of pronoun use as indicators of progress in an online collaborative project. In N. Rummel, M. Kapur, M. Nathan, & S. Puntambekar (Eds.), *To See the World and a Grain of Sand: Learning across Levels of Space, Time, and Scale — Proceedings of the International Conference on Computer-Supported Collaborative Learning* (Vol. 1, pp. 486–493). Madison, WI: International Society of the Learning Sciences.
- Trausan-Matu, S., Dascalu, M., & Dessus, P. (2012). Textual complexity and discourse structure in computer-supported collaborative learning. In *Intelligent Tutoring Systems* (Vol. 7315, pp. 352–357). Berlin: Springer. http://dx.doi.org/10.1007/978-3-642-30950-2_46
- Wegerif, R. (2011). Towards a dialogic theory of how children learn to think. *Thinking Skills and Creativity*, 6(3), 179–190. <http://dx.doi.org/10.1016/j.tsc.2011.08.002>
- Wegerif, R., Littleton, K., Dawes, L., Mercer, N., & Rowe, D. (2004). Widening access to educational opportunities through teaching children how to reason together. *Westminster Studies in Education*, 27(2), 143. <http://dx.doi.org/10.1080/0140672040270205>
- Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science*, 33(1), 1–30. <http://dx.doi.org/10.1007/s11251-004-2322-4>
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, 46(1), 71–95. <http://dx.doi.org/10.1016/j.compedu.2005.04.003>
- Wei, Z., He, Y., Buckingham Shum, S., Ferguson, R., Gao, W., & Wong, K.-F. (2013). A self-training framework for automatic identification of exploratory dialogue. Paper presented at the *14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2013)*, 24–30 March 2013, Samos, Greece. Retrieved from <http://www.qcri.com/app/media/2028>
- Winne, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition and Learning*, 9(2), 229–237. <http://dx.doi.org/10.1007/s11409-014-9113-3>
- Wise, A. F., Zhao, Y., Hausknecht, S., & Chiu, M. M. (2013). Temporal considerations in analyzing and designing online discussions in education: Examining duration, sequence, pace, and salience. In E. Barbera & P. Reimann (Eds.), *Assessment and evaluation of time factors in online teaching and learning* (pp.198-231). IGI Global.