Computer-Supported Collaborative Learning
DOI 10.1007/s11412-013-9172-5

Measuring prevalence of other-oriented transactive contributions using an automated measure of speech style accommodation

Gahgene Gweon · Mahaveer Jain · John McDonough · Bhiksha Raj · Carolyn P. Rosé Q2 /Q1

Received: 18 June 2012 / Accepted: 18 April 2013

Abstract This paper contributes to a theory-grounded methodological foundation for automatic collaborative learning process analysis. It does this by illustrating how insights from the social psychology and sociolinguistics of speech style provide a theoretical framework to inform the design of a computational model. The purpose of that model is to detect prevalence of an important group knowledge integration process in raw speech data. Specifically, this paper focuses on assessment of transactivity in dyadic discussions, where a transactive contribution is operationalized as one where reasoning is made explicit, and where that reasoning builds on a prior reasoning statement within the discussion. Transactive contributions can be either self-oriented, where the contribution builds on the speaker's own prior contribution, or other-oriented, where the contribution builds on a prior contribution of a conversational partner. Other-oriented transacts are particularly central to group knowledge integration processes. An unsupervised Dynamic Bayesian Network model motivated by concepts from Speech Accommodation Theory is presented and then evaluated on the task of estimating prevalence of other-oriented transacts in dyadic discussions. The evaluation demonstrates a significant positive correlation between an automatic measure of speech style accommodation and prevalence of other-oriented transacts (R=.36, p<.05).

Keywords Transactivity · Speech-based assessment · Machine learning · Speech style accommodation

Introduction Q4

Applications of machine learning to automatic collaborative learning process analysis are growing in popularity within the computer supported collaborative learning (CSCL) community. Automatic analysis of collaborative processes has value for real time assessment during collaborative learning, for dynamically triggering supportive interventions in the midst of collaborative learning sessions, and for facilitating efficient analysis of

G. Gweon (✉) · M. Jain · J. McDonough · B. Raj · C. P. Rosé
Q3 KAIST (Korea Advanced Instituted of Science and Technology), Daejeon, Republic of Korea
e-mail: ggweon@kaist.ac.kr

collaborative learning processes at a grand scale. Early work in automated collaborative learning process analysis focused on text based interactions and key-click data (Soller and Lesgold 2000; Erkens and Janssen 2008; Rosé et al. 2008; McLaren et al. 2007; Mu et al. 2012). This work has enabled a whole series of studies where interactive support for collaborative learning was triggered by real time analysis of collaborative processes and yielded significant positive impact on learning (Kumar et al. 2007; Chaudhuri et al. 2008; Ai et al. 2010; Kumar and Rosé 2011).

While existing approaches to automated collaborative learning process analysis have had impact in the context of online group learning, even face-to-face group learning could potentially benefit from such technology in the future. For example, analysis of data from an interview study and classroom study with project based course instructors provides evidence that supporting assessment of group processes would add value to such courses (Gweon et al. 2011a). That interview study demonstrated that project course instructors are concerned about the extent to which students engage in productive knowledge sharing and knowledge integration in their working groups, but they are unable to accurately evaluate the extent to which this is happening or not in those working groups because the students do most of their work outside of class. Recently, interest in group learning supported by robots has also begun to emerge (Kanda et al. 2012). These shifts towards face-to-face group interactions in the three dimensional world around us rather than online require a corresponding shift in analysis technology from text-based input to multi-modal input, including text, speech, and gesture.

Closer to the current reality, as communication technologies such as cell phones and voice over IP become more ubiquitous and allow for communication and collaboration over multiple modalities including video, audio, and text to be accessible any time and any place, the line between online group learning and face-to-face group learning begins to blur. Thus, as more and more collaboration takes place over video and audio channels, the need grows for the CSCL community to think about how to extend collaboration support technologies from the text realm into audio and eventually video. To begin meeting this challenge, early work towards analysis of collaborative processes from speech has begun to emerge as well (Gweon et al. 2011b), although the early results showed predictive value that was just above random. In this paper we take the next step.

Where the burgeoning area of automated collaborative learning process analysis is still in its infancy is in regard to its engagement with theoretical constructs from social and cognitive psychology. The problem with neglecting to engage is that the models that are built miss the deep, underlying structure in the data that would enable the models to generalize effectively. Where this paper makes its contribution beyond a proof of concept for speech analysis is in illustrating how insights from the social psychology and sociolinguistics of speech style are able to provide a theoretical framework to inform the design of computational models for automated assessment of collaborative learning processes applied to acoustic data. While it might be easy to think of psychology and machine learning as being in two distinct worlds, the truth is that 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. We use as an example automated assessment of one specific type of valuable student contribution to group knowledge construction (namely other-oriented transacts (Berkowitz and Gibbs 1983, 1987), described below). We illustrate Q5 how to motivate the design of a data representation and model structure that together yield a positive proof of concept that collaborative processes can be assessed automatically in acoustic data.

The necessity for this methodology can be argued from a very basic understanding of how machine learning is applied. Machine learning algorithms are designed with the goal of finding mappings between sets of input features and output categories. When it comes to applications of machine learning to speech or text, the algorithms are not applied to the language data in its raw form. Instead, it must first be represented in terms of a list of attribute-value pairs referred to collectively as a vector space representation of the language data. Thus, first the researcher must select a set of features for use in representing every segment of speech or text. And then for each segment, these features must be extracted so that each attribute is associated with a value that was extracted from the data. Supervised machine learning algorithms find stable patterns within these feature vector representations by examining collections of hand-coded "training examples" for each output category, then using statistical techniques to find characteristics that exemplify each category and distinguish it from the other categories. The goal of such an algorithm is to learn general rules from these examples, which can then be applied effectively to new data. In order for this to work well, the set of input features must be sufficiently expressive, and the training examples must be representative.

One limitation of the state-of-the-art in machine learning applied to analysis of conversational interactions is the tendency to learn overly specific models that don't work well in new contexts (Mu et al. 2012). The problem of learning generalizable models is of great interest in the machine learning community, although it continues to pose challenges that remain to be overcome (Arnold et al. 2008; Daumé 2007; Finkel and Manning 2009; Joshi et al. 2012). Mu et al. addressed the problem in the context of analysis of text based interactions in threaded discussion environments using a preprocessing step that replaces some context specific portions of text, such as names, with more general tags. This offers the model features that apply in more than one context, which then enables a higher level of generalization. In this paper, we take a different approach. Instead of explicitly including more abstract features, we include simple generic speech features but include enough of them to offer the model the opportunity to choose the most strategic subset in context. Because we designed the structure of the model using theories from the social psychology of speech style, the model is able to leverage those theoretical insights in interpreting patterns of features. The model then is able to identify which subset of features has significance in a context sensitive way based on how they behave over the course of a conversation. This is done using an unsupervised approach, which requires neither hand labeled data nor hand crafted features. Generalization comes from the ability to learn a context specific model without labeled training data.

In the remainder of the paper, we first situate our work in the midst of current directions in collaborative process analysis and speech processing and review the literature on speech style accommodation in order to motivate our hypothesis. Next, we present both our manual and automatic approach for measuring the prevalence of other-oriented transactive contributions in debate discussions. After presenting an evaluation of the predictive validity of our model, we conclude with a discussion of future directions.

Theoretical framework

The area of automatic collaborative process analysis has focused on discussion processes associated with knowledge integration. Frameworks for analysis of group knowledge building are plentiful and include examples such as Transactivity (Berkowitz and Gibbs 1983; Teasley 1997; Weinberger and Fischer 2006), Inter-subjective Meaning Making (Suthers

2006), and Productive Agency (Schwartz 1998). In this paper we are focusing specifically on transactivity. More specifically, our operationalization of transactivity is defined as the process of building on an idea expressed earlier in a conversation using a reasoning statement. Research has shown that such knowledge integration processes provide opportunities for cognitive conflict to be triggered within group interactions, which may eventually result in cognitive restructuring and learning (de Lisi and Golbeck 1999). While the value of this general class of processes in the learning sciences has largely been argued from a cognitive perspective, these processes undoubtedly have a social component, which we explain below and use to motivate our technical approach.

Transactivity

Despite differences in orientation between the cognitive and socio-cultural learning communities, the conversational behaviors that have been identified as valuable are very similar. Schwartz and colleagues (Schwartz 1998) and de Lisi and Golbeck (1999) make very similar arguments for the significance of these behaviors from the Vygotskian and Piagetian theoretical frameworks respectively. The idea of transactivity comes originally from a Piagetian framework. However, it is important to note that when Schwartz describes from a Vygotskian framework the kind of mental scaffolding that collaborating peers offer one another, he describes it in terms of one student using words that serve as a starting place for the other student's reasoning and construction of knowledge. This implies explicit articulations of reasoning, so that the reasoning can be known by the partner and then built upon by that partner. Thus, the process is explained similarly to what we describe for the production of transactive contributions. In both cases, mental models are articulated, shared, mutually examined, and possibly integrated.

Building on these common understandings, Weinberger and Fischer have developed and successfully evaluated scaffolding for collaborative learning that addresses observed weaknesses in conversational behavior related to their operationalization of transactivity, which they refer to as Social Modes of Co-Construction (Weinberger and Fischer 2006), and which they distinguish as a separate dimension from micro (Toulmin 1958) and macro level argumentation (Kuhn 1991). Nevertheless, while they consider their Social Modes of Co-construction framework as being primarily an operationalization of the idea of transactivity, they describe how they draw from a variety of related frameworks rather than narrowly situating themselves within a single theoretical tradition.

There are a variety of subtly different definitions of transactivity in the literature, however, they frequently share 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 utterance involves one person operating on or thinking with some previously articulated reasoning. If a reasoning statement does not operate on some previously articulated reasoning it is an externalization. The most popular formalization of the construct of transactivity (Berkowitz and Gibbs 1979) has 18 types of transactive moves, which characterize each student's conversational turn, as long as it is considered an explicit reasoning display that connects with some previously articulated reasoning display. Before considering which of these codes, if any, is appropriate for a contribution, one must first determine whether that contribution constitutes an explicit articulation of reasoning, or at least a reasoning attempt. Beyond this, transacts have been divided along multiple different dimensions. However, for our work, we focus mainly on one, specifically the dimension that represents whether the transact might be self-oriented (ego, operates on the speaker's own reasoning) or other-oriented (alter, operates on the reasoning of a partner, dyad shared opinion) (Teasley 1997; Berkowitz and Gibbs 1979).

The important message behind our work is that effective application of machine learning requires insight into what social processes are transpiring in the data. In the case of transactivity specifically, the Piagetian roots of the concept argue that the associated social intentions should be maintaining relative equality and exerting effort towards building common ground. Those attitudes are consistent with maintaining a balance of assimilation and accommodation (de Lisi and Golbeck 1999), which goes hand in hand with the occurrence of productive sociocognitive conflict. While typically operationalizations of transactivity are expressed in terms of content level distinctions, the above discussion argues for a social interpretation that predicts the occurrence of other-oriented transacts in the presence of underlying processes of showing respect both for one's own views as well as of those of the interlocutor. Consistent with this idea, Azmitia and Montgomery (1993) have demonstrated that friends exhibit higher levels of transactive conversational moves than pairs who are not friends. Furthermore, it makes sense to consider that to build on a partner's reasoning, one must be attending to the partner's reasoning in the first place, and deem it worth referring to in the articulation of one's own reasoning.

Thus, with respect to the goal of automatic analysis of transactivity from speech data, targeting other-oriented transacts specifically, we hypothesize that designing a model in a theoretically informed way will improve our predictive validity. Specifically, by combining a feature representation that offers flexibility in the way style is encoded in speech as well as a model structure that reflects what is known about processes used to build social balance into an interaction we will be able to build a model that will positively correlate with the prevalence of other-oriented transacts in that interaction.

Speech style accommodation

We motivate our representation of the speech observations and the structure of the model from the sociolinguistic literature on speech style specifically (Coupland 2007; Eckert and Rickford 2001; Jaffe 2009) and language style more generally (Fina et al. 2006). It has long been established that, while some speech style shifts are subconscious, some speakers may also choose to adapt their way of speaking to achieve social effects within an interaction (Sanders 1987). Specifically we leverage the sociolinguistic notion of Speech Style Accommodation (Giles and Coupland 1991), which is very similar to the notion of interactive alignment (Garrod and Pickering 2004), both of which occur when interlocutors are working to build rapport and where speakers are treating one another with respect. From more of a computational perspective, we refer to one very specific process, which has been previously been referred to as "entrainment," "priming," "accommodation," or "adaptation" in other computational work (e.g., Levitan et al. 2011). From both of these perspectives, we are leveraging constructs that describe how shifts in language behavior within interactions reflect relational dynamics between conversational participants that reflect a very similar underlying balance of power to what we have described above in connection with transactivity (Giles and Coupland 1991).

Stylistic shifts may occur at a variety of levels of speech or language representation. For example, much of the early work on speech style accommodation focused on regional dialect variation, and specifically on aspects of pronunciation, such as the occurrence of post-vocalic r in New York City, that reflected differences in age, regional identification, and socioeconomic status (Labov 2010). Distribution of backchannels and pauses have also been the target of prior computational work on accommodation (Levitan et al. 2011).

One of the main motives for accommodation is to manipulate perceived social distance. If the amount of shift is asymmetric between speakers, it is typical for the speaker perceived as lower power or lower status to shift towards the speaker perceived as higher power or higher status. In that way, the lower status speaker shifts to close the gap in vertical social distance. Differences in power may originate from multiple sources, including persistent social roles and transitory relational dynamics, such as that one speaker is trying to persuade another speaker of something, which places that other speaker temporarily in a higher power position in the interaction.

On a variety of levels, speech style accommodation has been found to affect the impression that speakers give within an interaction. This is the mechanism through which speech style affects social distance. For example, Welkowitz and Feldstein (1970) found that when speakers shift to become more similar to their partners, they are liked more by partners. Another study by Putman and Street (1984) demonstrated that interviewees who converge to the speaking and response rates of their interviewers are rated more favorably. Giles and colleagues (1987) found that more accommodating speakers were also rated as more intelligent and supportive by their partners. Conversely, social and cultural factors in a group context affect the extent to which interlocutors engage with one another in the first place, if at all. For example, Purcell (1984) found that Hawaiian children exhibit more convergence in interactions with peer groups that they like more. Bourhis and Giles (1977) found that Welsh speakers, while answering to an English surveyor, broadened their Welsh accent when their ethnic identity was challenged. Scotton (1985) also found that few people hesitated to repeat lexical patterns of their partners to maintain integrity. These effects may be moderated by other social factors. For example, Bilous and Krauss (1988) found that females accommodated to their male partners in conversation in terms of average number of words uttered per turn. Hecht et al. (1989) also reported that extroverts are more listener adaptive than introverts, and so extroverts converged more in their data.

Prior research has attempted to quantify accommodation computationally by measuring similarity of speech and lexical features either over full conversations or by comparing the similarity in the first half and the second half of the conversation. For example, Edlund and colleagues (2009) measured accommodation in pause and gap length, using measures such as synchrony and convergence. Levitan and colleagues (2011) found that accommodation is also found in backchannel rituals. They show that speakers in conversation tend to use similar kinds of speech cues, such as high pitch at the end of utterance, to invite a back channel from their partner. In order to measure accommodation on these cues, researchers usually compute the correlation between the numerical measures of cue usage by interlocutors.

When stylistic shifts focus on specific linguistic features, then measuring the extent of the stylistic accommodation is simple because a speaker's style may be represented within a one or two dimensional space, and its movement can then be measured precisely within this space using simple linear functions. However, the rich sociolinguistic literature on speech style accommodation highlights a much greater variety of speech style characteristics that could be associated with social status. Unfortunately, within any given context, the linguistic features that have these status associations, generally referred to as "indexical" features, are only a small subset of all the linguistic features that are being used by a speaker in some way. Furthermore, the choice of which features carry this indexicality is frequently specific to a context. So separating the socially-meaningful variation from variation in other linguistic features occurring for other reasons can be like searching for a needle in a haystack. To meet this challenge, accommodation is measured with Dynamic Bayesian Networks (DBNs) in our work (Jain et al. 2012; Jensen 1996; Pearl 1988). This allows us to include a wide range of speech features extracted using acoustic processing techniques to represent the speech observations so that the contextually salient features have a greater chance of being included within the state space learned by the DBN.

The unsupervised Dynamic Bayesian Network Model allows one to model speech style accommodation without narrowly specifying the targeted linguistic features (more details on this model can be found in the Methods section). Because accommodation reflects social processes that extend over time within an interaction, one may expect a certain consistency of motion within the stylistic shift. A model that captures this insight is able to identify meaningful structure within the speech. Specifically, one can leverage this consistency of style shift to identify socially-meaningful variation, without specifying ahead of time what particular stylistic elements are the focus.

Insights related to language accommodation have important implications for computational work related to collaborative learning process analysis. The prevalence of other-oriented transacts in an interaction is said to reflect a balance of perceived power within an interaction. It is consistent with prior work on style accommodation to expect to observe this accommodation when interlocutors are working to build common ground with one another. Therefore, we hypothesize that an automatically generated assessment of speech style accommodation would positively correlate with the prevalence of hand coded other-oriented transactive contributions. Prior work has also revealed a consistent pattern in text based interactions. For example, in many earlier efforts towards automated analysis of transactivity in text based interactions we have achieved higher performance when our feature based representation of the text used for machine learning included a feature that represents language similarity (Rosé et al. 2008; Ai et al. 2010). This confirms that consideration of basic language processes and how they relate to categories of behavior informs the design of effective representations for making a coding scheme learnable.

Method

Our hypothesis is that a measure of speech style accommodation should positively correlate with prevalence of other-oriented transacts in conversations. We have argued this in the theoretical discussion above. The significance of this finding from a methodological standpoint is that it highlights the importance of considering the theoretical foundation for a construct when setting up a machine learning model to use for automated assessment.

Experimental procedure

In order to test the hypothesis, we first need a corpus of conversations that have been hand coded for other-oriented transacts so that we will have a validated measure of prevalence of other-oriented transacts to use as a dependent measure. Our three step method for measuring this dependent variable is detailed in the "Corpus Preparation" section. In addition, we need an automated measure of speech style accommodation in order to provide the independent variable. This measurement is outlined in the "Measuring Speech Style Accommodation with a Dynamic Bayesian Network" section. In that section, we present an unsupervised model for measuring speech style accommodation in segmented speech. In the results section we will present a validation experiment that supports the interpretation of the result returned from the unsupervised model as a measure of speech style accommodation. We then Q8 conduct a correlational analysis to evaluate the extent to which a measure of speech style accommodation positively correlates with prevalence of other-oriented transacts. Note that if the hypothesis is confirmed, the result will be far from unfalsifiable. While it is true that an unsupervised model will always find some structure in the data, there is no reason to believe that structure should necessarily correlate with prevalence of other-oriented transacts specifically apart from the hypothesis being correct.

Corpus preparation

Step 1: **Data collection using speech recorders.** The corpus used in our investigation is taken from face-to-face debate discussions collected as part of research on arousal and learning (Nokes et al. 2010). The study was conducted in a laboratory setting where pairs of participants were engaged in a debate wherein they took opposing sides on a controversial topic. The specific task that the participants were asked to discuss was the cause of the decline of the Ottoman Empire, which has prompted some controversy among historians. One side of the debate emphasizes factors internal to the Empire, while the other side emphasizes external factors. Each of the participants was provided with a four page packet containing background materials that support the idea of an internal or external cause, and were then asked to argue for their side. Each debate lasted 8 min. The experiment had two conditions in terms of conversation patterns: blocked and freeform. In the freeform condition, the two speakers could talk freely for the duration of 8 min. In the blocked condition, each speaker was given a chance to speak for 2 min in each turn, resulting in two turns per speaker during the 8 min. In this experiment, we focus particularly on the data from the freeform condition.

Participants were male undergraduate students, between the ages of 18 and 25 who volunteered to participate in the experiment for pay. Apart from meeting the criteria of being male undergraduates within the stated age range, no filtering was done. Participants were randomly assigned to conditions and pairs. In prior studies, it has been shown that accommodation varies based on gender, age and familiarity between partners. Because this corpus controls for most of these factors, it is appropriate for this experiment. Furthermore, because the participants did not know each other before the debate, one can assume that if accommodation occurred, it was only during the conversation.

In order to collect clean speech with each student's voice on a separate channel, each student wore a directional microphone. It should be noted that although it was possible to clearly identify the main speaker from the audio file, crosstalk, which is the other participants' voice, could still be heard in the background. A total of 76 sessions (with 152 participants) were collected and used for further analysis, half of which were in the freeform condition.

Step 2: **Transcribing and segmenting the recorded data.** For each audio file, each of the eight-minute discussion sessions were transcribed and manually segmented for further analysis. The motivation for the segmentation was that most articulations of reasoning should fit within a single segment so that transactive segments should link back to one specific prior segment. In our formulation of the rules for segmentation, we make use of the linguistic distinction between independent clauses and dependent clauses. A clause typically consists of one main verb and its arguments (i.e., the subject, and any direct and/or indirect objects). Sentences typically include one main clause, termed the "matrix clause", where the main idea of the sentence is most succinctly expressed. But the sentence may consist of multiple clauses. Some of these additional clauses are dependent on other clauses. For example, dependent clauses may modify a noun phrase, such as in "the country where a person was born" where the clause "where a person was born" is dependent on the noun phrase "the country". Other clauses are independent of one another. For example, "The Ottoman Empire fell, and all its glory became something of the past." consists of two clauses, separated by a comma, which can stand independent of one another. Our observation of the data was that typically, articulations of reasoning were expressed in single independent clauses, sometimes with additional dependent clauses attached. Thus, it made sense to segment the corpus at independent clause boundaries.

Specifically, the data was segmented into independent clauses according to the following two rules:

- Analyze-from-beginning rule: sentences should be analyzed from the beginning of the sentence to the end; and a clause boundary should be placed as soon as enough text has been seen that the clause is complete (i.e., all of the arguments of the verb have been seen).
- Dependent-clause rule: a sentence fragment that cannot stand alone should be treated as a dependent clause either on the preceding segment or the following segment.

This segmentation resulted in 5,490 separate segments.

Step 3: **Manual coding of transactivity.** Our analysis of transactivity is based on a categorical coding scheme. The categories are designed to flag places where there is reflection wherein participants take the time to display their reasoning, and then self or others build on that reasoning. These moments are distinguished from other places where speakers are expressing new ideas, restating facts, or otherwise interacting at a more superficial level. We looked for evidence of transactivity across the units of speech that participants expressed during the conversation. In order to be coded as a transactive speech unit, a statement should first contain a display of reasoning. That display of reasoning should also be related to a previous statement. If that previous statement was contributed by the same participant, then it is coded as a "self-oriented transact", otherwise it is coded as an "other-oriented transact".

Determining whether a sentence contains a reasoning statement is quite subjective—especially in conversational data, which can be informal in its presentation and leave much implicit. Therefore, we divide the process of identifying transactive contributions into two steps where we begin by differentiating non-reasoning and reasoning statements. Next, we differentiate between reasoning statements that represent new directions, from those statements that build on prior contributions (i.e., externalizations versus transactive contributions respectively). Finally the statements that are labeled as transactive are further coded as self-oriented transacts or other-oriented transacts.

The first step of the coding process is to distinguish between non-reasoning statements and reasoning statements. We have adapted the notion of an epistemic unit from Weinberger and Fischer (2006) because the topic of our conversations is somewhat different in nature. As in Weinberger and Fischer's (2006) notion of "epistemic unit", we look for a connection between two or more concepts. We describe our operationalization in detail below.

We use as an example a segment of a conversation provided in Table 1. The fourth column indicates whether the given contribution contains an articulation of reasoning ("R") or no reasoning ("N"). The simple way of thinking about what constitutes a reasoning display is that it typically communicates an expression of some causal mechanism. Often that will come in the form of an explanation, such as X because Y. However, it can be more subtle than that, for example "Russian invasion in 1914 led to a decrease in their population." The basic premise was that a reasoning statement should reflect the process of drawing an inference or conclusion through the use of reason. Note that in the example with the Russian invasion, although there is no "because" clause, one could rephrase this in the following way, which does contain a "because" clause: "The population decreased because of the Russian invasion in 1914."

More generally, we defined a reasoning display as an expressed relationship between two or more concepts. A concept could be some generally known prior knowledge, or one of the facts provided to the participants. The presence of multiple concepts in a statement by itself does not determine whether a statement articulates reasoning so that it is made explicit. Rather, the relationship between multiple concepts is the determining factor. For example, a simple list of concepts (e.g., Russians invaded, population decreased) is information sharing, and not articulated reasoning. We identified two types of relationships that signal a reasoning articulation; (1) Compare & contrast, and (2) Cause & effect.

1. Compare and contrast, tradeoff: When the speaker compares two concepts, the speaker is making a judgment, which involves thinking about how two concepts are related to one another.

    - The speaker compares two time periods ("at the time" & "today"): "At the time if you look at the technology, it wasn't that advanced as we have today."
    - When a speaker makes an analogy, he is making a link due to the similarity between two concepts. *"Outside powers were like the match lighting the fire."*

2. Cause and effect: When the speaker uses a cause-and-effect relationship, this process involves establishing the relationship between two concepts through a reasoning process. The general relation in this category is "doing x helps you achieve y". Examples are illustrated below.

    - A because of B: "They forced the Empire to be economically dependent because they set up trading posts and banks"
    - A in order to achieve B: "Great Britain came in and introduced capitulations to control schools and health systems."

Occasionally a reasoning statement was expressed over a sequence consisting of more than one segment. In that case, only the final segment was coded as reasoning and all of the other segments in this sequence were coded as no reasoning.

Statements that display reasoning can be either (1) externalizations, which represent a new direction in the conversation, not building on prior contributions, or (2) transactive contributions, which operate on or build on prior contributions. In our distinction between externalizations and transactive contributions, we have attempted to take an intuitive approach by determining whether a contribution refers linguistically in some way to a prior statement, such as through the use of a pronoun or deictic expression. Note that this does not mean that any deictic expression that refers to an entity mentioned in an earlier contribution is an indicator of a transactive contribution. Rather, what we mean is that the deictic expression should refer back to the idea of the earlier statement, i.e., "That means that a war would be more likely as a result." Furthermore, sharing a common subject between sentences can be a linguistic indicator that the focus of the two sentences remains consistent. For example, "Economic dependence of one country on another means the dependent country is weaker." And "Economic dependence can limit the agility of a country to respond to difficulties that arise." In this case, the shared subject is a linguistic indicator of the building relationship between these two statements.

The final step in the coding process is to distinguish between self-oriented and other-oriented transacts. This is usually a trivial matter of determining whether the prior statement on which a statement builds was contributed by the same speaker or a different speaker. In some cases, however, determining which prior statement a statement builds on is subjective. Ambiguous cases were very infrequent, however, as can be seen in the agreement measure reported below.

Table 1 shows a segment of conversation from the corpus used in this study. The fourth column indicates whether the given contribution contains reasoning ("R") or no reasoning ("N"). The last column of the table is marked as either an externalization (E), or as transactive, which can be self-oriented transacts (ST) or other-oriented transacts (OT) for the statements marked as (R). The first statement by speaker A is an externalization, since A starts a new topic; thus this contribution is not building on a prior contribution. Subsequent reasoning contributions in this discussion are coded as (ST) because they each build on statements that directly precede them, which in both cases were contributed by the same speaker. Table 2 shows an example where a speaker builds on an idea contributed by a different speaker.

This coding process was learned by two coders, initially trained using a manual that describes the above operationalization of reasoning displays and transactivity, along with an extensive set of examples. After each coding session, coders discussed disagreements and refined the manual as needed. Most of their disagreements were due to the interpretation of what the students meant rather than with the definition of reasoning itself. Therefore, later efforts focused more on defining how much the context of a statement could be brought to bear on its interpretation. In a final evaluation of reliability for reasoning coding, the kappa agreement was 0.72 between two coders over all of the data. After calculation of the kappa, disagreements were settled by discussion between the two coders. For distinguishing instances of transactivity and externalization, the coding yielded a kappa value of 0.7. For the distinction between self-oriented and other-oriented transacts, the kappa value was 0.95.

Based on the dichotomous coding of other-oriented transact or not, we computed a prevalence of other-oriented transacts per session by summing the number of other-

Q9 **Table 1** Sample contribution with self-oriented transacts in one turn from speaker A

Line	Speaker	Contribution	R/N	E/ST/OT
14	A	I think that the economic downfall of the Ottoman Empire was due to internal problems because of the first World War uh, and other civil wars going on uh, beforehand which took place over the hundreds and thousands of years that people have been in that area.	R	E
15		Um, this lead to, these wars lead to population problems.	R	ST
16		Uh, people were either being killed or they couldn't farm,	N	
17		and if you can't farm, you can't feed people	R	ST

Table 2 Sample contribution with both self-oriented and other-oriented transacts in an interaction between speakers A and B

Line	Speaker	Contribution	R/N	E/ST/OT
14	A	They couldn't, they didn't have any way to defend against the Europeans.	N	
15		It wasn't that the Europeans were so tough,	R	E
16		It was that they had already defeated themselves.	R	ST
17	B	Well I think another part is they uh, they just couldn't handle the Europeans.	R	OT

oriented transacts contained therein. This resulted in an average score of 36 per session. The minimum score for a session was 22, the maximum score was 60.

Measuring speech style accommodation with a Dynamic Bayesian Network

The goal of our modeling work is to develop an approach to measuring speech style accommodation that has the potential for easy adaptation to different contexts. For this purpose, an unsupervised approach is ideal since it does not require labeled training data. Dynamic Bayesian Network models provide the right mixture of formal properties for accomplishing this, as we detail in this section.

The theory of Bayesian networks is well documented and understood (Jensen 1996; Pearl 1988). A Bayesian network is a probabilistic model that represents statistical relationships between random variables via a directed acyclic graph (DAG). Thus, one can consider them a form of structural equation model (Loehlin 1998). Formally, it is a directed acyclic graph whose nodes represent random variables (which may be observable quantities, *latent* unobservable variables, or hypotheses to be estimated). *Dynamic* Bayesian networks (DBNs) represent time-series data through a recurrent formulation of a basic Bayesian network that represents the relationship between variables. Within a DBN, a set of random variables at each time instance t is represented as a static Bayesian Network with temporal dependencies to variables at other instants. Namely, the distribution of a variable x_t^i at time t is dependent on other variables at previous time points through conditional probabilities. For simplicity, in the discussion that follows we do not explicitly specify the random variables and the form of the associated probability distributions, but only present them graphically. We employ expectation maximization algorithm to learn the parameters of the models from training data, and the junction tree algorithm (Lauritzen and Spiegelhalter 1988) to perform inference.

The states and links that make up a DBN embody the assumptions behind the way the phenomenon of interest works. The idea is that when the probabilities are estimated from the data, they are most likely to be instantiated in such a way that any pattern found in the data by the network reflects those assumptions. Thus, if the assumptions are properly encoded in the structure of the network, then the pattern found by the network is likely to reflect the phenomenon of interest from which those assumptions were inspired. Our model embodies two premises. First, a person's speech in any turn is a function of his/her speaking style in that turn, which is influenced by their speech style in their previous turn. Second, a person's speaking style at any turn depends not only on their own personal tendencies, but also by their accommodation to their partner. We represent these dependencies as the DBN displayed in Fig. 1.

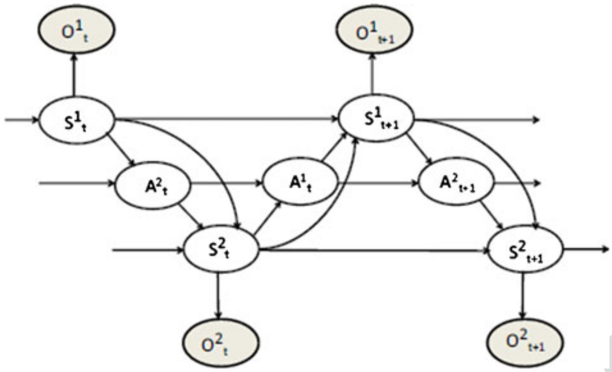

Fig. 1 DBN for modeling speech style accommodation (Jain et al. 2012)

Our model is constructed from two types of latent states in addition to observed vectors of speech feature:

1. *Speaking Style State*: These states represent the speaking styles of the partners in a conversation. We represent these states as s_t^i , where t represent turn index and i represents speaker index. These states are assumed to belong to a finite, discrete set.
2. *Accommodation State*: An accommodation state represents the indirect influence of partners on each other in a conversation. In our present design, it can take a value of either 1 or 0. These states are represented as A_t^i , where t is the turn index and i represents the speaker index.
3. *Observation Vector*: The observation vectors are the feature vectors o_t^i computed for each turn, where again where t is the turn index and i represents the speaker index.

The foundation of the model represents the production of speech (i.e. speech features) by a speaker in the absence of other influences. As in other state-of-the-art approaches to applying machine learning technology to speech data, the speech signal is first processed using basic audio processing techniques. The signal is processed in order to extract features from the segments of speech, which are then used for classification using a machine learning model. For example, one may use acoustic and prosodic features typically used for measuring emotion in speech (Ranganath et al. 2009; Ang et al. 2002; Kumar et al. 2006; Liscombe et al. 2005). This research makes use of signal processing techniques that are able to extract the basic acoustic and prosodic features used frequently in prior work; for example, variation and average levels of pitch, intensity of speech, or the amount of silence and duration of the speech. Acoustic and prosodic features are frequently associated with intuitive interpretations, and this makes them an attractive choice to play a role in baseline techniques for stylistic classification tasks. For example, increased variation in pitch might indicate that the speaker wants to deliver his ideas more clearly. Likewise, volume and duration of speech may signal that a speaker is explaining his ideas in detail, presenting his point of view about the subject matter.

The speech features o_t^i in any turn are caused by the speaking style s_t^i in that turn. The style s_t^i in any turn depends on the style in the previous turn, to capture the speaker-specific patterns of variation in speaking style. Specifically, we characterize conversations as a series of spoken turns by the partners. Thus, from a technical perspective we characterize the speech in each turn through a vector o_t^i that captures several aspects of the signal that are salient to style.

We add to that basic model the influence the conversational partner's speech style has on the speaker's style. These are conditional probability links that point from one speaker's style state to that of the other speaker. In addition to this we introduce binary valued accommodation states, A_t^i , into the model that indicate whether a speaker i is in a state of accommodating to his partner or not at time t. The accommodation state in one time point influences the accommodation state in the next time point. We see this both in (1) the links from speaking style states to accommodation states as well as (2) between accommodation state from one time point to the accommodation state in the next time point. We expect that the likelihood of a speaker accommodating in one time point is higher if the other speaker was in a state of accommodating on the last time point. The value of the accommodation state interacts with the influence of a partner's speech style on the speaker's speech style. In other words, the partner's style should have a greater influence on the speaker when they are accommodating than when they are not. We see this in the links from the accommodation states to speaking style states.

Using the model components introduced in this section, a space of possible models has been systematically explored in our prior work on speech style accommodation (Jain et al. 2012). And while we justify the model structure proposed in this paper from a theoretical perspective, we acknowledge that the link between theory and model structure could be further explored, and there may be alternative model structures that would perform better than the one we propose in this paper.

Results

In this section we present two types of results. First, we present a validation study in which we evaluate the extent to which the DBN model can be said to measure speech style accommodation. Next, we test the hypothesis that speech style accommodation positively correlates with prevalence of transactivity.

Model estimation and validation

The purpose of the DBN described in the previous section is to obtain a measure of speech style accommodation from the raw speech (i.e., audio signal) collected in a session to use for testing the hypothesis that speech style accommodation positively correlates with transactivity. In the last section, we described how the theory behind how accommodation works was used to inspire the structure of the model that we specified. In this section we describe how we used data to estimate the parameters for that model as well as to validate the model's measurement as a predictor of speech style accommodation. The validation experiment was conducted on the Ottoman Empire corpus mentioned earlier.

Preparing the speech data As mentioned above, the speech from each participant was recorded on a separate channel. As a first step, we segmented the speech data from each student into turn length segments. We did this by aligning the speech recordings automatically to their transcriptions at the word and turn level. After aligning the corpus at the word level, we identify each turn interval of each partner in the conversation. Using this method, we split the set of 76 segmented conversations into two sets of 38 conversations. We extracted features from each segment, and we trained the model on one set of 38 multi-segment conversations and tested on the other.

In this paragraph, we will explain the specifics of the features extracted from the speech from a technical perspective. Casual readers may skip this paragraph. The goal of the speech data representation was to enable modeling style in a general way, without making a strong assumption about what aspect of the speech signal would carry the socially significant style indicators. Thus, a rather broad range of feature types was included while keeping the total feature space size to a manageable level for the small amount of data that we had available for training. Within each turn the speech was segmented into analysis windows of 50 ms, where adjacent windows overlapped by 40 ms. From each analysis window a total of seven features were computed: voice probability, harmonic to noise ratio, voice quality, three measures of pitch $\left(F_0, F_{0^{raw}}, F_0^{env}\right)$, and loudness. A 10-bin histogram of feature values was computed for each of these features, which was then normalized to sum to 1.0. The normalized histogram effectively represents both the values and the fluctuation in the features. For instance, a histogram of loudness values captures the variation in the loudness of the speaker within a turn. The logarithms of the normalized 10-bin histograms for the seven features were concatenated to result in a single 70-dimensional observation vector for the turn. These 70 dimensional observation vectors for each turn of any speaker are represented in our model as o_t^i where t is turn index and i is speaker index. We used the OPENSmile toolkit (OpenSmile 2012) to compute the features. Q10

Real versus constructed pairs We set up the validation experiment in such a way as to isolate speech style convergence from lexical convergence when we evaluate the performance of our model. We accomplished this by measuring accommodation between (1) Real pairs: pairs of humans who had a real conversation and (2) Constructed pairs: constructed pairs in which one person from a real conversation is paired with a constructed partner, where the partner's side of the conversation was constructed from turns that occurred in other conversations. In particular, for each of the 38 Real pairs in the test corpus, we composed two Constructed pairs. Each Constructed pair comprised one student from the corresponding Real pair (i.e., the real student) and a Constructed partner that resembled the real partner in content but not necessarily style. We did this by iterating through the real partner's turns, replacing each with a turn that matched as well as possible in terms of lexical content but came from a different conversation. Lexical content match was measured in terms of word overlap. Turns were selected from the other Real pairs. Thus, the Constructed partner had similar content to the corresponding real partner on a turn by turn basis, but the style of expression could not be influenced by the Real student. Thus, any similarity that existed in style would be by chance or because of lexical similarity rather than from speech style accommodation.

Accommodation is a phenomenon that occurs within interactions between speakers; we can expect not to observe accommodation occurring between individuals that have never met and are not interacting. On average, then, we expect to see more evidence of speech style accommodation in pairs of individuals who really interacted than in pairs of individuals who did not interact and have never met. Thus, we may evaluate the extent to which our model is sensitive to social dynamics within pairs by the extent to which it is able to distinguish between true conversations between Real pairs of speakers and synthetic conversation between Constructed pairs. A similar experimental paradigm has been adopted in prior work on speech style accommodation (Levitan et al. 2011). The extent to which the model returns a higher score for the Real pair than the Constructed pair can be seen as a sign of success.

We computed an accommodation score for each of the Real pairs and Constructed pairs. In order to obtain a measure that can be used to compute the extent of accommodation for a session, we compute the most probable style state for each turn from the model using by means of the maximum likelihood estimate. The accommodation value as then the fraction of turns in a session where the most likely style state of the two partners on adjacent turns was computed as the same. We then compared the extent to which the model predicted higher accommodation for the Real pair versus the Constructed pairs using an ANOVA model with Conversation type (Real vs Constructed) nested within Conversations as the Independent variable and Accommodation score as the Dependent variable. In this way we make a controlled comparison between real and constructed pairs such that we hold constant random factors that vary between conversations. The difference was significant $F(1, 76)=1.88$, $p<.05$, with the average score for Constructed pairs being .52 with a standard deviation of .27, and for Real pairs .62 with a standard deviation of .31. The computed accommodation score for each session is what we use in the experiment to test the extent to which speech style accommodation positively correlates with prevalence of other-oriented transacts below.

Hypothesis test

Now we evaluate the correlation between the accommodation score and prevalence of other-oriented transacts using a linear regression. Rather than trying to locate the exact position of transactive statements, we measured the prevalence of other-oriented transacts. It makes sense to believe that extent of accommodation says something about the effort participants in a conversation are making towards building mutual understanding, which should be reflected in prevalence of other-oriented transacts (de Lisi and Golbeck 1999). For this analysis, accommodation scores were assigned to conversations through three-fold cross-validation where on each fold, 2/3 of the data was used as training data and 1/3 for testing, so that all of the freeform conversations could be used in the correlational analysis. Beyond hypothesizing that we should see a significant positive correlation between the accommodation score and prevalence of other-oriented transacts, we further hypothesized that there will not be a significant correlation between amount of accommodation and non-social categories of reasoning including reasoning statements that are not transactive or transacts that are self-oriented.

Indeed, as displayed in Fig. 2, the finding is exactly what we predict. Since prevalence of reasoning, prevalence of transactivity in general (including both self and other-oriented transacts) and other-oriented transacts are highly correlated, we do see positive correlations between the accommodation score and all three of these measures, however, the correlation is only significant in the case of other-oriented transacts (R=.36, p<.05). It is not significant in the case of reasoning statements (R=.18, p=n.s.) or transacts in general (R=.13, p=n. s.).

Note that we are not arguing that there is a causal relationship between speech style accommodation and other-oriented transacts. Rather, we are saying that speech style accommodation is useful for assessment of other-oriented transacts because both are caused by the same underlying social processes. In support of this, Table 3 illustrates an extended example from a conversation where we see a high degree of speech style accommodation using the Dynamic Bayesian Network model. We see in this interaction that the two speakers are each working hard to understand where the other is coming from. We see this particularly in markers such as "you mean" and "you're talking about". Thus, although the two speakers are intensely involved in the discussion, and they don't agree with one another, they are working to understand one another, and this

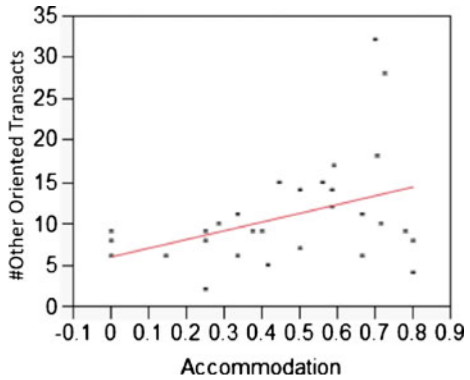

Fig. 2 Correlation between speech style accommodation and prevalence of other-oriented transacts

is reflected both in their high degree of speech style accommodation and in their high prevalence of other-oriented transacts.

Discussion

In this article, we presented our work toward an automatic detection of transactive contributions in speech data. As argued above, where this paper makes its contribution beyond a

t3.1 **Table 3** Example interaction between speakers A and B where speech style accommodation is high, and there is a high prevalence of other-oriented transacts

	Line	Speaker	Contribution	R/N	E/ST/OT
t3.2	Line	Speaker	Contribution	R/N	E/ST/OT
t3.3	52	A	It's based off of internal factors [like the economy] that are the main cause	R	E
t3.4	54	B	So, so you mean that the external factors…I mean this is, but this is what I don't understand.	N	
t3.5	55		You're talking about how the exter, the internal factors had external factors just play off,	N	
t3.6	56		but maybe are you sure its not also the ex, the internal factors helped the external factors?	R	OT
t3.7	57	A	They may help them	N	
t3.8	58		but there's no, (PB interrupts) the external factors don't come in if there's not the internal strife.	R	OT
t3.9	59	B	In…Internal factors already, the internal problems were already there	N	
t3.10	60		and the external factors, the external problems such as the country, all the European nations helped insight revolutions and rebellions,	R	OT
t3.11	61		and thus, thus they tear, they tore it apart.	R	ST
t3.12	62	A	It's their own religious differences between the communities that tears it apart because they all have a sense of nationalism and pride that they don't want to be under the Turkish sultan.	R	OT
t3.13	63		They want to be their own (B interrupts) place.	N	
t3.14	64	B	And thus they turned to the European nations because the European nations offered help to gain their own countries.	R	ST

proof of concept for speech analysis is in illustrating how insights from the social psychology and sociolinguistics of speech style are able to provide a theoretical framework to inform the design of computational models for automated assessment of collaborative learning processes. As an illustration, we have demonstrated the possibility of measuring prevalence of other-oriented transactive contributions in speech recordings from face-to-face discussions. This research shows promise that automatically detectable properties of speech, such as evidence of stylistic convergences between speakers, can be useful indicators of prevalence of other-oriented transacts (r=0.36).

More importantly we have illustrated a methodology for guarding against learning shallow models that miss the underlying structure in the data so that enable the models are able to generalize effectively. This work demonstrates that applying machine learning for an automated collaborative process analysis task can productively leverage insights from social psychology and sociolinguistics. Our future work will build on this initial demonstration and seek other ways that we can improve our ability to monitor social processes that operate through linguistic communication by using theoretically motivated applications of machine learning technology. For example, our reading of this literature points to the importance of considering how social interpretation of language requires comparisons between properties of an utterance and expectations that arise from individual and group norms. However, these norms are also a moving target. And thus as we focus on more challenging assessment tasks over large periods of time, we may need to leverage ideas related to social emergence in our computational models (Sawyer 2005).

Earlier work laying a foundation for detection of transactivity in speech (Gweon et al. 2011b) began by using a straightforward application of frameworks from prior language technologies research that focused on the related problem of emotion detection is speech (Kumar et al. 2006), or detection of social processes such as flirting (Ranganath et al. 2009). While the results of this earlier work showed a non-random correlation between simple speech features used in prior work and a distinction between transactive and non-transactive contributions, this paper presents more convincing results. Specifically, we leverage insights from the sociolinguistics of speech style, which is a literature that explores social interpretations of stylistic shifts within an interaction (Eckert and Rickford 2001; Giles 1984). We have discussed the theoretical connection between speech style accommodation and transactivity above, and that theoretical motivation lead to a positive result of our technical approach, as demonstrated in the results we presented above.

One limitation of the current work is that it was conducted using data from short argumentative interactions between pairs of male students who were close to one another in age. The very narrowly defined scope of contextual factors might very well have affected the amount of speech style accommodation we see, and might also affect the strength of connection between speech style accommodation and prevalence of other-oriented transacts. In our future work we will investigate the generality of the finding across a much wider variety of tasks and interaction contexts in terms of group composition with respect to age and gender. Furthermore, it would be interesting to investigate the extent to which the pattern we have identified might be specific to certain cultures.

Finally, although a major advantage of the unsupervised DBN modelling approach we have used is generality across contexts, we have only evaluated the predictive validity of its computed Accommodation score in this one context. Thus, and important part of our follow-up work will be testing the generality of this approach across contexts.

Acknowledgments This work was supported in part by NSF grant SBE 0836012 to the Pittsburgh Science of Learning Center. We gratefully acknowledge John Levine and Timothy Nokes from the University of Pittsburgh for sharing their data with us for these experiments.

References

Ai, H., Kumar, R., Nguyen, D., Nagasunder, A., Rosé, C.P. (2010). Exploring the effectiveness of social capabilities and goal alignment in computer supported collaborative learning. In *Proc. ITS 2010. LNCS* (vol. 6095, pp. 134–143). Heidelberg: Springer.

Ang, J., Dhillon, R., Krupski, A., Shriberg, E., Stolcke, A. (2002). Prosody-based automatic detection of annoyance and frustration in human-computer dialog. In *Proc. International Conference on Spoken Language Processing* (*ICSLP 2002)* (pp. 2037–2039). Denver, CO.

Arnold, A., Nallapati, R., Cohen, W. (2008). Exploiting feature hierarchy for transfer learning in named entity recognition. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (pp. 245–253).

Azmitia, M., & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social Development, 2*, 202–221.

Berkowitz, M., & Gibbs, J. (1979). Unpublished manual for coding transactivity.

Berkowitz, M., & Gibbs, J. (1983). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly, 29*, 399–410.

Bilous, F., & Krauss, R. (1988). Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language and Communication, 8*(3), 183–194.

Bourhis, R., & Giles, H. (1977). The language of intergroup distinctiveness. In H. Giles (Ed.), *Language, ethnicity and intergroup relations* (pp. 119–135). London: Academic.

Chaudhuri, S., Kumar, R., Joshi, M., Terrell, E., Higgs, F., Aleven, V., et al. (2008). It's not easy being green: Supporting collaborative "green design" learning. In *Proceedings of Intelligent Tutoring Systems, Lecture Notes in Computer Science Volume 5091* (pp. 807–809).

Coupland, N. (2007). *Style: Language variation and identity*. Cambridge: Cambridge University Press.

Daumé, H. III. (2007). Frustratingly easy domain adaptation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics* (pp. 256–263).

de Lisi, R., & Golbeck, S.L. (1999). Implications of the Piagetian Theory for peer learning. In *Cognitive perspectives on peer learning* (pp. 3–37).

Eckert, P., & Rickford, J. (2001). *Style and sociolinguistic variation*. Cambridge Univ Press.

Edlund, J., Heldner, M., Hirschberg, J. (2009). Pause and gap length in face-to-face interaction. In *Proceedings of Interspeech* (pp. 2779–2782).

Erkens, G., & Janssen, J. (2008). Automatic coding of dialogue acts in collaboration protocols. *International Journal of Computer Supported Collaborative Learning* (3), 447–470.

Fina, A., Schiffrin, D., Bamberg, M. (2006). *Discourse and identity*, Cambridge University Press.

Finkel, J., & Manning, C. (2009). Hierarchical Bayesian domain adaptation. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 602–610).

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8*, 8–11.

Giles, H. (1984). *The dynamics of speech accomodation*. Amsterdam: Mouton.

Giles, H., & Coupland, N. (1991). *Language: Contexts and consequences*. Thomson Brooks/Cole Publishing Co.

Giles, H., Mulac, A., Bradac, J., & Johnson, P. (1987). Speech accommodation theory: The next decade and beyond. *Communication Yearbook, 10*, 13–48.

Gweon, G., Jeon, S., Lee, J., Rosé, C.P. (2011a). Diagnosing problems in student project groups. In S. Puntambekar, G. Erkens, C. Hmelo-Silver (Eds.) *Analyzing collaborative interactions in CSCL: Methods, approaches and issues* (pp. 293–318). Springer.

Gweon, G., Agarwal, P., Udani, M., Raj, B., Rosé, C.P. (2011b). The automatic assessment of knowledge integration processes in project teams. In *Proceedings of Computer Supported Collaborative Learning*, Hong Kong, China (pp. 462–469).

Hecht, M., Boster, F., & LaMer, S. (1989). The effect of extroversion and differentiation on listener adapted communication. *Communication Reports, 2*(1), 1–8.

Jaffe, A. (2009). *Stance: Sociolinguistic perspectives*. Oxford University Press.

Jain, M., McDonogh, J., Gweon, G., Raj, B., Rosé, C.P. (2012). An unsupervised Dynamic Bayesian Network approach to measuring speech style accommodation. In *EACL 2012 Proceedings of the*

13th Conference of the European Association for Computational Linguistics, Avingon, France, April 23–27, 2012 (pp. 787–797).

Jensen, F.V. (1996). *An introduction to Bayesian networks*. UCL Press.

Joshi, M., Dredze, M., Cohen, W., Rosé, C.P. (2012). Multi-domain learning: When do domains matter. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CoNLL 2012*, July 12–14, 2012, Jeju Island, Korea (pp. 1302–1312).

Kanda, T., Shimada, M., Koizumi, S. (2012). Children learning with a social robot. In *The Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction*, ACM, New York (pp. 351–358).

Kuhn, D. (1991). *The skills of argument*. New York: Cambridge University Press.

Kumar, R., & Rosé, C. P. (2011). Architecture for building conversational agents that support collaborative learning. *IEEE Transactions on Learning, 4*(1), 21–34.

Kumar, R., Rosé, C.P., Litman, D. (2006). Identification of confusion and surprise in spoken dialogusing prosodic features. In *Proceedings of Interspeech*, Pittsburgh, PA.

Kumar, R., Rosé, C.P., Wang, Y.C., Joshi, M., Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. In *Proceedings of the 2007 Conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work* (pp 383–390).

Labov, W. (2010). *Principles of linguistic change: Internal factors, volume 1*. Wiley-Blackwell.

Lauritzen, S. L., & Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, 50*, 157–224.

Levitan, R., Gravano, A., Hirschberg, J. (2011). Entrainment in speech preceding backchannels. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*: *Short Papers-Volume 2* (pp. 113–117). Association for Computational Linguistics.

Liscombe, J., Venditti, J, Hirschberg, J. (2005). Detecting certainness in spoken tutorial dialogues. In *Proceedings Interspeech* (pp. 1837–1840).

Loehlin, J. (1998). *Latent variable models*: *An introduction to factor, path, and structural equation analysis*. Psychology Press.

McLaren, B., Scheuer, O., De Laat, M., Hever, R., de Groot, R., Rosé, C.P. (2007). Using machine learning techniques to analyze and support mediation of student e-discussions. In *Proceedings of Artificial Intelligence in Education* (pp. 331–338). IOS Press.

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.

Nokes, T., Levine, J.M., Belenky, D., Gadgil, S. (2010). Investigating the impact of dialectical interaction on engagement, affect, and robust learning. In *Proceedings of the 9th International Conference of the Learning Sciences* (pp. 215–218). International Society of the Learning Sciences.

OpenSmile. (2012). Available at http://opensmile.sourceforge.net/.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.

Purcell, A. (1984). Code shifting hawaiian style: Childrens accommodation along a decreolizing continuum. *International Journal of the Sociology of Language, 1984*(46), 71–86.

Putman, W., & Street, R., Jr. (1984). The conception and perception of noncontent speech performance: Implications for speech-accommodation theory. *International Journal of the Sociology of Language, 46*, 97–114.

Ranganath, R., Jurafsky, D., McFarland, D. (2009). It's not you, it's me: Detecting flirting and its misperception in speed-dates. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 334–342). Singapore.

Rosé, C. P., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., et al. (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.

Sanders, R. (1987). *Cognitive foundations of calculated speech*. State University of New York Press.

Sawyer, R.K. (2005). *Social emergence: Societies as complex systems*. Cambridge University Press.

Schwartz, D. (1998). The productive agency that drives collaborative learning. In P. Dillenbourg (Ed.) *Collaborative learning: Cognitive and computational approaches*. Emrald Group Publishing.

Scotton, C. (1985). What the heck, sir: Style shifting and lexical colouring as features of powerful language. In *Sequence and pattern in communicative behaviour* (pp. 103–119).

Soller, A., & Lesgold, A. (2000). Modeling the process of collaborative learning. In *Proceedings of the International Workshop on New Technologies in Collaborative Learning*. Japan: Awaiji–Yumebutai.

Suthers, D. (2006). Technology affordances for inter-subjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning, 1*, 315–337.

Teasley, S. D. (1997). Talking about reasoning: How important is the peer in peer collaborations? In L. B. Resnick, C. Pontecorvo, & R. Saljo (Eds.), *Discourse, tools, and reasoning: Situated cognition and technologically supported environments*. Heidelberg: Springer.

Toulmin. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer supported collaborative learning. *Computers in Education, 46*, 71–95.

Welkowitz, J., & Feldstein, S. (1970). Relation of experimentally manipulated interpersonal perception and psychological differentiation to the temporal patterning of conversation. In *Proceedings of the 78th Annual Convention of the American Psychological Association* (volume 5, pp. 387–388).