An Affect-Enriched Dialogue Act Classification Model for Task-Oriented Dialogue

Kristy Elizabeth Boyer | **Joseph F. Grafsgaard** | **Eun Young Ha** | **Robert Phillips*** | **James C. Lester**

Department of Computer Science
North Carolina State University
Raleigh, NC, USA

* Dual Affiliation with Applied Research Associates, Inc.
Raleigh, NC, USA

{keboyer, jfgrafsg, eha, rphilli, lester}@ncsu.edu

Abstract

Dialogue act classification is a central challenge for dialogue systems. Although the importance of emotion in human dialogue is widely recognized, most dialogue act classification models make limited or no use of affective channels in dialogue act classification. This paper presents a novel affect-enriched dialogue act classifier for task-oriented dialogue that models facial expressions of users, in particular, facial expressions related to confusion. The findings indicate that the affect-enriched classifiers perform significantly better for distinguishing user requests for feedback and grounding dialogue acts within textual dialogue. The results point to ways in which dialogue systems can effectively leverage affective channels to improve dialogue act classification.

1 Introduction

Dialogue systems aim to engage users in rich, adaptive natural language conversation. For these systems, understanding the role of a user's utterance in the broader context of the dialogue is a key challenge (Sridhar, Bangalore, & Narayanan, 2009). Central to this endeavor is dialogue act classification, which categorizes the intention behind the user's move (e.g., asking a question, providing declarative information). Automatic dialogue act classification has been the focus of a large body of research, and a variety of approaches, including sequential models (Stolcke et al., 2000), vector-based models (Sridhar, Bangalore, & Narayanan, 2009), and most recently, feature-enhanced latent semantic analysis (Di Eugenio, Xie, & Serafin, 2010), have shown promise. These models may be further improved by leveraging regularities of the dialogue from both linguistic and extra-linguistic sources. Users' expressions of emotion are one such source.

Human interaction has long been understood to include rich phenomena consisting of verbal and nonverbal cues, with facial expressions playing a vital role (Knapp & Hall, 2006; McNeill, 1992; Mehrabian, 2007; Russell, Bachorowski, & Fernandez-Dols, 2003; Schmidt & Cohn, 2001). While the importance of emotional expressions in dialogue is widely recognized, the majority of dialogue act classification projects have focused either peripherally (or not at all) on emotion, such as by leveraging acoustic and prosodic features of spoken utterances to aid in online dialogue act classification (Sridhar, Bangalore, & Narayanan, 2009). Other research on emotion in dialogue has involved detecting affect and adapting to it within a dialogue system (Forbes-Riley, Rotaru, Litman, & Tetreault, 2009; López-Cózar, Silovsky, & Griol, 2010), but this work has not explored leveraging affect information for automatic user dialogue act classification. Outside of dialogue, sentiment analysis within discourse is an active area of research (López-Cózar et al., 2010), but it is generally lim-

Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 1190–1199,
Portland, Oregon, June 19-24, 2011.

ited to modeling textual features and not multimodal expressions of emotion such as facial actions. Such multimodal expressions have only just begun to be explored within corpus-based dialogue research (Calvo & D'Mello, 2010; Cavicchio, 2009).

This paper presents a novel affect-enriched dialogue act classification approach that leverages knowledge of users' facial expressions during computer-mediated textual human-human dialogue. Intuitively, the user's affective state is a promising source of information that may help to distinguish between particular dialogue acts (e.g., a *confused* user may be more likely to ask a question). We focus specifically on occurrences of students' confusion-related facial actions during task-oriented tutorial dialogue.

Confusion was selected as the focus of this work for several reasons. First, confusion is known to be prevalent within tutoring, and its implications for student learning are thought to run deep (Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005). Second, while identifying the "ground truth" of emotion based on any external display by a user presents challenges, prior research has demonstrated a correlation between particular facial action units and confusion during learning (Craig, D'Mello, Witherspoon, Sullins, & Graesser, 2004; D'Mello, Craig, Sullins, & Graesser, 2006; McDaniel et al., 2007). Finally, automatic facial action recognition technologies are developing rapidly, and confusion-related facial action events are among those that can be reliably recognized automatically (Bartlett et al., 2006; Cohn, Reed, Ambadar, Xiao, & Moriyama, 2004; Pantic & Bartlett, 2007; Zeng, Pantic, Roisman, & Huang, 2009). This promising development bodes well for the feasibility of automatic real-time confusion detection within dialogue systems.

2 Background and Related Work

2.1 Dialogue Act Classification

Because of the importance of dialogue act classification within dialogue systems, it has been an active area of research for some time. Early work on automatic dialogue act classification modeled discourse structure with hidden Markov models, experimenting with lexical and prosodic features, and applying the dialogue act model as a constraint to aid in automatic speech recognition (Stolcke et al., 2000). In contrast to this sequential modeling approach, which is best suited to offline processing, recent work has explored how lexical, syntactic, and prosodic features perform for online dialogue act tagging (when only partial dialogue sequences are available) within a maximum entropy framework (Sridhar, Bangalore, & Narayanan, 2009). A recently proposed alternative approach involves treating dialogue utterances as documents within a latent semantic analysis framework, and applying feature enhancements that incorporate such information as speaker and utterance duration (Di Eugenio et al., 2010). Of the approaches noted above, the modeling framework presented in this paper is most similar to the vector-based maximum entropy approach of Sridhar et al. (2009). However, it takes a step beyond the previous work by including multimodal affective displays, specifically facial expressions, as features available to an affect-enriched dialogue act classification model.

2.2 Detecting Emotions in Dialogue

Detecting emotional states during spoken dialogue is an active area of research, much of which focuses on detecting frustration so that a user can be automatically transferred to a human dialogue agent (López-Cózar et al., 2010). Research on spoken dialogue has leveraged lexical features along with discourse cues and acoustic information to classify user emotion, sometimes at a coarse grain along a positive/negative axis (Lee & Narayanan, 2005). Recent work on an affective companion agent has examined user emotion classification within conversational speech (Cavazza et al., 2010). In contrast to that spoken dialogue research, the work in this paper is situated within textual dialogue, a widely used modality of communication for which a deeper understanding of user affect may substantially improve system performance.

While many projects have focused on linguistic cues, recent work has begun to explore numerous channels for affect detection including facial actions, electrocardiograms, skin conductance, and posture sensors (Calvo & D'Mello, 2010). A recent project in a map task domain investigates some of these sources of affect data within task-oriented dialogue (Cavicchio, 2009). Like that work, the current project utilizes facial action tagging, for

which promising automatic technologies exist (Bartlett et al., 2006; Pantic & Bartlett, 2007; Zeng, Pantic, Roisman, & Huang, 2009). However, we leverage the recognized expressions of emotion for the task of dialogue act classification.

2.3 Categorizing Emotions within Dialogue and Discourse

Sets of emotion taxonomies for discourse and dialogue are often application-specific, for example, focusing on the frustration of users who are interacting with a spoken dialogue system (López-Cózar et al., 2010), or on uncertainty expressed by students while interacting with a tutor (Forbes-Riley, Rotaru, Litman, & Tetreault, 2007). In contrast, the most widely utilized emotion frameworks are not application-specific; for example, Ekman's Facial Action Coding System (FACS) has been widely used as a rigorous technique for coding facial movements based on human facial anatomy (Ekman & Friesen, 1978). Within this framework, facial movements are categorized into facial action units, which represent discrete movements of muscle groups. Additionally, facial action descriptors (for movements not derived from facial muscles) and movement and visibility codes are included. Ekman's basic emotions (Ekman, 1999) have been used in recent work on classifying emotion expressed within blog text (Das & Bandyopadhyay, 2009), while other recent work (Nguyen, 2010) utilizes Russell's core affect model (Russell, 2003) for a similar task.

During tutorial dialogue, students may not frequently experience Ekman's basic emotions of *happiness*, *sadness*, *anger*, *fear*, *surprise*, and *disgust*. Instead, students appear to more frequently experience cognitive-affective states such as *flow* and *confusion* (Calvo & D'Mello, 2010). Our work leverages Ekman's facial tagging scheme to identify a particular facial action unit, Action Unit 4 (AU4), that has been observed to correlate with confusion (Craig, D'Mello, Witherspoon, Sullins, & Graesser, 2004; D'Mello, Craig, Sullins, & Graesser, 2006; McDaniel et al., 2007).

2.4 Importance of Confusion in Tutorial Dialogue

Among the affective states that students experience during tutorial dialogue, confusion is prevalent, and its implications for student learning are significant. Confusion is associated with *cognitive disequilibrium*, a state in which students' existing knowledge is inconsistent with a novel learning experience (Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005). Students may express such confusion within dialogue as *uncertainty*, to which human tutors often adapt in a context-dependent fashion (Forbes-Riley et al., 2007). Moreover, implementing adaptations to student uncertainty within a dialogue system can improve the effectiveness of the system (Forbes-Riley et al., 2009).

For tutorial dialogue, the importance of understanding student utterances is paramount for a system to positively impact student learning (Dzikovska, Moore, Steinhauser, & Campbell, 2010). The importance of frustration as a cognitive-affective state during learning suggests that the presence of student confusion may serve as a useful constraining feature for dialogue act classification of student utterances. This paper explores the use of facial expression features in this way.

3 Task-Oriented Dialogue Corpus

The corpus was collected during a textual human-human tutorial dialogue study in the domain of introductory computer science (Boyer, Phillips, et al., 2010). Students solved an introductory computer programming problem and carried on textual dialogue with tutors, who viewed a synchronized version of the students' problem-solving workspace. The original corpus consists of 48 dialogues, one per student. Each student interacted with one of two tutors. Facial videos of students were collected using built-in webcams, but were not shown to the tutors. Video quality was ranked based on factors such as obscured foreheads due to hats or hair, and improper camera position resulting in students' faces not being fully captured on the video. The highest-quality set contained 14 videos, and these videos were used in this analysis. They have a total running time of 11 hours and 55 minutes, and include dialogues with three female subjects and eleven male subjects.

3.1 Dialogue act annotation

The dialogue act annotation scheme (Table 1) was applied manually. The kappa statistic for inter-annotator agreement on a 10% subset of the corpus was κ=0.80, indicating good reliability.

Table 1. Dialogue act tags and relative frequencies across fourteen dialogues in video corpus

Student Dialogue Act	Example	Rel. Freq.
EXTRA-DOMAIN (EX)	*Little sleep deprived today*	.08
GROUNDING (G)	*Ok* or *Thanks*	.21
NEGATIVE FEEDBACK WITH ELABORATION (NE)	*I'm still confused on what this next for loop is doing.*	.02
NEGATIVE FEEDBACK (N)	*I don't see the diff.*	.04
POSITIVE FEEDBACK WITH ELABORATION (PE)	*It makes sense now that you explained it, but I never used an else if in any of my other programs*	.04
POSITIVE FEEDBACK (P)	*Second part complete.*	.11
QUESTION (Q)	*Why couldn't I have said if (i<5)*	.11
STATEMENT (S)	*i is my only index*	.07
REQUEST FOR FEEDBACK (RF)	*So I need to create a new method that sees how many elements are in my array?*	.16
RESPONSE (RSP)	*You mean not the length but the contents*	.14
UNCERTAIN FEEDBACK WITH ELABORATION (UE)	*I'm trying to remember how to copy arrays*	.008
UNCERTAIN FEEDBACK (U)	*Not quite yet*	.008

3.2 Task action annotation

The tutoring sessions were task-oriented, focusing on a computer programming exercise. The task had several subtasks consisting of programming modules to be implemented by the student. Each of those subtasks also had numerous fine-grained goals, and student task actions either contributed or did not contribute to the goals. Therefore, to obtain a rich representation of the task, a manual annotation along two dimensions was conducted (Boyer, Phillips, et al., 2010). First, the subtask structure was annotated hierarchically, and then each task action was labeled for correctness according to the requirements of the assignment. Inter-annotator agreement was computed on 20% of the corpus at the leaves of the subtask tagging scheme, and resulted in a simple kappa of κ=.56. However, the leaves of the annotation scheme feature an implicit ordering (subtasks were completed in order, and adjacent subtasks are semantically more similar than subtasks at a greater distance); therefore, a weighted kappa is also meaningful to consider for this annotation. The weighted kappa is $\kappa_{weighted}$=.80. An annotated excerpt of the corpus is displayed in Table 2.

Table 2. Excerpt from corpus illustrating annotations and interplay between dialogue and task

13:38:09	Student:	How do I know where to end? **[RF]**
13:38:26	Tutor:	Well you told me how to get how many elements in an array by using .length right?
13:38:26	Student:	**[Task action: Subtask 1-a-iv, Buggy]**
13:38:56	Tutor:	Great
13:38:56	Student:	**[Task action: Subtask 1-a-v, Correct]**
13:39:35	Student:	Well is it "array.length"? **[RF]** ****Facial Expression: AU4**
13:39:46	Tutor:	You just need to use the correct array name
13:39:46	Student:	**[Task action: Subtask 1-a-iv, Buggy]**

3.3 Lexical and Syntactic Features

In addition to the manually annotated dialogue and task features described above, syntactic features of each utterance were automatically extracted using the Stanford Parser (De Marneffe et al., 2006). From the phrase structure trees, we extracted the top-most syntactic node and its first two children. In the case where an utterance consisted of more than one sentence, only the phrase structure tree of the first sentence was considered. Individual word tokens in the utterances were further processed with the Porter Stemmer (Porter, 1980) in the NLTK package (Loper & Bird, 2004). Our prior work has shown that these lexical and syntactic features are highly predictive of dialogue acts during task-oriented tutorial dialogue (Boyer, Ha et al. 2010).

4 Facial Action Tagging

An annotator who was certified in the Facial Action Coding System (FACS) (Ekman, Friesen, & Hager, 2002) tagged the video corpus consisting of fourteen dialogues. The FACS certification process requires annotators to pass a test designed to analyze their agreement with reference coders on a set of spontaneous facial expressions (Ekman & Rosenberg, 2005). This annotator viewed the videos continuously and paused the playback whenever notable facial displays of Action Unit 4 (AU4: Brow Lowerer) were seen. This action unit was chosen for this study based on its correlations with confusion in prior research (Craig, D'Mello, Witherspoon, Sullins, & Graesser, 2004; D'Mello, Craig, Sullins, & Graesser, 2006; McDaniel et al., 2007).

To establish reliability of the annotation, a second FACS-certified annotator independently annotated 36% of the video corpus (5 of 14 dialogues), chosen randomly after stratification by gender and tutor. This annotator followed the same method as the first annotator, pausing the video at any point to tag facial action events. At any given time in the video, the coder was first identifying whether an action unit event existed, and then describing the facial movements that were present. The annotators also specified the beginning and ending time of each event. In this way, the action unit event tags spanned discrete durations of varying length, as specified by the coders. Because the two coders were not required to tag at the same point in time, but rather were permitted the freedom to stop the video at any point where they felt a notable facial action event occurred, calculating agreement between annotators required discretizing the continuous facial action time windows across the tutoring sessions. This discretization was performed at granularities of 1/4, 1/2, 3/4, and 1 second, and inter-rater reliability was calculated at each level of granularity (Table 3). Windows in which both annotators agreed that no facial action event was present were tagged by default as *neutral*. Figure 1 illustrates facial expressions that display facial Action Unit 4.

Table 3. Kappa values for inter-annotator agreement on facial action events

	Granularity			
	¼ sec	½ sec	¾ sec	1 sec
Presence of AU4 (Brow Lowerer)	*.84*	*.87*	*.86*	*.86*

Figure 1. Facial expressions displaying AU4 (Brow Lowerer)

Despite the fact that promising automatic approaches exist to identifying many facial action units (Bartlett et al., 2006; Cohn, Reed, Ambadar, Xiao, & Moriyama, 2004; Pantic & Bartlett, 2007; Zeng, Pantic, Roisman, & Huang, 2009), manual annotation was selected for this project for two reasons. First, manual annotation is more robust than automatic recognition of facial action units, and manual annotation facilitated an exploratory, comprehensive view of student facial expressions during learning through task-oriented dialogue. Although a detailed discussion of the other emotions present in the corpus is beyond the scope of this paper, Figure 2 illustrates some other spontaneous student facial expressions that differ from those associated with confusion.

Figure 2. Other facial expressions from the corpus

5 Models

The goal of the modeling experiment was to determine whether the addition of confusion-related facial expression features significantly boosts dialogue act classification accuracy for student utterances.

5.1 Features

We take a vector-based approach, in which the features consist of the following:

Utterance Features

- *Dialogue act features*: Manually annotated dialogue act for the past three utterances. These features include tutor dialogue acts, annotated with a scheme analogous to that used to annotate student utterances (Boyer et al., 2009).
- *Speaker*: Speaker for past three utterances
- *Lexical features*: Word unigrams
- *Syntactic features*: Top-most syntactic node and its first two children

Task-based Features

- *Subtask*: Hierarchical subtask structure for past three task actions (semantic programming actions taken by student)
- *Correctness*: Correctness of past three task actions taken by student
- *Preceded by task*: Indicator for whether the most recent task action immediately preceded the target utterance, or whether it was immediately preceded by the last dialogue move

Facial Expression Features

- *AU4_1sec:* Indicator for the display of the brow lowerer within 1 second prior to this utterance being sent, for the most recent three utterances
- *AU4_5sec:* Indicator for the display of the brow lowerer within 5 seconds prior to this utterance being sent, for the most recent three utterances
- *AU4_10sec:* Indicator for the display of the brow lowerer within 10 seconds prior to this utterance being sent, for the most recent three utterances

5.2 Modeling Approach

A logistic regression approach was used to classify the dialogue acts based on the above feature vectors. The Weka machine learning toolkit (Hall et al., 2009) was used to learn the models and to first perform feature selection in a best-first search. Logistic regression is a generalized maximum likelihood model that discriminates between pairs of output values by calculating a feature weight vector over the predictors.

The goal of this work is to explore the utility of confusion-related facial features in the context of particular dialogue act types. For this reason, a specialized classifier was learned by dialogue act.

5.3 Classification Results

The classification accuracy and kappa for each specialized classifier is displayed in Table 4. Note that kappa statistics adjust for the accuracy that would be expected by majority-baseline chance; a kappa statistic of zero indicates that the classifier performed equal to chance, and a positive kappa statistic indicates that the classifier performed better than chance. A kappa of 1 constitutes perfect agreement. As the table illustrates, the feature selection chose to utilize the AU4 feature for every dialogue act except STATEMENT (S). When considering the accuracy of the model across the ten folds, two of the affect-enriched classifiers exhibited statistically significantly better performance. For GROUNDING (G) and REQUEST FOR FEEDBACK (RF), the facial expression features significantly

improved the classification accuracy compared to a model that was learned without affective features.

6 Discussion

Dialogue act classification is an essential task for dialogue systems, and it has been addressed with a variety of modeling approaches and feature sets. We have presented a novel approach that treats facial expressions of students as constraining features for an affect-enriched dialogue act classification model in task-oriented tutorial dialogue. The results suggest that knowledge of the student's confusion-related facial expressions can significantly enhance dialogue act classification for two types of dialogue acts, GROUNDING and REQUEST FOR FEEDBACK.

Table 4. Classification accuracy and kappa for specialized DA classifiers. Statistically significant differences (across ten folds, one-tailed *t*-test) are shown in bold.

Dialogue Act	Classifier with AU4		Classifier without AU4		*p*-value
	% acc	κ	% acc	κ	
EX	90.7	.62	89.0	.28	>.05
G	**92.6**	**.76**	**91**	**.71**	**.018**
P	93	.49	92.2	.40	>.05
Q	94.6	.72	94.2	.72	>.05
S	Not chosen in feat. sel.		93	.22	n/a
RF	**90.7**	**.62**	**88.3**	**.53**	**.003**
RSP	93	.68	95	.75	>.05
NE	*		*		
N	*		*		
PE	*		*		
U	*		*		
UE	*		*		

*Too few instances for ten-fold cross-validation.

6.1 Features Selected for Classification

Out of more than 1500 features available during feature selection, each of the specialized dialogue act classifiers selected between 30 and 50 features in each condition (with and without affect features). To gain insight into the specific features that were useful for classifying these dialogue acts, it is useful to examine which of the AU4 history features were chosen during feature selection.

For GROUNDING, features that indicated the presence of absence of AU4 in the immediately preceding utterance, either at the 1 second or 5 second granularity, were selected. Absence of this confusion-related facial action unit was associated with a higher probability of a grounding act, such as an acknowledgement. This finding is consistent with our understanding of how students and tutors interacted in this corpus; when a student experienced confusion, she would be unlikely to then make a simple grounding dialogue move, but instead would tend to inspect her computer program, ask a question, or wait for the tutor to explain more.

For REQUEST FOR FEEDBACK, the predictive features were presence or absence of AU4 within ten seconds of the longest available history (three turns in the past), as well as the presence of AU4 within five seconds of the current utterance (the utterance whose dialogue act is being classified). This finding suggests that there may be some lag between the student experiencing confusion and then choosing to make a request for feedback, and that the confusion-related facial expressions may re-emerge as the student is making a request for feedback, since the five-second window prior to the student sending the textual dialogue message would overlap with the student's construction of the message itself.

Although the improvements seen with AU4 features for QUESTION, POSITIVE FEEDBACK, and EXTRA-DOMAIN acts were not statistically reliable, examining the AU4 features that were selected for classifying these moves points toward ways in which facial expressions may influence classification of these acts (Table 5).

Table 5. Number of features, and AU4 features selected, for specialized DA classifiers

Dialogue Act	# features selected	AU4 features selected
G	43	One utterance ago: AU4_1sec, AU4_5sec
RF	37	Three utterances ago: AU4_10sec Target utterance: AU4_5sec
EX	50	Three utterances ago: AU4_1sec
P	36	Current utterance: AU4_10sec
Q	30	One utterance ago: AU4_5sec

6.2 Implications

The results presented here demonstrate that leveraging knowledge of user affect, in particular of spontaneous facial expressions, may improve the performance of dialogue act classification models. Perhaps most interestingly, displays of confusion-related facial actions prior to a student dialogue move enabled an affect-enriched classifier to recognize requests for feedback with significantly greater accuracy than a classifier that did not have access to the facial action features. Feedback is known to be a key component of effective tutorial dialogue, through which tutors provide adaptive help (Shute, 2008). Requesting feedback also seems to be an important behavior of students, characteristically engaged in more frequently by women than men, and more frequently by students with lower incoming knowledge than by students with higher incoming knowledge (Boyer, Vouk, & Lester, 2007).

6.3 Limitations

The experiments reported here have several notable limitations. First, the time-consuming nature of manual facial action tagging restricted the number of dialogues that could be tagged. Although the highest quality videos were selected for annotation, other medium quality videos would have been sufficiently clear to permit tagging, which would have increased the sample size and likely revealed statistically significant trends. For example, the performance of the affect-enriched classifier was better for dialogue acts of interest such as positive feedback and questions, but this difference was not statistically reliable.

An additional limitation stems from the more fundamental question of which affective states are indicated by particular external displays. The field is only just beginning to understand facial expressions during learning and to correlate these facial actions with emotions. Additional research into the "ground truth" of emotion expression will shed additional light on this area. Finally, the results of manual facial action annotation may constitute upper-bound findings for applying automatic facial expression analysis to dialogue act classification.

7 Conclusions and Future Work

Emotion plays a vital role in human interactions. In particular, the role of facial expressions in human-human dialogue is widely recognized. Facial expressions offer a promising channel for understanding the emotions experienced by users of dialogue systems, particularly given the ubiquity of webcam technologies and the increasing number of dialogue systems that are deployed on webcam-enabled devices. This paper has reported on a first step toward using knowledge of user facial expressions to improve a dialogue act classification model for tutorial dialogue, and the results demonstrate that facial expressions hold great promise for distinguishing the pedagogically relevant dialogue act REQUEST FOR FEEDBACK, and the conversational moves of GROUNDING.

These early findings highlight the importance of future work in this area. Dialogue act classification models have not fully leveraged some of the techniques emerging from work on sentiment analysis. These approaches may prove particularly useful for identifying emotions in dialogue utterances. Another important direction for future work involves more fully exploring the ways in which affect expression differs between textual and spoken dialogue. Finally, as automatic facial tagging technologies mature, they may prove powerful enough to enable broadly deployed dialogue systems to feasibly leverage facial expression data in the near future.

Acknowledgments

This work is supported in part by the North Carolina State University Department of Computer Science and by the National Science Foundation through Grants REC-0632450, IIS-0812291, DRL-1007962 and the STARS Alliance Grant CNS-0739216. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the participants, and do not necessarily represent the official views, opinions, or policy of the National Science Foundation.

References

A. Andreevskaia and S. Bergler. 2008. When specialists and generalists work together: Overcoming domain dependence in sentiment tagging. Proceedings of the Annual Meeting of the Association for Computational Linguistics and Human Language Technologies (ACL HLT), 290-298.

M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. 2006. Fully Automatic Facial Action Recognition in Spontaneous Behavior. 7th International Conference on Automatic Face and Gesture Recognition (FGR06), 223-230.

K.E. Boyer, M. Vouk, and J.C. Lester. 2007. The influence of learner characteristics on task-oriented tutorial dialogue. Proceedings of the International Conference on Artificial Intelligence in Education, 365–372.

K.E. Boyer, E.Y. Ha, R. Phillips, M.D. Wallis, M. Vouk, and J.C. Lester. 2010. Dialogue act modeling in a complex task-oriented domain. Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), 297-305.

K.E. Boyer, R. Phillips, E.Y. Ha, M.D. Wallis, M.A. Vouk, and J.C. Lester. 2009. Modeling dialogue structure with adjacency pair analysis and hidden Markov models. Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Short Papers, 49-52.

K.E. Boyer, R. Phillips, E.Y. Ha, M.D. Wallis, M.A. Vouk, and J.C. Lester. 2010. Leveraging hidden dialogue state to select tutorial moves. Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications, 66-73.

R.A. Calvo and S. D'Mello. 2010. Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. IEEE Transactions on Affective Computing, 1(1): 18-37.

M. Cavazza, R.S.D.L. Cámara, M. Turunen, J. Gil, J. Hakulinen, N. Crook, et al. 2010. How was your day? An affective companion ECA prototype. Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), 277-280.

F. Cavicchio. 2009. The modulation of cooperation and emotion in dialogue: the REC Corpus. Proceedings of the ACL-IJCNLP 2009 Student Research Workshop, 43-48.

J.F. Cohn, L.I. Reed, Z. Ambadar, J. Xiao, and T. Moriyama. 2004. Automatic Analysis and Recognition of Brow Actions and Head Motion in Spontaneous Facial Behavior. IEEE International Conference on Systems, Man and Cybernetics, 610-616.

S.D. Craig, S. D'Mello, A. Witherspoon, J. Sullins, and A.C. Graesser. 2004. Emotions during learning: The first steps toward an affect sensitive intelligent tutoring system. In J. Nall and R. Robson (Eds.), E-learn 2004: World conference on E-learning in Corporate, Government, Healthcare, & Higher Education, 241-250.

D. Das and S. Bandyopadhyay. 2009. Word to sentence level emotion tagging for Bengali blogs. Proceedings of the ACL-IJCNLP Conference, Short Papers, 149-152.

S. Dasgupta and V. Ng. 2009. Mine the easy, classify the hard: a semi-supervised approach to automatic sentiment classification. Proceedings of the 46th Annual Meeting of the ACL and the 4th IJCNLP, 701-709.

B. Di Eugenio, Z. Xie, and R. Serafin. 2010. Dialogue Act Classification, Higher Order Dialogue Structure, and Instance-Based Learning. Dialogue & Discourse, 1(2): 1-24.

M. Dzikovska, J.D. Moore, N. Steinhauser, and G. Campbell. 2010. The impact of interpretation problems on tutorial dialogue. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Short Papers, 43-48.

S. D'Mello, S.D. Craig, J. Sullins, and A.C. Graesser. 2006. Predicting Affective States expressed through an Emote-Aloud Procedure from AutoTutor's Mixed- Initiative Dialogue. International Journal of Artificial Intelligence in Education, 16(1): 3-28.

P. Ekman. 1999. Basic Emotions. In T. Dalgleish and M. J. Power (Eds.), Handbook of Cognition and Emotion. New York: Wiley.

P. Ekman, W.V. Friesen. 1978. Facial Action Coding System. Palo Alto, CA: Consulting Psychologists Press.

P. Ekman, W.V. Friesen, and J.C. Hager. 2002. Facial Action Coding System: Investigator's Guide. Salt Lake City, USA: A Human Face.

P. Ekman and E.L. Rosenberg (Eds.). 2005. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS) (2nd ed.). New York: Oxford University Press.

K. Forbes-Riley, M. Rotaru, D.J. Litman, and J. Tetreault. 2007. Exploring affect-context dependencies for adaptive system development. The Conference of the North American Chapter of the Association for Computational Linguistics and Human Language Technologies (NAACL HLT), Short Papers, 41-44.

K. Forbes-Riley, M. Rotaru, D.J. Litman, and J. Tetreault. 2009. Adapting to student uncertainty improves tutoring dialogues. Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED), 33-40.

A.C. Graesser, S. Lu, B. Olde, E. Cooper-Pye, and S. Whitten. 2005. Question asking and eye tracking during cognitive disequilibrium: comprehending illustrated texts on devices when the devices break down. Memory & Cognition, 33(7): 1235-1247.

S. Greene and P. Resnik. 2009. More than words: Syntactic packaging and implicit sentiment. Proceedings of the 2009 Annual Conference of the North American Chapter of the ACL and Human Language Technologies (NAACL HLT), 503-511.

M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten. 2009. The WEKA data mining software: An update. SIGKDD Explorations, 11(1): 10–18.

R. Iida, S. Kobayashi, and T. Tokunaga. 2010. Incorporating extra-linguistic information into reference resolution in collaborative task dialogue. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 1259-1267.

M.L. Knapp and J.A. Hall. 2006. Nonverbal Communication in Human Interaction (6th ed.). Belmont, CA: Wadsworth/Thomson Learning.

C.M. Lee, S.S. Narayanan. 2005. Toward detecting emotions in spoken dialogs. IEEE Transactions on Speech and Audio Processing, 13(2): 293-303.

R. López-Cózar, J. Silovsky, and D. Griol. 2010. F2– New Technique for Recognition of User Emotional States in Spoken Dialogue Systems. Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), 281-288.

B.T. McDaniel, S. D'Mello, B.G. King, P. Chipman, K. Tapp, and A.C. Graesser. 2007. Facial Features for Affective State Detection in Learning Environments. Proceedings of the 29th Annual Cognitive Science Society, 467-472.

D. McNeill. 1992. Hand and mind: What gestures reveal about thought. Chicago: University of Chicago Press.

A. Mehrabian. 2007. Nonverbal Communication. New Brunswick, NJ: Aldine Transaction.

T. Nguyen. 2010. Mood patterns and affective lexicon access in weblogs. Proceedings of the ACL 2010 Student Research Workshop, 43-48.

M. Pantic and M.S. Bartlett. 2007. Machine Analysis of Facial Expressions. In K. Delac and M. Grgic (Eds.), Face Recognition, 377-416. Vienna, Austria: I-Tech Education and Publishing.

J.A. Russell. 2003. Core affect and the psychological construction of emotion. Psychological Review, 110(1): 145-172.

J.A. Russell, J.A. Bachorowski, and J.M. Fernandez-Dols. 2003. Facial and vocal expressions of emotion. Annual Review of Psychology, 54, 329-49.

K.L. Schmidt and J.F. Cohn. 2001. Human Facial Expressions as Adaptations: Evolutionary Questions in Facial Expression Research. Am J Phys Anthropol, 33: 3-24.

V.J. Shute. 2008. Focus on Formative Feedback. Review of Educational Research, 78(1): 153-189.

V.K.R Sridar, S. Bangalore, and S.S. Narayanan. 2009. Combining lexical, syntactic and prosodic cues for improved online dialog act tagging. Computer Speech & Language, 23(4): 407-422. Elsevier Ltd.

A. Stolcke, K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, et al. 2000. Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. Computational Linguistics, 26(3): 339-373.

C. Toprak, N. Jakob, and I. Gurevych. 2010. Sentence and expression level annotation of opinions in user-generated discourse. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, 575-584.

T. Wilson, J. Wiebe, and P. Hoffmann. 2009. Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis. Computational Linguistics, 35(3): 399-433.

Z. Zeng, M. Pantic, G.I. Roisman, and T.S. Huang. 2009. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(1): 39-58.