Comments of Journalism Mentors on News Stories: Classification and Epistemic Status of Mentor Contributions

Art Graesser[1], Zhiqiang Cai[1], Jonathan Wood[1], David Hatfield[2], Elizabeth Bagley[2], Padraig Nash[2], David Shaffer[2]

[1] Institute for Intelligent Systems, 365 Innovation Drive, University of Memphis, Memphis, TN 38152, {a-graesser, zcai, jwood}@memphis.edu
[2] Educational Sciences, Department of Educational Psychology, 1025 W. Johnson Street, University of Wisconsin, Madison, WI 53706, {ebagley, pnash, dhatfield, dws}@ef-games.com

Abstract. We identified the speech act categories and clusters of discourse comments of journalism mentors who interact with students editing news stories. Two important speech act categories are evaluations and suggestions. Latent semantic analysis and principal components analyses helped us discover clusters of comments involving evaluations and suggestions. The comments of mentors were also significantly aligned with epistemic frame elements that motivate the comments at a deeper level of discourse and pedagogy. Such alignments were validated by logistic regression analyses on a sample of hand-coded judgments of the frame elements. There was some modest transfer from a journalism practicum corpus to a game corpus. These analyses provide an important first step in building a virtual AutoMentor for multiparty epistemic games.

Keywords: virtual agents, serious games, epistemic games, discourse, latent semantic analysis

1 Introduction

Our long-term goal is to build an automated virtual mentor, called AutoMentor, that provides guidance to students as they interact in groups in serious games. AutoMentor will vigilantly observe the game states and interactions between players and will periodically offer comments and suggestions to promote learning and productive conversation. Shaffer and his colleagues [1, 2, 3] have established the need for human mentors to promote learning when students interact with multiparty games, such as *Urban Science*, *Science.net*, or the *Land Science* game under current development and testing. These games help students understand the kinds of problems and problem solving that socially valued professions routinely engage in, such as how the development of cities and suburbs are influenced by zoning, roads, parks, housing, and economic investment, or how important developments in science can be

communicated through narrative details and attention to accuracy and source attribution. Student learning is severely limited, however, if there is no mentorship and expertise from professional stakeholders. Our epistemic games group is therefore analyzing the verbal interactions between human mentors and students, with the hope of automating their language and discourse contributions in AutoMentor.

AutoMentor can be viewed as an augmentation of AutoTutor [4, 5, 6], a pedagogical agent that helps students learn by holding a conversation in natural language. The original AutoTutor was developed for one-on-one tutoring in language, but more recent versions have involved trialogs with two agents interacting with one student [7] and interactions among the student, an AutoTutor agent, and an external simulation environment [8]. AutoMentor moves beyond these AutoTutor versions by having a single virtual agent interact with multiple students in groups as they interact with a complex simulation game on urban planning and environmental science.

The quality of such conversational agents depends on their ability to understand and generate discourse. Discourse has multiple levels of analysis that have been identified by researchers in discourse processes [9, 10, 11] and computational linguistics [12]. According to one multilevel discourse framework [11], the levels include the surface code (wording and syntax), the explicit textbase, the referential situation model, genre and rhetorical structure, and pragmatic communication. Discourse becomes more complex as one moves from dialogues to multiparty conversations [9] and from minimal external environments to complex dynamic external environments (such as simulation games). The field of computational linguistics has not advanced to the point of accurately understanding language and appropriately generating language in a broad landscape of discourse worlds [12]. However, a combination of symbolic and statistical architectures have been quite successful in handling conversational interactions in some conversational contexts, including tutoring. Such systems sometimes help student learning and motivation even though the conversation is not perfect [4, 13]. The hope is that these successes will extend to AutoMentor.

2 Structure of Games and Mentor Contributions

A few words need to be said about the structure of our games and discourse, although it is beyond the scope of this paper to provide a full specification. The games consist of an ordered sequence of major activities (i.e., game phases) in service of a goal, with a group of players and a mentor in each activity. Each activity also has external media, controls, or products that players can view, manipulate, or create. An activity in one game may involve students being science journalists with the goal of creating an on-line science news magazine (the external product). An activity in another game might involve a group of players making decisions on how to change resources in a city to reduce pollution. Whatever the activity, the quality of interactions among players should be superior with the intervention of a mentor.

The mentor guides or enters the group discussion through conversational turns and each turn has one or more speech acts. Some speech acts fulfill politeness norms for

conversation, such as greetings ("hello"), introductions ("I'm your mentor"), and closing acts ("Let me leave you to work on this by yourselves"). Two important substantive speech acts of the mentor consist of *evaluations* ("The report is fine", "That is a good idea") and *suggestions* ("Make the graph more precise," "Improve the first sentence."). Evaluations consist of judgment about a person, product, or external referent. Suggestions are recommendations, directives, requests, or hints on what actions the students should perform. Table 1 shows some examples of these evaluation expressions. The discourse analysis system needs to segment and classify the contributions of the mentor into speech act categories. One major objective of the present study is to identify the different clusters of evaluations and of suggestions expressed by mentors.

Table 1. Example feedback and suggestion speech acts by human mentors.

Exemplar	Summary description	Template structure
EVALUATION		
Nice try in the originality of the lead.	The lead is good.	The lead is X.
Readers would want to get to know him better.	Readers want to know more about a person.	Readers want to know more about X.
SUGGESTION		
Strive to get more sizzle into that critical first sentence.	Make the first sentence better.	Make sentence X better.
Ideally, too, there'd be descriptions of action, from watching Bucky at the game.	You need to know what an ideal story is like in order to improve your writing.	Stories should ideally have X, so maybe you should have Y.

There is another important level of discourse course analysis called *epistemic frames* [1, 2, 3]. An epistemic frame is a set of norms, virtues, or criteria that guides the mentor's decisions and actions in the activity. The frame consists of a specific description of a way of talking, listening, writing, reading, acting, interacting, believing, valuing, and feeling (and using various objects, symbols, images, tools, and technologies). The particular frame elements are categorized into skills, knowledge, identity, values, and epistemology (what we call the SKIVE components), as will be elaborated shortly (see Table 2). The epistemic frame represents the vision of the mentor and hopefully the group of game players after some time while interacting with the mentor. It is indeed a deeper and more abstract level of discourse analysis than the particular speech acts. It is essential to have some alignment between the mentor's acts of evaluations or suggestions and the epistemic frame elements. A second major objective of the present study is to examine the alignments between mentor comments and elements of the SKIVE epistemic frame.

3 Corpora of Mentor Contributions in Journalism

We conducted some analyses on two corpora in order to investigate (a) the categories of mentor evaluations and suggestions and (b) the alignment between these comment categories and 18 SKIVE elements of an epistemic frame. A journalism practicum corpus consisted of a journalism professor providing copyedit comments on news stories submitted by student journalists over the course of a semester-long practicum, i.e., line-by-line reactions and suggestions to a news story. The *journalism practicum corpus* consisted of 426 comments. These comments were segregated into 443 evaluation segments and 620 suggestion segments by a graduate research assistant. The comments tended to be copious, detailed, and blunt. A second *game corpus* consisted of copyedit comments from *science.net,* a game designed for middle school students to work as science reporters and publish news stories in an online newspaper. Graduate students outside of the field of journalism were trained to give feedback in the game. These comments were less blunt, but they allegedly retained the salient features of the professional journalists. There were 1620 comments in the game corpus. A sample of comments in these corpora were analyzed on 18 SKIVE elements by two graduate students. They achieved a respectable interjudge reliability score (kappa =.76) when averaging over the 18 SKIVE elements.

Table 1 gives examples of some of the evaluation and suggestion segments. We segregated evaluations and suggestions, crossed with three descriptions: (a) an actual comment verbatim (exemplar), (b) our succinct summary statement that represents a cluster category of exemplars, as discussed later, and (c) a symbolic specification that could be used as a template to generate comments in AutoMentor when values are bound to elements or parameters within a particular game activity context.

Table 2. Elements of Epistemic Frames.

SKIVE element	Description
Skill: investigating	Ability to gather information for a story.
Skill: detail	Need to provide useful information, specific details, and facts in stories.
Knowledge: story	Terms concerning language used by journalists about their stories.
Knowledge: reporting	Terms about reporting, finding, gathering, and analyzing information for stories
Knowledge: Reader	Knowledge about what the reader wants and reader attributes.
Identity: Writer	Feedback on being a writer, including projective identity references that position journalist as a writer.
Value: Informing public	Informing the public on what they want to know about important issues in the community.
Value: Engaging reader	Maintaining the readers' attention by phrases that hook the reader.
Epistemology: Rich details	Guiding principle to tell stories by showing rather than telling and using details to bring the story alive.

Table 2 presents 9 SKIVE elements in an epistemic frame for an expert in journalism. These were based on ethnographic notes prepared by a graduate student

on the comments of the journalism professor. There were 18 SKIVE elements altogether, but the frequencies of some of the elements were too low to analyze. For the 9 elements in Table 2, the proportions of comments that possessed the element in the journalism practicum were .29, .34, .79, .36, .32, 10, .21, 22, and .25, respectively.

4 Analyses of Journalism Comments

Evaluation and Suggestion Categories. This analysis was conducted to discover the clusters of comments within the evaluation and suggestion speech acts. In essence, what were the fundamental clusters of comments expressed by the mentor in the journalism practicum corpus? We computed a similarity matrix between all possible pairs of the 443 evaluation segments using latent semantic analysis (LSA)[14]. In the same fashion we computed a similarity matrix on the 620 suggestion segments.

Each similarity score between segment A and B was computed as the geometric cosine between the two expressions via the dimensions in an LSA space. Our LSA space was based on the Touchstone Applied Sciences Associates (TASA) corpus of 37,651 documents with approximately 11 million words. This is a frequently used corpus to represent what typical high school students have read over their lifetimes. LSA is a useful method of computing similarity because it considers implicit knowledge in addition to the explicit words. LSA is a statistical technique for representing world knowledge, based on a large corpus of documents [14]. A single value decomposition technique is performed on the large document-by-word matrix (from the TASA corpus) that specifies the number occurrences of particular words in particular documents. It reduces the large sparse matrix to approximately 300 dimensions. The conceptual similarity between any two text excerpts is computed as the cosine between the values and weighted dimensions of the two text excerpts.

A principal components (PC) analysis was then conducted on the similarity matrix of the 443 evaluation segments. The top 20 components accounted for 79.2% of the variance in similarity scores. The evaluation segments are clustered according to the sorted loadings on the top 20 components with Varimax rotation. The same analysis was conducted on the similarity matrix of the 620 suggestion segments, with the top 30 components accounting for 85% of the variance, generating 30 clusters of suggestion segments. Table 1 shows exemplars of a few of these clusters.

The PC analysis is useful because we can induce what specific comment categories are relevant in a particular discourse context, in this case the journalism practicum course. This is indeed a useful discovery technique but hardly the end of the story in our analysis of games and mentors. There are at least three fundamental follow-up questions. First, what states of the game and interactions among students systematically trigger a particular category of evaluation or suggestion? Second, how are these evaluation and suggestion moves (illustrated in Table 1) aligned with the epistemic frame elements illustrated in Table 2? Third, how well can we generalize these categories form the journalism practicum corpus to both a similar corpus and to a different game corpus? It is beyond the scope of this paper to answer the first question, but we did conduct analyses relevant to the second and third questions.

Aligning Evaluation and Suggestion Moves to Epistemic Frame Elements. We conducted some binary logistic regression (BLR) analyses that attempted to predict each epistemic frame element from the principle component loadings of the 20 evaluation and 30 suggestion clusters. We focused on the 9 epistemic frame elements in Table 2 because they had a large enough frequency of occurrences.

Consider first the journalism practicum corpus. We randomly split the observations in half to prepare a training and test set. A BLR analysis was performed on the training set with component scores as predictors and the binary presence of a frame element as the criterion variable. The coefficients from this test set were applied to the test set to generate predictions about presence or absence of a frame element. The predictions from the BLR analysis (binary 0/1 values) were compared with the binary decisions of the human judges. A kappa score served as an index of the accuracy of the predictions; kappa adjusts for base rates and varies from 0 (chance) to 1 (perfect accuracy).

Table 3 shows the kappa scores for the journalism practicum corpus. The left four columns of numbers segregate training versus test sets for suggestion versus evaluation segments. The results support the claim that the component scores indeed can significantly predict the epistemic frame elements. The mean kappa scores were .76, .63, .54, and .50 for training-suggestions, training-evaluations, test-suggestions, and test-evaluations, respectively.

Table 3. Prediction (kappa) of Epistemic Frame Elements from Mentor Comments.

	Journalism Practicum				Game			
	Suggestion		Evaluation		Suggestion		Evaluation	
SKIVE element	Train	Test	Train	Test	Train	Test	Train	Test
Skill: investigating	.79	.57	.68	.53	.83	.44	.83	.56
Skill: detail	.66	.53	.58	.45	.53	.16	.33	.09
Knowledge: story	.55	.40	.58	.43	.59	.49	.57	.54
Knowledge: reporting	.68	.48	.58	.50	.58	.38	.52	.42
Knowledge: Reader	.91	.81	.88	.75	.89	.64	.84	.76
Identity: Writer	1.00	.36	.63	.35	.76	.56	.82	.62
Value: Informing public	.78	.60	.75	.69	.63	.51	.65	.38
Value: Engaging reader	.71	.54	.49	.44	.57	.40	.31	.12
Epistemology: Rich details	.74	.59	.50	.36	.64	.16	.38	.11

We next considered whether our PC categories in the journalism corpus can generalize to the game corpus. For the game corpus, 332 observations were coded on epistemic frame elements by graduate students. We used the coefficients derived from the journalism corpus to compute component scores and predict these coded elements in the game corpus. Once again, we randomly segregated training and test observations for suggestions and evaluations. In the training BLR we computed a new set of regression coefficients on the game sample; these new regression coefficients were applied to the test sample of the game corpus.

The right 4 columns of Table 3 support the claim that there is some transfer of the PC speech act clusters to the game corpus. The mean kappa scores were .67, .58, .42, and .40 for training-suggestions, training-evaluations, test-suggestions, and test-evaluations, respectively. Therefore, we could imagine a methodology in which

researchers (a) hand code epistemic frame elements of a modest sample of observations in the new discourse context, (b) compute LSA similarity scores on the new corpus (segregating evaluation and suggestion segments), (c) compute component score coefficients, (d) derive BLR formulas for the observations in a, and (e) compute the predicted elements for other segments in the new corpus.

It should be noted, however, that transfer is quite modest when we do not go through the above process that has humans hand code a sample of observations in the new corpus on epistemic frame elements. For example, we computed a BLR analysis on the 443 evaluation segments in the journalism corpus as a training set and used the same regression coefficients to predict the elements in the game corpus as a test set. The kappa scores for the training and transfer test set were .61 and .23, respectively. It appears that some hand coding in a new game activity is needed for these judgments of epistemic frame elements. However, it is an open question as to how much hand coding is needed. It is also conceivable that the amount of necessary hand coding will decrease substantially as we explore a greater number and diversity of game activities.

5 Discussion

This paper has identified the speech act categories and clusters of comments of journalism mentors who interact with students editing news stories. The basic speech act categories are evaluations and suggestions, in addition to the normal greetings, introductions, and closing moves that speech participants normally perform in multiparty conversation. LSA and principal components analyses helped us discover the different types of evaluations and suggestions. These comments of mentors were also significantly aligned with epistemic frame elements (see Table 2). Such alignments were discovered by logistic regression and validated on a sample of hand-coded judgments of the frame elements. There was some modest transfer from a journalism practicum corpus to a game corpus, but more research is needed to explore how the transfer can improve.

These analyses provide an important first step in building a virtual AutoMentor for multiparty epistemic games. We now have a sketch of AutoMentor's speech act categories, comments, and the epistemic functions that they serve. The next step is to formulate algorithms that generate particular speech acts and discourse moves in a fashion that is sensitive to the game states and interactions among students. The generation algorithms need to be sensitive to top-down goals (i.e., what epistemic frame elements are operating in the current game activity) and bottom-up constraints (e.g., conflicts between students, progress in the game activity). Moreover, the generation templates in column 3 of Table 1 provide one approach to binding a discourse move to the setting and game parameters at particular points in time. The success of these approaches await the future development and testing of AutoMentor.

Acknowledgments. This research was supported by the National Science Foundation (BCS 0904909, DRK-12-0918409) and the Institute of Education Sciences (R305B070349, R305A080594). The opinions, findings, and conclusions do not

reflect the views of the funding agencies, cooperating institutions, or other individuals.

References

1. Shaffer, D.W.: How Computer Games Help Children Learn. Palgrave, New York (2007)
2. Hatfield, D., Shaffer, D.W.: The Epistemology of Journalism 335: Complexity in developing journalistic experience. In: ICLS, Chicago, IL (2010)
3. Bagley, E., Shaffer, D.W.: When People Get in the Way: Promoting Civic Thinking Through Epistemic Game Play. International Journal of Gaming and Computer-Mediated Simulations. 1, 36--52 (2009)
4. Graesser, A.C., Lu, S., Jackson, T.T., Mitchell, H., Ventura, M., Olney, A., Louwerse, M.M.: AutoTutor: A Tutor with Dialogue in Natural Language. Behavior Research Methods, Instruments, and Computers. 36, 180--193 (2004)
5. Graesser, A.C., Jeon, M., Dufty, D.: Agent Technologies Designed to Facilitate Interactive Knowledge Construction. Discourse Processes, 45, 298--322 (2008)
6. D'Mello, S., King, B., Chipman, P., Graesser, A.C.: Towards Spoken Human-Computer Tutorial Dialogues. Human Computer Interaction, In Press
7. Millis, K., Cai, Z., Graesser, A., Halpern, D., Wallace, P.: Learning Scientific Inquiry By Asking Questions in an Educational Game. In: Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2009, pp. 2951--2956. AACE, Chesapeake, VA (2009)
8. Jackson, G.T., Olney, A., Graesser, A.C., Kim, H.T.: AutoTutor 3-D Simulations: Analyzing User's Actions and Learning Trends. In: Proceedings of the 28th Annual Meeting of the Cognitive Science Society, pp. 1557--1562. ErlBaum, Mahway, NJ (2006)
9. Clark, H.H.: Using Language. Cambridge University Press, Cambridge (1996)
10. Gee, J.P.: An Introduction to Discourse Analysis: Theory and Method. Routledge, London (1999)
11. Graesser, A.C., McNamara, D.S.: Computational Analyses of Multilevel Discourse Comprehension. Topics in Cognitive Science, In Press
12. Jurafsky, D., Martin, J.H.: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice-Hall, Upper Saddle River, NJ (2008)
13. VanLehn, K., Graesser, A.C., Jackson, G.T., Jordan, P., Olney, A., Rose, C.P.: When Are Tutorial Dialogues More Effective Than Reading?. Cognitive Science, 31, 3--62 (2007)
14. Landauer, T., McNamara, D.S., Dennis, S., Kintsch, W. (eds.): Handbook of Latent Semantic Analysis. Erlbaum, Mahwah, NJ (2007)