

Typed versus Spoken Conversations in a Multi-Party Epistemic Game

Brent Morgan¹, Candice Burkett¹, Elizabeth Bagley²,
and Arthur Graesser¹

¹ University of Memphis, Psychology, Institute for Intelligent Systems,
365 Innovation Drive, Memphis, TN 38152, USA

² University of Wisconsin-Madison, Educational Psychology, Educational Sciences
Building, Room 1078D, 1025 West Johnson Street, Madison, WI 53711

Abstract. Multi-party chat is a standard feature of popular online games and is increasingly available in collaborative learning environments, including epistemic games. However, little is known about the linguistic qualities of group conversation, especially regarding modality of communication. This paper addresses the differences between spoken and typed conversations as high school students interacted with the epistemic game Urban Science. Coh-Metrix analyses showed that speech was associated with more global aspects of text (narrativity, cohesion) whereas typed input was associated with more local aspects (syntactic simplicity, word concreteness). The spoken conversations were also more verbose than typed conversations. These findings suggest that the modality in group communication should be considered, particularly with respect to the breaks in discourse cohesion that are more prominent in typed group chat.

Keywords: computational linguistics, distance learning, epistemic games, natural language processing

1 Introduction

Multi-party collaborative learning environments are becoming increasingly popular in a wide range of formal and informal applications [1]. The prototypical formal application involves collaborative problem-solving in a problem-based curriculum on science or technology [1, 2], while the prototypical informal application is a multi-party serious game [3, 4]. Although online chat is a relatively new area of research in education, chat has long been the standard for group communication in online video games (e.g., Ultima, or more recently, World of Warcraft). With the increased attention to serious games (which frequently use recreational games as a model), multi-party chat will likely become a critical focus of Natural Language Processing (NLP) research in the coming years.

Within these group conversations, both the typed chat medium and oral face-to-face interactions offer certain affordances and constraints, so designers question

whether to have the exchanges face-to-face or virtual in multi-party applications. For example, in distance learning, the collaborative learning or round-table discussion format of traditional classrooms is replaced with online chat, wherein the instructor and students interact by sending typed messages that the entire group or a selected subgroup can view. What are the consequences on language and discourse when the group conversations move from face-to-face to chats? Since the linguistic differences and implications between speaking and typing in a multi-party setting have not been adequately studied in educational contexts, this study focuses on the comparative impact of the two communication media on language and discourse.

There is a large body of research on differences between oral and written communication [5, 6, 7]. The early research in linguistics and discourse processing required researchers to annotate the discourse samples by hand. However, advances in computational linguistics, corpus linguistics, and natural language understanding in AI [8] have given us the opportunity to automate many of the levels of language and discourse. An early automated corpus analysis by Biber (1988) examined a sample of oral versus written language. Biber assembled a corpus of 481 spoken and written texts (including conversations, speeches, letters, fiction, etc.) and analyzed them across seven dimensions identified by a factor analysis. The results showed no single predominant difference between oral spoken and written texts, but there were differences in the factors between various text categories (genre). For example, the differences between narrative and non-narrative language were similar in both spoken and written language.

More recently, Louwse, McCarthy, McNamara, & Graesser [9] analyzed the Biber corpus using a system called Coh-Metrix [10, 11, 12]. Coh-Metrix automatically analyzes texts on multiple levels of language and discourse (as will be discussed in section 1.2). Whereas Biber's analysis was exclusively focused at the word level, Louwse et al. analyzed syntax, discourse cohesion, and other levels of language and discourse. Their analysis revealed that differences between speech and writing accounted for the majority of the variability among the Biber texts. These analyses are important first steps, but the discourse segments in the sample did not include the interactive dialogues of interest in the present study.

Within interactive dialogues, researchers have compared spoken and chat interactions in one-on-one tutoring sessions. For example, Van Lehn, Graesser, Jackson, Jordan, Olney, & Rose [13] reported slightly higher student learning gains for spoken as opposed to typed human-to-human tutoring sessions. Litman et al. (2006) [14] reported a similar finding for both a human tutor and their ITSpoke intelligent tutoring system (ITS), although the effect was diminished for the computer tutor. Whereas Van Lehn et al. and Litman et al. varied the communication mode of both the student and the tutor together, D'Mello, Dowell, & Graesser [15] only varied whether the students spoke or typed as they interacted with AutoTutor, a system with an animated conversational agent that helps students learn by holding a conversation in natural language [13, 16]. D'Mello et al. [15] found no differences in learning gains across two experiments. Additionally, differences in the language and discourse of spoken versus typed interactions with AutoTutor were reported by Graesser [17].

Although the findings regarding spoken versus typed student communication in a one-on-one tutoring session are extremely informative, it remains to be seen how they translate to a group setting. Multi-party chat constitutes a new frontier in NLP, with a

host of new challenges. One important step is to understand how humans adapt their conversational style to accommodate the affordances and constraints of a chat room. Accordingly, this study compared speech and chat in the context of the epistemic game *Urban Science* [18].

1.1 Urban Science

Urban Science is an epistemic game created by education researchers at the University of Wisconsin-Madison [19], designed to simulate an urban planning practicum experience [20]. Young people role-play as professional urban planners in an ecologically-rich neighborhood to develop new ways of observing and acting in the world they inhabit. Students are assigned to one of three planning teams, each of which represents a stakeholder group (e.g., People for Greenspace). Students conduct a virtual site visit to learn about the issues their Non-Player Character (NPC) stakeholders care about. The students ultimately submit and defend a new plan for the city that aims to meet the needs of the community. During the game, players communicate with other members of their planning team, as well as with an adult mentor role-playing as a professional planning consultant. These conversations between the mentor and students were analyzed on various measures of Coh-Metrix.

1.2 Coh-Metrix

Coh-Metrix is a computational linguistic tool that measures text cohesion and text difficulty on a range of word, sentence, paragraph, and discourse dimensions. Coh-Metrix was designed for analyzing meaning at multiple levels of language and discourse beyond the word level, as opposed to letters, phonemes, and units within words.

Recently, a principal components analysis (PCA) reduced 53 Coh-Metrix measures to five major dimensions of text: narrativity, referential cohesion, situation model cohesion, syntactic simplicity, and word concreteness [11]. Using these deep-level components in conjunction with more superficial aspects of the text (i.e., total number of sentences, total number of words, and words per sentence), we aim to better understand whether and how oral and typed communication differ in *Urban Science* [6, 17, 21].

We hypothesize that narrativity should favor the spoken condition because it is associated with familiar (easy to produce) words and oral language. Additionally, the two cohesion measures, referential cohesion and situation model cohesion, should also favor the spoken condition because the prevalence of multiple conversational threads in chat run the risk of creating cohesion breaks between contributions when the contributions are not delivered fast enough. Finally, since students in the typed condition are able to edit their contributions before sharing them, we hypothesize there will be greater syntactic simplicity and more concrete or specific selection to content words to promote grounding and minimize confusion.

These five components can be further divided at a conceptual level. Narrativity, referential cohesion, and situation model cohesion are concerned with connections

between ideas and thus represent *global* aspects of communication. In contrast, syntactic simplicity and word concreteness focus on the *local* characteristics of individual words and sentences. The typed condition should be superior for individual contributions because of ability to edit, but spoken should be superior across contributions because the faster response time will more directly connect to topics of conversation.

The superficial aspects of the texts include the number of sentences, number of words, and words per sentence. The first two are production volume indices which should favor the spoken condition since speech is easier to produce. The number of words per sentence, however, might favor either condition because while speech is easier to produce, typing allows more time to compose well-formed sentences.

2 Method

Students played the epistemic game, Urban Science, which enabled them to complete an urban planning internship for a fictitious urban planning firm. During the game, participants worked in teams and interacted with mentors who were trained in the urban planning profession, the game's activities, and preferred mentoring strategies. The players' primary task was to redesign the Northside neighborhood in Madison, WI.

2.1 Participants

This study analyzed data from 21 high school-aged participants and 2 mentors who played Urban Science for 10 hours over 3 days as a part of a week-long Conservation Leadership Program. Participants had no prior experience in Urban Planning and were recruited by outreach specialists at the Massachusetts Audubon Society's Drumlin Farm Wildlife Sanctuary.

2.2 Procedure

Upon arrival, students were randomly assigned to one of two mentoring conditions: typed or spoken. In the typed condition, players and mentors interacted entirely through an internal chat program within the game interface. In the spoken condition, players and mentors communicated orally, and all interactions were audio recorded. Human transcribers converted conversations in the spoken condition to text for analysis. All other aspects of the game remained constant across the conditions. Participants completed three phases of the game: Introduction, Stakeholder, and Final Plan. The three phases were subdivided into stages, with each stage requiring different tasks, skills, and goals.

The Introductory phase consisted of two stages. The first stage introduced the participants to the game, and in the second stage each participant was assigned to act as liaisons for one of three fictitious stakeholder groups concerned with the development of the Northside neighborhood: Madison Developers' Consortium,

Northside Neighbors, and People for Greenspace. Each stakeholder group consisted of three students and one mentor.

The Stakeholder phase consisted of ten stages, wherein participants conducted a virtual site visit of the Northside neighborhood and collected information from stakeholders in the community (NPCs) to learn about the types of issues they cared about. Players then proposed land changes based on their stakeholder input using *iPlan*, an interactive GIS model of the planning site that let them assess the ramifications of those changes. Players submitted their proposals to their stakeholder group, and received their feedback via email.

The Final Plan phase consisted of five stages and began with participants forming new planning teams (with one player from each of the previous stakeholder teams). Following a similar process as in the Stakeholder Phase, each player ultimately submitted and justified a final plan that would incorporate the needs of all of the stakeholder groups.

3 Results and Discussion

Our analyses focused on identifying the linguistic differences between typed and spoken language during student and mentor conversations in Urban Science. As noted in the Methods section, since the 6 planning teams were regrouped for the Final Plan phase, to preserve continuity, the present Coh-Metrix analyses focus only on 13 stages in the Introduction and Stakeholder phases. The unit of analysis was total group conversation that occurred in a particular stage by a particular planning team. There were 13 stages of focus and 6 teams, with half being spoken and half typed.

The dependent variables for this analysis contained 5 deep-level dimensions (narrativity, referential cohesion, situation model cohesion, syntax simplicity, and word concreteness) and 3 superficial aspects (number of words, number of sentences, and number of words per sentence) of the texts. For each dependent variable, a mixed analysis of variance was conducted on the z-scores (deep-level) or numerical counts (superficial). The independent variables included one between-subjects variable (typed vs. spoken) and one within-subject variable (stage, with 13 levels), with 6 texts for each stage. Presently, we have no theoretical concern for the differences between stages, so only the main effects between typed and spoken will be presented. The results are displayed in Table 1.

For the deep-level dimensions, we predicted that the global aspects of the text (narrativity, referential cohesion, and situation model cohesion) would favor the spoken condition, and the results supported this hypothesis. Conversations in the spoken conditions were significantly higher in narrativity and situation model cohesion and marginally significantly higher in referential cohesion. The predictions for the local aspects of the text were also supported, with typed conversations rating significantly higher on word concreteness and marginally higher on syntactic ease.

The superficial aspects of the text contained 2 aggregate measures (number of words and number of sentences) of each text and 1 rate measure (words per sentence). We predicted that the spoken condition would have higher values for the aggregate measures than the typed conditions, and the results confirmed the hypotheses.

However, we were uncertain as to how the rate measure would be affected by speaking versus typing. In fact, the spoken condition contained a significantly higher rate of words per sentence than the typed condition.

Table 1. Coh-Matrix Analysis of Stages 1-13.

	Typed		Spoken		<i>F</i>	<i>p</i>	η^2
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Narrativity	0.94	0.33	1.46	0.21	28.17	0.01	0.88*
Referential Cohesion	-0.62	0.78	-0.09	0.49	6.24	0.07	0.61
Situation Model Cohesion	-0.26	0.73	1.03	0.84	60.97	0.00	0.94*
Syntactic Simplicity	0.57	0.44	0.12	0.54	4.93	0.09	0.55
Word Concreteness	-1.76	0.82	-2.43	0.53	63.53	0.00	0.94*
Total Number of Words	342.31	173.20	1,001.33	659.70	13.80	0.02	0.78*
Total Number of Sentences	32.05	18.15	71.97	49.00	8.81	0.04	0.69*
Words per Sentence	11.38	1.81	14.11	4.08	10.93	0.03	0.73*

* $p < .05$

We analyzed the interaction between stage and the typed vs. spoken variables. Each interaction was either statistically significant at $p < .05$ (narrativity, referential cohesion, situation model cohesion, word concreteness, total number of words, words per sentence) or at $p < .10$ (syntactic simplicity, total number of sentences). A Wilcoxon Sign Test on each of the dependent measures indicated that the means were significantly in the same direction across the 13 stages; all sign tests were significant at $p < .05$ except for one being $p < .10$. Therefore, variations in the magnitude of the differences rather than the direction of the differences accounted for the significant interactions between stage and communication medium. Overall, our results indicate a preference for speech with regards to global aspects of text, whereas typed is superior with regards to local aspects. The spoken condition was also much more verbose than the typed condition with respect to all the superficial aspects of the text. Additionally, while there were interactions between typed versus spoken and

individual stages, these were mostly due to magnitude variations and not a reversal in the means.

4 General Discussion

Our Coh-Matrix analyses examined distinctive characteristics of typed and spoken language in group interactions as high school students role-played as urban planners in an epistemic game. Our analyses revealed that conversations in the spoken condition were more narrative and cohesive representing connectivity across contributions (global), whereas typed conversations contained more concrete words and simpler syntax (local). The superficial aspects of the texts indicated that the players were more verbose in the spoken condition.

This paper provides a preliminary insight into the differences between speaking and typing in a multi-party setting. Furthermore, as distance learning and epistemic games scale up in order to be accessible to more students, there will be a greater reliance on NLP to process student contributions. These results show that the implementation of the AI needs to take the medium of communication into consideration. In particular, if multiple conversational threads in group communication are creating breaks or discontinuities in discourse cohesion, then the interface and discourse management facilities must find ways to connect the content of conversational turns to the appropriate points in conversations in the electronic media.

Finally, as NLP conquers group discussion, numerous educational possibilities will emerge. For example, one can imagine a virtual environment in which NPCs allow students to explore individually or engage in peer tutoring, but can engage in the discussion when guidance is needed. To the extent that distance learning and epistemic games continue to emulate online recreational games, they will become more collaborative and focus on group conversation, making this a focus of research in the years to come.

Acknowledgments. This work was funded in part by the MacArthur Foundation, the National Science Foundation (REC-0347000, DUE-091934, DRL-0918409, DRL-0946372, BCS 0904909, and DRK-12-0918409), the Institute of Education Sciences (R305G020018, R305A080589), The Gates Foundation, and U.S. Department of Homeland Security (Z934002/UTAA08-063). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of these funding agencies, cooperating institutions, or other individuals. We also thank Zhiqiang Cai for the Coh-Matrix software development.

References

1. Dede, C.: Immersive interfaces for engagement and learning. *Science*. 323, 66--69 (2009)

2. Kali, Y.: Collaborative knowledge-building using the Design Principles Database. *International Journal of Computer Support for Collaborative Learning*. 1, 187--201 (2006)
3. Dieterle, E., Clarke, J.: Multi-user virtual environments for teaching and learning. In: Pagani, M. (ed.), *Encyclopedia of multimedia technology and networking* (2nd ed). Idea Group, Hershey (in press)
4. Ketelhut, D., Dede, C., Clarke, J., Nelson, B., Bowman, C.: Studying situated learning in a multi-user virtual environment. In: Baker, E., Dickieson, J., Wulfeck, W., & O'Neil, H. (eds.), *Assessment of problem solving using simulations*, pp. 37--58. Earlbaum, Mahweh (2007)
5. Biber, D.: *Variation across speech and writing*. Cambridge University Press, Cambridge (1988)
6. Clark, H.: *Using Language*. Cambridge University Press, Cambridge (1996)
7. Tannen, D.: Oral and literate strategies in spoken and written narratives. *Language*. 58, 1--21 (1982)
8. Jurafsky, D., Martin, J.: *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Prentice-Hall, Upper Saddle River (2008)
9. Louwerse, M.M., McCarthy, P.M., McNamara, D.S., Graesser, A.C.: Variation in language and cohesion across written and spoken registers. In: Forbus, K., Gentner, D., Regier, T. (eds.), *Proceedings of the twenty-sixth annual conference of the Cognitive Science Society*, pp. 843--848. Erlbaum, Mahweh. (2004)
10. Graesser, A.C., McNamara, D.S., Louwerse, M.M., Cai, Z. Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers*. 36, 193--202 (2004)
11. Graesser, A.C., McNamara, D.S. Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science* (in press)
12. McNamara, D. S., Graesser, A., Louwerse, M. Sources of text difficulty: Across the ages and genres. To appear in: Sabatini, J. P., Albro, E. (eds.), *Assessing reading in the 21st century: Aligning and applying advances in the reading and measurement sciences* (in press)
13. Van Lehn, K., Graesser, A., Jackson, G., Jordan, P., Olney, A., Rose, C. P. When are tutorial dialogues more effective than reading? *Cognitive Science*. 31, 3-62 (2007)
14. Litman, D., Rose, C., Forbes-Riley, K., VanLehn, K., Bhembe, D., Silliman, S. Spoken versus typed human and computer dialogue tutoring. *International Journal of Artificial Intelligence In Education*. 16, 145--170 (2006)
15. D'Mello, S. K., Dowell, N., Graesser, A.C. Does it really matter whether students' contributions are spoken versus typed in an Intelligent Tutoring System with natural language? *Journal of Experimental Psychology: Applied* (in press)
16. Graesser, A. C., Jeon, M., Dufty, D. Agent technologies designed to facilitate interactive knowledge construction. *Discourse Processes*. 45, 298--322 (2008)
17. Graesser, A. C., Jeon, M., Yang, Y., Cai, Z. Discourse cohesion in text and tutorial dialogue. *Information Design Journal*. 15, 199--213 (2007)
18. Shaffer, D. W. (2007). *How computer games help children learn*. New York: Palgrave.
19. Bagley, E. S., 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)
20. Bagley, E. Epistemography of an urban and regional planning practicum: Appropriation in the face of resistance. (WCER Working Paper 2010-8). Madison: University of Wisconsin-Madison, Wisconsin Center for Education Research (2010)
21. Clark, H. H, Brennan, S. E. Grounding in communication. In: Resnick, L. B., Levine, J. M, Teasley, S.D. (eds.), *Perspectives on Socially Shared Cognition*, American Psychological Association, pp. 127--149 (1991)