

# Modeling Learners' Cognitive, Affective, and Social Processes through Language and Discourse

Nia M. M. Dowell and Arthur C. Graesser

University of Memphis and Institute for Intelligent Systems, United States

[ndowell@memphis.edu](mailto:ndowell@memphis.edu)

An emerging trend toward computer-mediated collaborative learning environments promotes lively exchanges between learners in order to facilitate learning. Discourse can play an important role in enhancing epistemology, pedagogy, and assessments in these environments. In this paper, we highlight some of our recent work showing the advantages using theoretically grounded automated linguistics tools to identify pedagogically valuable discourse features that can be applied in collaborative learning, intelligent tutoring systems (ITS), computer-mediated collaborative learning (CMCL), and MOOC environments.

**Keywords:** Coh-Metrix, learning analytics, computer-mediated communication, online learning, educational data mining

## 1. INTRODUCTION

Current educational practices suggest an emerging trend toward computer-mediated collaborative learning environments and groupware tools, such as email, chat, threaded discussion, massive open online courses (MOOCs), and dialog-based intelligent tutoring systems (ITSs). This has stimulated recent discussion among educational data mining and learning analytics researchers about how best to model learners' cognitive, motivational, affective, and social processes and incorporate pedagogically beneficial, adaptive strategies into these environments. In this paper, we highlight some of the recent work showing the advantages of using theoretically grounded automated linguistics tools to identify pedagogically valuable discourse features that can be applied in collaborative learning, ITS, CMCL, and MOOC environments.

### 1.1 Theoretical Framework

Collaborative language is the factor that sets CMCL learning apart from individual learning (Dowell, Cade, Tausczik, Pennebaker, & Graesser, 2014). In this context, language, discourse and communication play a critical and complex role that can provide insight regarding social processes (i.e., establishing a common ground and vision), individual and group cognitive processes (i.e., knowledge construction), and affective processes (i.e., confusion, frustration, boredom, flow/engagement). Psychological frameworks of comprehension and learning have identified the representations, structures, strategies, and processes at multiple levels of discourse (Graesser & McNamara, 2011; Kintsch, 1998; Snow, 2002). Five levels have frequently been proposed in these frameworks: 1) *words*, 2) *syntax*, 3) the explicit *textbase*, 4) the *situation model* (sometimes called the mental model), and 5) the *discourse genre and rhetorical structure* (the type of discourse and its composition). In the educational context, learners can experience communication misalignments and comprehension breakdowns at different levels. These breakdowns and misalignments have important implications for cognitive processing.

## 2. METHODS

### 2.1 Computational Linguistic Analysis Tool

Coh-Matrix is a computational linguistics facility that analyzes higher-level features of language and discourse (Graesser, McNamara, & Kulikowich, 2011; McNamara, Graesser, McCarthy, & Cai, 2014). Coh-Matrix includes sophisticated methods of natural language processing, providing over 100 measures at multiple levels, including genre, cohesion, syntax, words, as well as other characteristics of language and discourse. Coh-Matrix also offers measures of linguistic complexity and formality, characteristics of words, and readability scores. There was a need to reduce the large number of measures provided by Coh-Matrix into a more manageable number of measures. This was achieved in a study that examined 53 Coh-Matrix measures for 37,520 texts in the TASA (Touchstone Applied Science Association) corpus, which represents what typical high school students have read throughout their lives (Graesser et al., 2011). A principal components analysis was conducted on the corpus, yielding eight components that explained an impressive 67.3% of the variability among texts; the top five components explained over 50% of the variance. Importantly, the components aligned with the discourse levels previously proposed in multilevel theoretical frameworks of cognition and comprehension (Graesser & McNamara, 2011; Kintsch, 1998; Snow, 2002) and thus are ideal for investigating trends in learning-oriented interactions. The five major dimensions are succinctly defined below, starting with the most global level (genre):

- **Narrativity:** The extent to which the text is in the narrative genre, which conveys a story, a procedure, or a sequence of episodes of actions and events with animate beings. Informational texts on unfamiliar topics are at the opposite end of the continuum.
- **Deep Cohesion:** The extent to which the ideas in the text are cohesively connected at a deeper conceptual level that signifies causality or intentionality.
- **Referential Cohesion:** The extent to which explicit words and ideas in the text are connected with each other as the text unfolds.
- **Syntactic Simplicity:** Sentences with few words and simple, familiar syntactic structures. At the opposite pole are structurally embedded sentences that require the reader to hold many words and ideas in working memory.
- **Word Concreteness:** The extent to which content words are concrete, meaningful, and evoke mental images as opposed to abstract words.

## 3. RECENT FINDINGS

Our recent work used Coh-Matrix to explore cognitive, affective, social, and socio-affective processes during collaborative learning, ITS, and MOOC interactions (Cade, Dowell, Graesser, Tausczik, & Pennebaker, 2014; D’Mello, Dowell, & Graesser, 2009; Dowell et al., 2014; Joksimović et al., under review). For instance, Dowell et al. (2014) explored the possibility of using discourse features to predict student and group performance during collaborative learning interactions. We investigated the linguistic patterns of group chats, within an online collaborative learning exercise, on five discourse dimensions using Coh-Matrix. Our results indicated that students who engaged in deeper cohesive integration and generated more complicated syntactic structures performed significantly better. The overall group level results indicated collaborative groups who engaged in deeper cohesive and expository style interactions performed significantly better on post-tests. In line with this, Cade et al. (2014) demonstrated that cognitive linguistic cues can be used in detecting students’ socio-affective attitudes towards fellow

students in CMCL environments, which may have long-term consequences for their motivation and continued use of such systems. Our current and future research focuses on exploring how these strategies transfer to increasingly larger, more culturally diverse populations of learners and extending our conclusions to practical applications that enhance learning and teaching.

#### 4. CONTRIBUTION TO LEARNING ANALYTICS

These results suggest that students' latent cognitive, affective, and social processes can be monitored by analyzing language and discourse. An interdisciplinary approach that combines psychological theories of discourse comprehension with computational linguistics methodologies holds the potential for enabling substantially improved learning environments by providing real-time detection of students and group performance and by using this information to develop student models and provide adaptive learning supports.

#### ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation under Grant BCS 0904909, DRK-12-0918409; the Air Force Office of Scientific Research under the Grant Minerva Initiative, 14RT1214; and the U.S. Department of Homeland Security under Grant Z934002/UTAA08-063.

#### REFERENCES

- Cade, W. L., Dowell, N. M., Graesser, A. C., Tausczik, Y. R., & Pennebaker, J. W. (2014). Modeling student socioaffective responses to group interactions in a collaborative online chat environment. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining*. (pp. 399–400). Berlin: Springer.
- D'Mello, S., Dowell, N., & Graesser, A. C. (2009). Cohesion relationships in tutorial dialogue as predictors of affective states. *Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling* (pp. 9–16). Amsterdam: IOS Press.
- Dowell, N. M., Cade, W. L., Tausczik, Y. R., Pennebaker, J. W., & Graesser, A. C. (2014). What works: Creating adaptive and intelligent systems for collaborative learning support. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *12th International Conference on Intelligent Tutoring Systems*. (pp. 124–133). Berlin: Springer.
- Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, 3(2), 371–398.
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223–234.
- Joksimović, S., Dowell, N. M., Oleksandra, S., Kovanović, V., Gašević, D., Dawson, S., & Graesser, A. C. (under review). How do you connect? Analysis of social capital accumulation in connectivist MOOCs.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, MA: Cambridge University Press.
- Snow, C. E. (2002). *Reading for understanding: Toward a research and development program in reading*

Do not touch this during review process. (xxxx). Paper title here. *Journal of Learning Analytics*, xx (x), xx–xx.

*comprehension*. Santa Monica, CA: Rand Corporation.