

Chapter 13

Statistical Discourse Analysis: Testing Educational Hypotheses with Large Datasets of Electronic Discourse

Ming Ming Chiu

University at Buffalo, State University of New York, USA

Gaowei Chen

University of Pittsburgh, USA

ABSTRACT

Educators are increasingly using electronic discourse for student learning and problem solving, partially due to its time and space flexibility and greater opportunities for information processing and higher order thinking. When researchers try to statistically analyze the relationships among electronic discourse messages however, they often face difficulties regarding the data (missing data, many codes, non-linear trees of messages), dependent variables (topic differences, time differences, discrete, infrequent, multiple dependent variables) and explanatory variables (sequences of messages, cross-level moderation, indirect effects, false positives). Statistical discourse analysis (SDA) addresses all of these difficulties as shown in analyses of social cues in 894 messages posted by 183 students during 60 online asynchronous discussions. The results showed that disagreements increased negative social cues, supporting the hypothesis that these participants did not save face during disagreements, but attacked face. Using these types of analyses and results, researchers can inform designs and uses of electronic discourse.

INTRODUCTION

Educators are increasingly using electronic discussions (e.g., discussion boards, knowledge building environments, blogs) to have students solve problems and to help them learn. These electronic discussions offer students greater flexibility to communicate at different times and in different

locations. Furthermore, the explicit and permanent displays of writing (and drawing) and longer time for reflection offer students greater opportunities for information processing and higher order thinking (Hara, Bonk, & Angeli, 2000; Tallent-Runnels, Thomas, Lan, Cooper, Ahern, & Shaw et al., 2006). As a result, students are not only creating voluminous amounts of electronic discourse data, but they are doing so at a faster rate.

DOI: 10.4018/978-1-4666-4426-7.ch013

While many studies have counted attributes of separate electronic discourse messages (e.g., Garrison, Anderson, & Archer, 2001; Gress, Fior, Hadwin, & Winne, 2008; Hara et al., 2000; Raffaelli & Sudweeks, 1998; Tallent-Runnels et al., 2006), they have not examined the relationships within sequences of messages. Within threads of messages, participants can respond to earlier actions and/or invite future actions. By analyzing these relationships across messages, researchers can improve their understanding of the social and cognitive dynamics of electronic discourses and help educators improve students' learning and discussions.

Suitable statistics methods can help researchers test hypotheses about these relationships on large data sets to build an empirically-supported knowledge base. However, statistically analyzing online interaction processes often requires addressing difficulties regarding the data set, outcome variables, and explanatory variables. Data issues include missing data, content analysis/coding, and non-linear trees of messages.

Difficulties involving dependent variables include differences across online discussion topics, similarities in adjacent messages, discrete (not continuous) variables, infrequent occurrences of focal events, and multiple dependent variables. Explanatory variable issues include sequences of messages, context-dependent effects, indirect effects across levels, and false positives.

In this chapter, we introduce how statistical discourse analysis (SDA; Chiu, 2008a, Chiu & Khoo, 2005) addresses the above analytic difficulties. To contextualize the methodological issues, we test whether earlier messages affect the likelihood of specific attributes in each message. For example, people often express personal affect or attitudes towards others during a discussion, a *social cue* ("Oh, I get it now"; "☺"). Hence, we can test which attributes of earlier messages affect the likelihood of a positive cue in the current message. In this study, we do so by applying SDA to 894 messages by 183 participants on 60

high school mathematics topics on a mathematics problem solving website. The analyses and results can help us build empirically-supported understanding to inform and improve educators' designs and uses of electronic discourse.

BACKGROUND

During an electronic discussion, a group of participants exchange ideas by sending and receiving messages. Like face-to-face discourse, electronic discourse involves individuals responding to others' recent messages and inviting others' future messages. When participants agree or disagree with the content of the previous message, they refer back into the past; in contrast, asking questions or issuing commands projects forward into future messages. These links connect series of sequential events (or time-series data). Moreover, if there are multiple discussion topics/groups, the electronic discourse also has multilevel structure (messages nested within topics).

Unlike face-to-face discourse, electronic discourse can have messages along many threads. Due to a weaker turn-taking mechanism, electronic discussants can respond to any previous message, especially during asynchronous discussions. As such, electronic discussions often proceed along multiple threads and generate non-linear trees of messages. While some of the messages receive many responses, others receive none (which would end the specific discussion branch; Hewitt, 2005; Thomas, 2002). See Figure 1 for an example of the tree relationships between a topic and its 13 responses. The number "0" denotes the initial problem; "1" through "13" indicates 13 chronological reply messages, where "1" refers to the earliest reply and "13" refers to the last reply. In this tree of messages, the topic and its replies were linked to one another by multiple threads and single connections. Messages in each thread were ordered by time, but they were not necessarily consecutive. For example, message #5 followed

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/statistical-discourse-analysis/76995

Related Content

Translation of Wine as a Culture-Bound Term From English Canon Text to a Language of Lesser Diffusion

Joseph Igono (2022). *International Journal of Translation, Interpretation, and Applied Linguistics* (pp. 1-12). www.irma-international.org/article/translation-of-wine-as-a-culture-bound-term-from-english-canon-text-to-a-language-of-lesser-diffusion/304078

Attention-Sharing Initiative of Multimodal Processing in Simultaneous Interpreting

Tianyun Liand Bicheng Fan (2020). *International Journal of Translation, Interpretation, and Applied Linguistics* (pp. 42-53). www.irma-international.org/article/attention-sharing-initiative-of-multimodal-processing-in-simultaneous-interpreting/257029

Morpho-Syntactic Marking of Inflectional Categories in English

Gulsat Aygen (2019). *Applied Linguistics for Teachers of Culturally and Linguistically Diverse Learners* (pp. 128-178). www.irma-international.org/chapter/morpho-syntactic-marking-of-inflectional-categories-in-english/236248

International Family Configurations in Tokyo and their Cross-Cultural Approaches to Language Socialization

Donna M. Velliaris (2015). *Handbook of Research on Cross-Cultural Approaches to Language and Literacy Development* (pp. 57-86). www.irma-international.org/chapter/international-family-configurations-in-tokyo-and-their-cross-cultural-approaches-to-language-socialization/134736

Second or Foreign Language Learning With Augmented Reality

Aubrey Stattiand Kelly M. Torres (2020). *Language Learning and Literacy: Breakthroughs in Research and Practice* (pp. 933-954). www.irma-international.org/chapter/second-or-foreign-language-learning-with-augmented-reality/233137