December 2001

Syntactic and Semantic Understanding of Conceptual Data Models

Cheryl Dunn
Florida State University

Severin Grabski
Michigan State University

SYNTACTIC AND SEMANTIC UNDERSTANDING OF CONCEPTUAL DATA MODELS

Cheryl L. Dunn
Florida State University
cheryl.dunn@fsu.edu

Severin V. Grabski
Michigan State University
grabski@msu.edu

Extended Abstract[1]

Conceptual data models are used for discovery and validation communication between analysts and users; as a communication tool between analysts and designers; as a basis for end-user developed applications; and as part of the systems documentation (e.g., Batra and Davis 1992; Juhn and Naumann 1985; Siau et al. 1997). A goal of creating a conceptual model is to develop a database schema to be used to implement a database that meets the information needs of intended users. To develop a suitable database schema, the designer must be able to use the conceptual data model as a communication tool to verify the assumptions made in its creation. Batra and Davis state that the conceptual model must be capable of providing a structure for the database along with the semantic constraints for communication with users. The conceptual data model also serves as a representation of the database after its completion: it is part of the systems documentation, and hence can be used for system evaluation by auditors or others. Conceptual data models include several components, each of which provides information content. Siau et al. examined the use of two components in entity-relationship data models: the surface semantics and the structural constraints (participation cardinality) of the relationships.

Siau et al. found that expert data modelers tended to ignore surface semantics in interpreting data models and instead relied almost exclusively on the structural constraints. They attributed the results to cognitive bias exhibited by the expert system modelers who participated in the task. However, alternative explanations can account for the attained results. The current study considers an alternative explanation and demonstrates that at least two factors influence the primacy of either surface semantics or structural constraints in decision-making. As in Siau et al. this study utilizes conceptual models in an entity-relationship format.

This study investigates decision makers' syntactic and semantic understanding of conceptual data modeling from the perspectives of text-centered theory (Faris and Smeltzer 1997) and schema theory (Anderson 1983, 1990; Mandler 1984; Rumelhart 1980). Text-centered theory posits that meaning is contained in the syntax and core word meanings of the written text and readers are passive recipients of the text; understanding based on text-centered theory is syntactic understanding. Schema theory contends that users assimilate text information into existing knowledge structures and sets of expectations (schemas) to derive meaning. How a user interprets a written (or other) communication depends on the information presented and on the user's schema. Understanding based on schema theory is semantic understanding. Semantic understanding is examined in both low and high information load contexts. When asked *what* the entity-relationship diagram portrays, the participants responded as predicted by text-centered theory, that is, they only looked at the syntax. When asked what the entity-relationship diagram *should* portray, the participants responded as predicted by schema theory, assimilating their pre-existing knowledge with the depicted information. These results were consistent across domains (general, acquisition, and revenue business processes). People apply different types of understanding based on their interpretation of different types of questions. This provides an alternative explanation for the results obtained by Siau et al. Participants' interpretation of the question asked, and the corresponding syntactic or semantic response, rather than cognitive bias, may have driven their results.

An interaction between information load and structural constraint type was also identified. When presented with mandatory structural constraints for a relationship whose underlying semantics are in conflict (i.e., optional participation should have been specified), the high information load group performed significantly worse than did low information load groups. This result was consistent across both the acquisition and revenue business processes. However, there was no significant difference in

---

[1]**Keywords:** Data models, REA models, system design, schema theory.

performance across the groups when presented with optional structural constraints when mandatory participation should have been specified. Participants were able to identify the appropriate participation as mandatory in low information loads when structural constraints depicted it as optional just as well as they could when no structural constraints were included. When the same relationships are examined in a high information load condition, participants are equally able to correct the incorrect structural constraints and identify the appropriate participation as mandatory. Bodart and Weber (1997) investigated the distinction between mandatory and optional structural constraints and predicted that individuals using conceptual models that employed subtypes with mandatory properties would better understand the underlying real-world domain than users of conceptual models with optional properties. Bodart and Weber were unable to support their hypothesis. The current study did not use subtypes, thus it did not specifically test Bodart and Weber's hypothesis; however, the results obtained in this research lend support to their premise that mandatory properties are more semantically meaningful than optional properties.

As a result of the information load finding, an implication of this study is that system evaluators should not be asked to assess diagrams of complete business processes, but should instead be presented with individual relationship clusters. The danger in this is that some participation cardinalities may be incorrectly identified as mandatory when optional participation by each of multiple alternative entities may be appropriate. Future research may examine whether a medium information load (i.e., enough multiple relationships to allow identification of alternative entities' participation, but smaller than an entire transaction cycle) changes the results, or whether alternative representations of relationships mitigate the negative effect of high information load.

References

Anderson, R. C. "Inferences About Word Meanings," in *The Psychology of Learning and Motivation*, A. C. Graesser and G. H. Bower (eds.), Academic Press, New York, 1990, pp. 1-16.

Anderson, R. C. "Some Reflections on the Acquisition of Knowledge," *Educational Researcher* (13), 1984, pp. 5-10.

Batra, D., and Davis, J. G. "Conceptual Data Modeling in Database Design: Similarities and Differences between Expert and Novice Designers," *International Journal of Man-Machine Studies* (37), 1992, pp. 83-101.

Bodart, F., and Weber, R. "Optional Properties Versus Subtyping in Conceptual Modeling: A Theory and Empirical Test," Working Paper, University of Queensland, 1997.

Faris, K. A., and Smeltzer, L. R. "Schema Theory Compared to Text-Centered Theory as an Explanation for the Readers' Understanding of a Business Message," *The Journal of Business Communication* (34:1), January, pp. 7-26.

Juhn, S., and Naumann, J. D. "The Effectiveness of Data Representation Characteristics on User Validation," in *Proceedings of the Sixth International Conference on Information Systems*, L. Gallegos, R. Welke, and J. Wetherbe (eds.), Indianapolis, IN, 1985, pp. 212-226.

Mandler, J. *Stories, Scripts, and Scenes: Aspects of Schema Theory*, Erlbaum, Hillsdale, NJ, 1984.

Rumelhart, D. E. "Schemata: The Building Blocks of Cognition," in *Theoretical Issues in Reading and Comprehension*, R. J. Spiro, B. Bruce, and W. F. Brewer (eds.), Erlbaum, Hillsdale, NJ.

Siau, K., Wand, Y., and Benbasat, I. "The Relative Importance of Structural Constraints and Surface Semantics in Information Modeling," *Information Systems* (22:2/3), 1997, pp. 155-170.