

Adding Discourse Statements to a Semantic Program

Robert B. Allen[✉]
independent
New York City, NY, USA
rba@boballen.info

Abstract

We have developed state-based semantic programs that describe scientific research phenomena. However, implementing interactive research reports requires more than a model of the phenomenon: it also requires discourse elements such as rhetorical and genre structures. We use the term discourse broadly to cover how material is presented. This paper treats those structures as an overlay on the semantic program of the base phenomenon. We discuss two examples related to the Snowball Earth Hypothesis.

CCS Concepts

• Information systems → Digital libraries and archives.

Keywords

Argumentation, Claims, Discourse States, Evidence, FrameNet, Ontology, Pragmatics, Research Genres, Research Reports, Rhetorical Structure Theory (RST), Structured Discourse

ACM Reference Format:

Robert B. Allen. 2026. Adding Discourse Statements to a Semantic Program. In *The 2026 ACM/IEEE Joint Conference on Digital Libraries (JCDL '26)*, October 13–16, 2026, Frisco, TX, USA. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3805696.3846494>

Introduction We have been exploring the creation of highly structured research reports (e.g., [1]). In earlier papers, we developed visualizations to explore mechanisms, causal threads, and experimental procedures. Recently, we introduced compilable and executable, state-based semantic programs for describing the models of the underlying research phenomena [8]. However, the semantic program did not include discourse elements that are needed to support presentations to users.

Discourse can be said to be actions by a speaker or author to adapt a message for a listener or reader. Pragmatics is closely related to discourse; it considers how context affects the presentation of relatively short passages. and we lump the two together. Here we emphasize high-level discourse structures such as Rhetorical Structure Theory (RST) [16] and Toulmin's structured argumentation [25]. Genre Theory [24] is also closely related; genres provide a relatively rigid structure that is recognized by a community of readers and sets up expectations for them.

Discourse can be used to organize formal presentations, such as scientific research reports. Typically, each section and even each

sentence has a purpose. We want to capture those statements using structured discourse to describe how each statement relates to the whole. Richer descriptions of the association should be useful. Here, we explore how to coordinate them with semantic programs (cf., [17]). However, none of the discourse frameworks introduced above is sufficient by itself; thus, we adopt and enhance features from each. Overall, our focus is on adequate representation rather than human cognition, complex inference, or text mining.

The plan for the remainder of the paper is as follows: Section 2 discusses knowledge structures and discourse structures. Section 3 discusses Rhetorical Structure Theory. Section 4 considers Genres of Research Reports. Section 5 discusses Structured Argumentation. Section 6 is the Conclusion.

1 From Knowledge Structures to Discourse Structures

Semantic programming is an executable, qualitative, state-based approach to representing semantic models [8]. The knowledge structures used in semantic programs build on ontologies such as BFO [10] and SUMO [18]. However, unlike those ontologies, we emphasize transitionals that are incorporated into a programming language and are used to make state changes.

We use Python as the platform for the semantic programs. We believe that executable programming languages are a viable alternative to static representations for knowledge structures. For instance, ontology terms may be treated as data types [4]. In addition, there are many tools for checking the validity of programs, and programs can directly support user interaction. If desired, the Python code could implement formal logic to represent the statements and propositions.

Consistent with our emphasis on transitionals, we obtain structured descriptions of them from linguistic resources such as FrameNet [19] and VerbNet [20]. In FrameNet, each Frame has specific Frame Elements that are integral to its usage. These Frame Elements are related semantic roles. Thus, frames are structured semantic units (cf., [9, 26]). However, the FrameNet corpus¹ was developed from a large, but still limited sample, of English usage. Thus, missed many words and word senses, especially in more technical uses (cf., [11]).

We assume an environment with a previously established knowledgebase and a microworld for the situation being investigated. Propositions are statements with a truth value. We consider Claims to be Propositions for which there is some degree of structured evidence. Because the term Claim is sometimes applied to any assertion, we refer to Claims with evidence as Final Claims. They can be integrated, at least tentatively, into the microworld and knowledgebase. Thus, we will have varying degrees of confidence about the processes asserted by the models, and the confidence level needs



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

JCDL '26, Frisco, TX, USA

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2597-5/2026/10

<https://doi.org/10.1145/3805696.3846494>

¹<https://framenet.icsi.berkeley.edu/>

to be captured by some parameters. It remains debatable when, if ever, a Final Claim should be treated as fact; though, as a practical matter, Final Claims that have met the test of time are generally accepted.

2 Rhetorical Structure Theory (RST)

RST describes relations for connecting blocks of text from an essay or report. According to the theory, there are main texts (nuclei) connected to supporting texts (satellites) by specific set of relationships [16].² Each nucleus and satellite can be considered a discourse state, and the relationships between them can also be considered as state transitions.

We use FrameNet-style schemas to structure RST and Genre transitions (see below).³ Just as the semantic program has state transitions that can be validated with FrameNet's semantic roles, transitions between RST states can also be validated. Note that these are not state changes in the microworld but in the information being presented to the message recipient. Some RST states have actions, others have situations, and still others have information. This simple typology imposes constraints on the content in each and can be validated. Other types of constraints can also be introduced and validated.

In some cases the distinction between semantic and discourse structures is blurred. For example, selecting the terms for the semantic model involves subjectivity. In addition, semantic programs include control structures like looping and conditionals. Consider the structured conditional statement in Java from one of our earlier papers [3]:

```
if ((NAD == LOW) && (Degeneration == LOW))
    then prefer ConceptualModel2;
else if ((NAD == LOW) && (Degeneration == NORMAL))
    then prefer ConceptualModel2;
```

This is a control structure in the program but can also be said to serve a discourse purpose can be modeled with RST Conditional and Anti-conditional nuclei.

Similarly, we considered causation in [6, 7]. There we made a distinction between mechanistic causation (as in the transitions of molecules in chemical reactions) and discourse-like assertions of causation that suggest a connection between seemingly remote events.

3 Research Report Genres

Here we move to modeling research reports. While our earlier work on semantic programming was applied to the dynamic systems of the phenomenon under investigation and was modeled in a constrained microworld, modeling entire research reports expands the scope of the entities and processes involved. There are information objects such as data sets and descriptions of established research procedures as well as the physical objects and contexts needed to record the information and to implement the procedures.

Furthermore, research reports typically have a formal structure. Swales [24] has proposed that the familiar IMRD (Introduction, Results, Methods, and Discussion) structure is a distinct stylistic genre. He discussed the processes associated with the Introduction in detail but did not go into detail about the other parts. Here, we hope to cover entire research reports.

IMRD Genres can be roughly modeled with RST and, in turn, RST can be represented with FrameNet-like schemas. A research report can be seen as a sequence of text blocks that provide different types of information. There are transitions both between and within genre sections.

Here, we consider several examples. We illustrate our approach with examples related to the Snowball Earth Theory. Early examination of the Elatina Formation north of Adelaide, Australia (relatively near the Equator) suggested that it was from the Neoproterozoic Era, but confusingly it showed evidence of extensive glaciation. The Fold Test is an established technique in geology that can provide evidence for latitude at the time of deposition. Sumner [23] applied the Fold Test to rock samples from the Elatina Formation.

When sediments are first deposited, magnetic particles in them align with Earth's magnetic field in a way that is directly related to latitude. Over time, tectonic movements may tilt or fold the rock layers. Researchers measure the magnetization and then apply a correction. If the corrected magnetic orientations are consistent, researchers can establish the latitude with confidence. Otherwise, the magnetization is likely acquired after folding. [23] concluded that the Neoproterozoic rocks in the Elatina Formation were near the Equator when they were deposited and glaciated. Later replications confirmed her finding. This raised the question of how rock formations near the equator could show extensive glaciation.

Here, we propose some FrameNet-like structures for RST-like relationships that can be applied to genre sections. These can also be considered as transitions because they move the reader's focus from nucleus to satellite. Note that, like FrameNet, they include both core (required) and non-core elements.

Research Question:

- Schema In
 - (non-core) Agent
 - (non-core) Need
 - (non-core) Approach to finding the solution.
 - (non-core) Agent
- Type In: The question
- Schema and Type Out
 - The Solution (Final Claim)
- Description: The Research Question is the primary nucleus of the research report. In RST, it is associated with a need. The response to the Research Question is the Final Claim and the entire report justifies the transition asserted by that Claim. In some cases, there are two levels of Research Questions: Broad and Narrow. Broad questions typically apply to issues raised by an application context while Narrow questions are based on the specific situation being examined.
- Example Application
 - Broad: Is there evidence of low-latitude glaciation during the Neoproterozoic Era?
 - Narrow: Have the rocks been in the same orientation since they were first deposited?

²<https://www.sfu.ca/rst/01intro/intro.html>

³The relationship between RST, Genres, and Structured Argumentation has been noted by [12, 15], but that work did not include schema links to an underlying notation.

Procedure:

- Schema In
 - Goal
 - Sequence of Volitional Causes
 - (non-core) expected Volitional Results associated with the Volitional Causes
 - Agent
- Type In: Situation
- Schema Out
 - Confirmation that the goal was reached, or the reason it was not reached
- Type Out: Situation
- Description: The Procedure in a research report is an Elaboration sequence of actions by an agent for obtaining evidence. It is a semi-ordered list of situated actions and expected results. Each step may be considered its own transition from Volitional Cause to Volitional Result. It must be consistent with the state of the world at the time it is triggered and help to accomplish the overall goal. The Procedure could be based on an established technique and may use standard instrumentation, both of which can also be taken from a knowledgebase. If, for some reason, the expected outcome is not obtained, additional procedures would be triggered. Note that in a research report, this is a description of the action; not the execution of the action.
- Example Application:
 - Data Collection Phase 1:
 - Identify rock formations of interest.
 - Identify individual rocks; for each, establish and record its coordinates and orientation.
 - Data Collection Phase 2:
 - In the laboratory, measure the magnetic orientation of the particles in the rocks.

The discourse segments and genre structures cascade constraints on the semantic content. For instance, the Research Question implies there is a Procedure that will address the Research Question and will provide relevant results. One cannot analyze data that has not yet been collected. And the results should provide at least a partial answer to the Research Question.

4 Limitations of Previous Research, Refining Claims, and Extended Argumentation

Many projects define argumentation discourse structures. However, most of these are very general and assert that a given Claim *supports* or *refutes* other findings (e.g., [13, 22]). Here, we describe the components of the argument based on Toulmin's well-known argumentation structure [25]. The overall framework is structured, but the contents of its elements (Facts, Warrant, Backing, etc.) are not structured.

In addition, as suggested by Swales' CARS model [24], the Introduction of most research reports includes a review of the literature. That review should present the Claims, limitations, and possible faults of relevant papers. As suggested in [5], the threats to internal and external validity can be applied for experiments and quasi-experiments [21]. For critiques of models (see below), the criticism

may be based on an inconsistency with the model or issues with its plausibility.

Assessing/Evaluating/Contrasting:

- Schema In
 - First alternative
 - Second alternative (possibly, more than two)
- Type In: Two (or more) situations
- Schema Out
 - Comparison of the alternatives
- Type Out: Information (Evaluative Comment)
- Description: This is synthesized from RST Evaluation and FrameNet's Assess/Evaluation
- Example Application: As an example, we consider competing Explanations of low-latitude glaciation. An early explanation [14, pp. 132-135] was that the Earth's axis of rotation could have been tilted obliquely toward the Sun. Thus, as the Earth revolves around the Sun, the hemispheres alternately experience extreme winters and during those extreme winters.

Evaluative Judgment/Criticism:

- Schema In
 - Object of the criticism: A statement in the model or a segment of the model itself.
 - Reason for the criticism
- Type In: Situation
- Schema Out
 - The criticism
- Type Out: Situation
- Description: Criticism is a type of Evaluative Judgment with a negative valence. We apply this to both situations and both final and non-final claims.
- Example Application: The oblique-axis model was criticized and rejected by [14, p. 135] because it also predicts very hot summers, which would prevent establishing long-term ice coverage. Note that this was not a matter of empirical evidence, but of the lack of consistency with other claims in the microworld.

Toulmin's model includes rebuttals, but those do not seem to richly represent recursive argumentation with the possibility of criticisms, elaboration, concessions, refinement, and reconceptualization. Such ongoing debates are relatively common in a set of related research reports in a digital library. Indeed, they could trace the evolution of an entire field.

5 Conclusion and Implications

We have proposed, and provide examples for, the development of highly structured discourse statements for research reports to accompany the structured semantic descriptions of the research phenomena.

We also believe that object-oriented, state-based, semantic programs are related to knowledge graphs. Some relationships between the entities are built into the program. Other relationships can be obtained by comparing the properties of objects, or by examining

object histories. For example, biological parent-of implies a past birthing relationship.

While RST and Genre Theory assume primarily linear presentation of material, our interactive interfaces provide more flexibility for users. Potentially, these programs can allow readers to examine different parts of the models much like navigating through a hypertext. As part of a digital library, this approach could change the nature of citations. Traditional citations with author, title, and publication could be replaced by concept-based linking [2]. However, we need additional work to maintain coherence with non-linear access.

Indeed, much work remains to reconcile the various discourse systems and semantic constraints. For instance, the corpora need to be greatly expanded. In addition, while we have indicated ways that discourse state changes can be used based on RST rules, a compiler for that has not been implemented. Overall, this paper should be considered a preliminary investigation with many details to be extended later.

References

- [1] Robert B. Allen. 2011. Model-Oriented Scientific Research Reports. *D-Lib Magazine* 17 (May 2011). doi:10.1045/may2011-allen
- [2] Robert B. Allen. 2012. Supporting Structured Browsing for Full-Text Scientific Research Reports. arXiv:1209.0036
- [3] Robert B. Allen. 2013. Rich Linking in a Digital Library of Full-Text Scientific Research Reports. columbia university research data symposium: <https://boballen.info/RBA/PAPERS/RDS/RDStext.pdf>
- [4] Robert B. Allen. 2019. Definitions and Semantic Simulations Based on Object-Oriented Analysis and Modeling. arXiv:1912.13186
- [5] Robert B. Allen. 2021. Yoked Flows for Direct Representation of Scientific Research Reports. *ceur-ws:2976_3*
- [6] Robert B. Allen. 2023. Using Causal Threads to Explain Changes in a Dynamic System. In *International Conference of Asian Digital Libraries (ICADL) (LNCS)*. Springer. doi:10.1007/978-981-99-8088-8_18
- [7] Robert B. Allen. 2026. Supporting Exploration of Structured Semantic Models. arXiv:in preparation Presented at ACM HT'26 workshop on Knowledge Graphs and Hypertext.
- [8] Robert B. Allen. 2026, in press. Semantic Models of Research Phenomena. In *IEEE eScience*. IEEE.
- [9] Robert B. Allen and Y-M. Chu. 2021. Semantic Modeling of Pottery Making. In *Pacific Neighborhood Consortium (PNC)*. doi:10.23919/PNC53575.2021.9672253
- [10] Robert Arp, Barry Smith, and Andrew D. Spear. 2015. *Building Ontologies with Basic Formal Ontology*. MIT Press, Cambridge, MA.
- [11] A.E. Dolbey. 2009. *BioFrameNet: a FrameNet extension to the domain of molecular biology*. Ph.D. Dissertation. University of California at Berkeley.
- [12] N. Ye. Doronkina. 2016. Argumentative Discourse of Research Papers in the Context of Rhetorical Structure Theory. *Bulletin of National Technical University of Ukraine (Kyiv Polytechnic Institute): Philology. Pedagogy: Collection of Scientific Papers* (2016). Issue 7. <https://ela.kpi.ua/items/4a2d7f68-a8da-4624-9671-8c633bfd0052>
- [13] Paul Groth, Andrew Gibson, and Jan Velterop. 2010. The Anatomy of a Nanopublication. *Information Services and Use* 30 (2010), 51–56.
- [14] Paul F. Hoffman and Daniel P. Schrag. 2002. The Snowball Earth hypothesis: Testing the limits of global Change. *Terra Nova* 14 (2002), 129–155.
- [15] O. V. Ivasiuk and N. Ye Doronkina. 2021. Argumentation Features of Scientific Review Paper in the Context of Rhetorical Structure Theory. *Scientific Bulletin of the International Humanitarian University: Series: Philology* 51 (2021).
- [16] William C. Mann and Sandra A. Thompson. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text—Interdisciplinary Journal for the Study of Discourse* 8 (1988), 243–281. doi:10.1515/text.1.1988.8.3.243
- [17] Johanna D. Moore and Martha E. Pollack. 1995. A Problem for RST: The Need for Multi-Level Discourse Analysis. In *Association for Computational Linguistics*.
- [18] Adam Pease. 2011. *Ontology: A Practical Guide*. Articulate Press.
- [19] Josef Ruppenhofer, Michael Ellsworth, Miriam R.L. Petruck, Christopher R. Johnson, and Jan Scheffczyk. 2010. *FrameNet II: Extended Theory and Practice*. https://my.eng.utah.edu/~cs6961/papers/FrameNet_book.pdf.
- [20] Karin K. Schuler. 2005. *Verbnet: A Broad-Coverage, Comprehensive Verb Lexicon*. Ph.D. Dissertation. University of Pennsylvania.
- [21] William R. Shadish, Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Research Designs for Generalized Causal Inference*. Houghton Mifflin.
- [22] Simon B. Shum. 2000. ScholOnto: An Ontology-based Digital Library Server for Research Documents and Discourse. *International Journal of Digital Libraries* 3 (2000), 237–248.
- [23] Dawn Y. Sumner, Joseph L. Kirschvink, and Bruce N. Runnegar. 1987. Soft-sediment Paleomagnetic Field Tests of Late Precambrian Glaciogenic Sediments. *EOS: Transactions of the American Geophysical Union* 68 (1987), 1251. Issue 44.
- [24] John M Swales. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.
- [25] Stephen Toulmin. 1958. *The Uses of Argument*. Cambridge University Press.
- [26] Lars Vogt. 2026. The Semantic Units Framework, a Technology-Agnostic Representational Approach to FAIR and CLEAR Knowledge Infrastructures. *Scientific Data* 13 (2026). doi:10.1038/s41597-026-07588-3