

A Tool for Deep Semantic Encoding of Narrative Texts

David K. Elson

Columbia University
New York City

delson@cs.columbia.edu

Kathleen R. McKeown

Columbia University
New York City

kathy@cs.columbia.edu

Abstract

We have developed a novel, publicly available annotation tool for the semantic encoding of texts, especially those in the narrative domain. Users can create formal propositions to represent spans of text, as well as temporal relations and other aspects of narrative. A built-in natural-language generation component regenerates text from the formal structures, which eases the annotation process. We have run collection experiments with the tool and shown that non-experts can easily create semantic encodings of short fables. We present this tool as a stand-alone, reusable resource for research in semantics in which formal encoding of text, especially in a narrative form, is required.

1 Introduction

Research in language processing has benefited greatly from the collection of large annotated corpora such as Penn PropBank (Kingsbury and Palmer, 2002) and Penn Treebank (Marcus et al., 1993). Such projects typically involve a formal model (such as a controlled vocabulary of thematic roles) and a corpus of text that has been annotated against the model. One persistent tradeoff in building such resources, however, is that a model with a wider scope is more challenging for annotators. For example, part-of-speech tagging is an easier task than PropBank annotation. We believe that careful user interface design can alleviate difficulties in annotating texts against deep semantic models. In this demonstration, we present a tool we have developed, SCHEHERAZADE, for deep annotation of text.¹

We are using the tool to collect semantic representations of narrative text. This domain occurs

frequently, yet is rarely studied in computational linguistics. Narrative occurs with every other discourse type, including dialogue, news, blogs and multi-party interaction. Given the volume of narrative prose on the Web, a system competent at understanding narrative structures would be instrumental in a range of text processing tasks, such as summarization or the generation of biographies for question answering.

In the pursuit of a complete and connected representation of the underlying facts of a story, our annotation process involves the labeling of verb frames, thematic roles, temporal structure, modality, causality and other features. This type of annotation allows for machine learning on the thematic dimension of narrative – that is, the aspects that unite a series of related facts into an engaging and fulfilling experience for a reader. Our methodology is novel in its synthesis of several annotation goals and its focus on content rather than expression. We aim to separate the narrative’s *fabula*, the content dimension of the story, from the rhetorical presentation at the textual surface (*sjuzet*) (Bal, 1997). To this end, our model incorporates formal elements found in other discourse-level annotation projects such as Penn Discourse Treebank (Prasad et al., 2008) and temporal markup languages such as TimeML (Mani and Pustejovsky, 2004). We call the representation a *story graph*, because these elements are embodied by nodes and connected by arcs that represent relationships such as temporal order and motivation.

More specifically, our annotation process involves the construction of propositions to best approximate each of the events described in the textual story. Every element of the representation is formally defined from controlled vocabularies: the verb frames, with their thematic roles, are adapted from VerbNet (Kipper et al., 2006), the largest verb lexicon available in English. When the verb frames are filled in to construct action

¹Available at <http://www.cs.columbia.edu/~delson>.

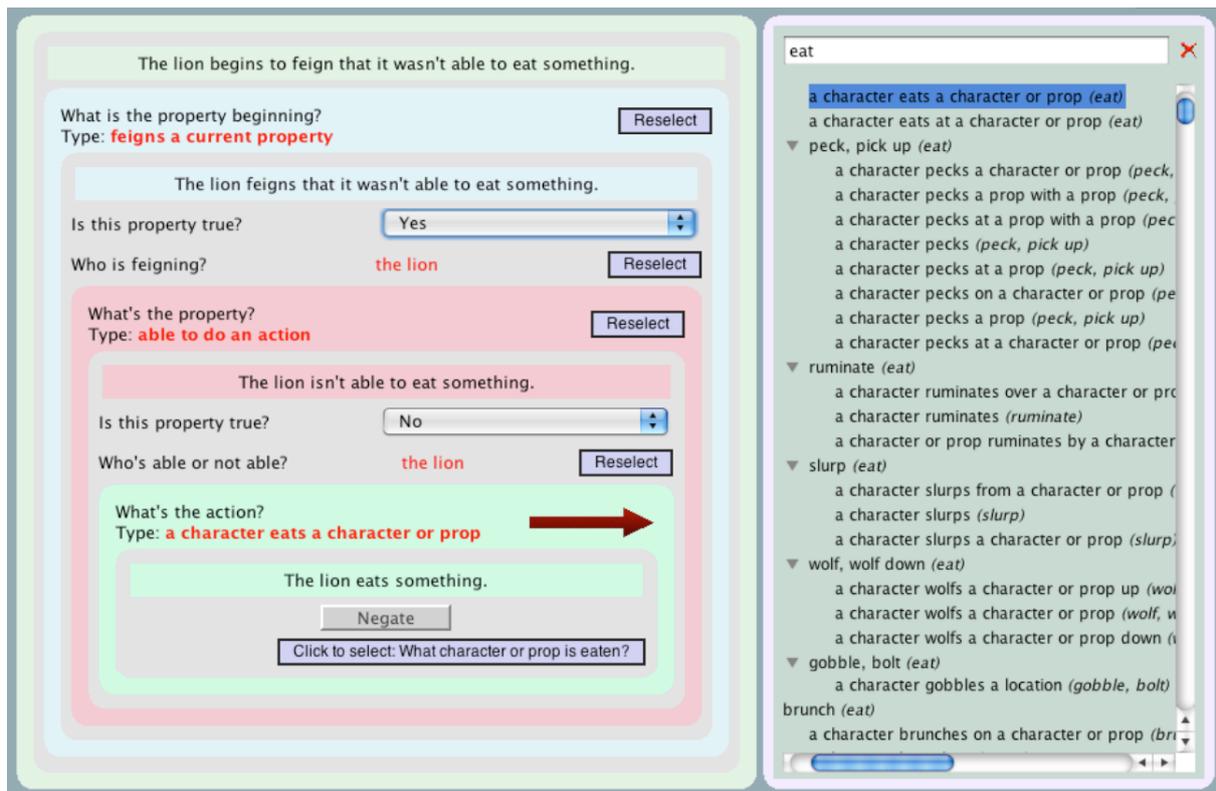


Figure 1: Screenshot from our tool showing the process of creating a formal proposition. On the left, the user is nesting three action propositions together; on the right, the user selects a particular frame from a searchable list. The resulting propositions are regenerated in rectangular boxes.

propositions, the arguments are either themselves propositions or noun synsets from WordNet (the largest available noun lexicon (Fellbaum, 1998)). Annotators can also write stative propositions and modifiers (with adjectives and adverbs culled from WordNet), and distinguish between goals, plans, beliefs and other “hypothetical” modalities. The representation supports connectives including causality and motivation between these elements. Finally, and crucially, each proposition is bound to a state (time slice) in the story’s main timeline (a linear sequence of states). Additional timelines can represent multi-state beliefs, goals or plans. In the course of authoring actions and statives, annotators create a detailed temporal framework to which they attach their propositions.

2 Description of Tool

The collection process is amenable to community and non-expert annotation by means of a graphical encoding tool. We believe this resource can serve a range of experiments in semantics and human text comprehension.

As seen in Figure 1, the process of creating a

proposition with our tool involves selecting an appropriate frame and filling the arguments indicated by the thematic roles of the frame. Annotators are guided through the process by a natural-language generation component that is able to realize textual equivalents of all possible propositions. A search in the interface for “flatter,” for example, offers a list of relevant frames such as *<A character> flatters <a character>*. Upon selecting this frame, an annotator is able to supply arguments by choosing actors from a list of declared characters. “The fox flatters the crow,” for one, would be internally represented with the proposition *<flatters>([Fox₁], [Crow₁])* where *flatters*, *Fox* and *Crow* are not snippets of surface text, but rather selected WordNet and VerbNet records. (The subscript indicates that the proposition is invoking a particular [Fox] instance that was previously declared.) In this manner an entire story can be encoded.

Figure 2 shows a screenshot from our interface in which propositions are positioned on a timeline to indicate temporal relationships. On the right side of the screen are the original text (used for reference) and the entire story as regenerated from

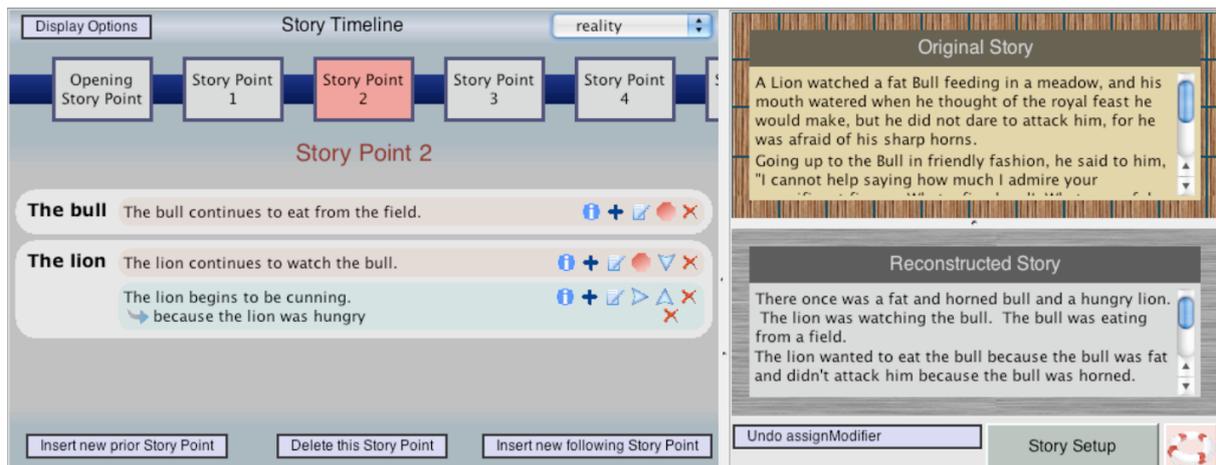


Figure 2: The main screen of our tool features a graphical timeline, as well as boxes for the reference text and the story as regenerated by the system from the formal model.

the current state of the formal model. It is also possible from this screen to invoke modalities such as goals, plans and beliefs, and to indicate links between propositions. Annotators are instructed to construct propositions until the resulting textual story, as realized by the generation component, is as close to their own understanding of the story as permitted by the formal representation.

The tool includes annotation guidelines for constructing the best propositions to approximate the content of the story. Depending on the intended use of the data, annotators may be instructed to model just the stated content in the text, or include the implied content as well. (For example, causal links between events are often not articulated in a text.) The resulting story graph is a unified representation of the entire *fabula*, without a story's beginning or end. In addition, the tool allows annotators to select spans of text and link them to the corresponding proposition(s). By indicating which propositions were stated in the original text, and in what order, the content and presentation dimensions of a story are cross-indexed.

3 Evaluation

We have conducted several formative evaluations and data collection experiments with this interface. In one, four annotators each modeled four of the fables attributed to Aesop. In another, two annotators each modeled twenty fables. We chose to model stories from the Aesop corpus due to several key advantages: the stories are mostly built from simple declaratives, which are within the expressive range of our semantic model, yet are rich

in thematic targets for automatic learning (such as dilemmas where characters must choose from between competing values).

In the latter collection, both annotators were undergraduates in our engineering school and native English speakers, with little background in linguistics. For this experiment, we instructed them to only model stated content (as opposed to including inferences), and skip the linking to spans of source text. On average, they required 35-45 minutes to encode a fable, though this decreased with practice. The 40 encodings include 574 propositions, excluding those in hypothetical modalities. The fables average 130 words in length (so the annotators created, on average, one proposition for every nine words).

Both annotators became comfortable with the tool after a period of training; in surveys that they completed after each task, they gave Likert-scale usability scores of 4.25 and 4.30 (averaged over all 20 tasks, with a score of 5 representing "easiest to use"). The most frequently cited deficiencies in the model were abstract concepts such as *fair* (in the sense of a community event), which we plan to support in a future release.

4 Results and Future Work

The end result from a collection experiment is a collection of story graphs which are suitable for machine learning. An example story graph, based on the state of the tool seen in Figure 2, is shown in Figure 3. Nodes in the graph represent states, declared objects and propositions (actions and statives). Each of the predicates (e.g., <lion>,

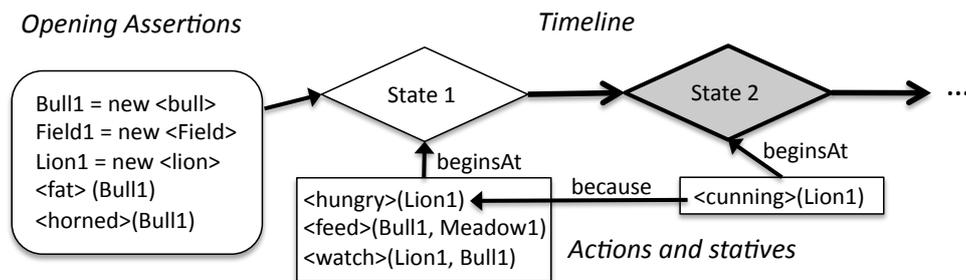


Figure 3: A portion of a story graph representation as created by SCHEHERAZADE.

<watch>, <cunning>) are linked to their corresponding VerbNet and WordNet records.

We are currently experimenting with approaches for data-driven analysis of narrative content along the “thematic” dimension as described above. In particular, we are interested in the automatic discovery of deep similarities between stories (such as analogous structures and prototypical characters). We are also interested in investigating the selection and ordering of content in the story’s telling (that is, which elements are stated and which remain implied), especially as they pertain to the reader’s affectual responses. We plan to make the annotated corpus publicly available in addition to the tool.

Overall, while more work remains in expanding the model as well as the graphical interface, we believe we are providing to the community a valuable new tool for eliciting semantic encodings of narrative texts for machine learning purposes.

5 Script Outline

Our demonstration involves a walk-through of the SCHEHERAZADE tool. It includes:

1. An outline of the goals of the project and the innovative aspects of our formal representation compared to other representations currently in the field.
2. A tour of the timeline screen (equivalent to Figure 2) as configured for a particular Aesop fable.
3. The procedure for reading a text for important named entities, and formally declaring these named entities for the story graph.
4. The process for constructing propositions in order to encode actions and statives in the text, as seen in Figure 1.

5. Other features of the software package, such as the setting of causal links and the ability to undo/redo.
6. A review of the results of our formative evaluations and data collection experiments, including surveys of user satisfaction.

References

- Mieke Bal. 1997. *Narratology: Introduction to the Theory of Narrative*. University of Toronto Press, Toronto, second edition.
- Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.
- Paul Kingsbury and Martha Palmer. 2002. From treebank to propbank. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC-02)*, Canary Islands, Spain.
- Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extensive classifications of english verbs. In *Proceedings of the 12th EURALEX International Congress*, Turin, Italy.
- Inderjeet Mani and James Pustejovsky. 2004. Temporal discourse models for narrative structure. In *Proceedings of the ACL Workshop on Discourse Annotation*, Barcelona, Spain.
- Mitchell P. Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a large annotated corpus of english: The penn treebank. *Computational Linguistics*, 19.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltasakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The penn discourse treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.