

Hand-coded *scripts* were used in the 1970-80s as knowledge backbones that enabled inference and other interpretation tasks requiring deep semantic knowledge. We propose unsupervised induction of similar schemata called *narratives* from raw newswire text.

Narratives are partially ordered sets of events centered around a common **protagonist**. They are related to structured sequences of participants and events that have been called *scripts* (Schank and Abelson, 1977) or *Fillmorean frames* (Fillmore, 1977). These participants and events can be filled in and instantiated in a particular text situation to draw inferences. Narratives focus on a single actor to facilitate learning, and thus this work addresses two tasks of narrative induction: *narrative event induction* and *structured selection* (pruning the event space into discrete sets).

Learning these prototypical schematic sequences of events is important for rich understanding of text. Scripts were central to natural language understanding research in the 1970s and 1980s for proposed tasks such as summarization, coreference resolution and question answering. For example, Schank and Abelson (1977) proposed that understanding text about restaurants required knowledge about the Restaurant Script, including the participants (Customer, Waiter, Cook, Tables, etc.), the events constituting the script (entering, sitting down, asking for menus, etc.), and the various preconditions, ordering, and results of each of the constituent actions.

Consider these two distinct narratives (letter variables used in place of named entities):

1. y accused X, X claimed, X argued, z dismissed X
2. W joined W, served W, W oversaw, W resigned

It would be useful for question answering or textual entailment to know that ‘**X denied v**’ is also a likely event in the left narrative, while ‘**u replaces W**’ temporally follows the right. Narratives offer the structure and power to directly infer these new subevents by providing the critical background knowledge. In part due to their complexity, automatic induction has not been addressed since the early non-statistical work of Mooney and DeJong (1985).

The first step to narrative induction uses an entity-based model for learning narrative relations by following a protagonist. Our model is inspired by Centering (Grosz et. al, 1995) and other entity-based models of coherence (Barzilay and Lapata, 2005). As a narrative progresses through a series of events, each event is characterized by the grammatical role played by the protagonist, and by the protagonist's shared connection to surrounding events. Our algorithm is an unsupervised distributional learning approach that uses coreferring arguments as evidence of a narrative relation. We show, using a new cloze evaluation task, **narrative cloze**, that our protagonist-based method leads to better induction than current statistical approaches.

Finally, the space of narrative events can be clustered and pruned to create discrete sets of narratives. We compare this automatic induction with current hand built linguistic resources such as FrameNet (Baker et. al, 1997).

References

Baker, Collin F., Charles J. Fillmore, and John B. Lowe. 1998. "The Berkeley FrameNet project." In *Proceedings of COLING ACL*, pages 86–90, Montreal, Canada.

Regina Barzilay and Mirella Lapata. 2005. Modeling local coherence: an entity-based approach. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 141–148.

Fillmore, Charles J. 1977. The need for a frame semantics in linguistics. In Hans Karlgren (ed.), *Statistical Methods in Linguistics*. Scriptor. 5-29.

J.B. Lowe, C.F. Baker, and C.J. Fillmore. 1997. A frame-semantic approach to semantic annotation. in *Proceedings of the SIGLEX workshop "Tagging Text with Lexical Semantics: Why, What, and How?"* held April 4-5, in Washington, D.C., USA in conjunction with ANLP-97.

Barbara Grosz, Aravind Joshi, and Scott Weinstein, "Centering: A Framework for Modelling the Local Coherence of Discourse." *Computational Linguistics*, 2(21), June (1995).

Raymond Mooney and Gerald DeJong. 1985. Learning schemata for natural language processing. In *Ninth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 681–687.

Roger C. Schank and Robert P. Abelson. 1977. *Scripts, plans, goals and understanding*. Lawrence Erlbaum.