Abs-199

Leon Derczynski and Robert Gaizauskas (both University of Sheffield)

RTMBank: A corpus annotated according to Reichenbach's tense model

Tense and time have a significant role in language. They allow the reader to locate and relate events in time. Hence the ability to automatically process and interpret tense is a crucial part of the long term programme aiming at computation language understanding.

In his 1947 account, Reichenbach (Reichenbach 1947) offered an analysis of the tenses of verbs, in terms of abstract time points. He posited that each tensed verb can be modelled with a _speech time_ for when the verb was uttered, an _event time_ that is the point where the events described by the verb occur, and a _reference time_ from which events are viewed. For example, in ``She will have eaten", if the speech time is the present, we are referring to an event that happens in the future -- that is, speech time is before event time -- and this is described from a viewpoint after the ``eating" event, so that the reference time is later than event time. Thus, we can say that for this tensed verb, speech time < event time < reference time.

RTMML (Derczynski 2011) is a markup language for annotating the tenses of verbs and temporal relations between verbs, that aims to support automated processing of tense and temporal relations in language. RTMML differs from TimeML (Pustejovsky 2004) in that (1) it chiefly only annotates verbs that indicate events, (2) the information annotated about verbs is more nuanced, and (3) inter-verb links are defined using Reichenbach's three abstract points instead of event boundaries. In this paper we describe the annotation scheme, introduce an RTMML annotation tool, and document the creation of an RTMML-annotated corpus, RTMBank.

There are other temporally annotated corpora, the largest and most detailed of which is TimeBank (Pustejovsky 2003) -- a TimeML annotated corpus of 183 newswire articles from US outlets. However, TimeML leaves out some information which is critical to the Reichenbachian model. The three time points are also useful for some tasks that TimeBank was intended to help with, such as the anchoring of temporal expressions on a calendrical timeline. The documents in RTMBank are chosen from those already in TimeBank. Since people using RTMBank may want additional information only contained in TimeBank. Finally, chosing documents for RTMML annotation that are already annotated in TimeML permits partial verification of our annotation effort.

In our full paper, we discuss the composition of RTMBank and include some summary statistics. We also examine temporal context and the temporal relations between verbs, where a temporal context is defined as a time frame shared by one or more events.

Finally, we conclude with potential applications of the corpus, including the training and evaluation of automated systems.

L. Derczynski and R. Gaizauskas. 2011. An Annotation Scheme for Reichenbachs Verbal Tense Structure. In Sixth Joint ACL - ISO Workshop on Interoperable Semantic Annotation. ACL.

J. Pustejovsky, P. Hanks, et al. 2003. The TimeBank Corpus. In Corpus Linguistics, volume 2003, page 40.

J. Pustejovsky, B. Ingria, R. Sauri, J. Castano, J. Littman, R. Gaizauskas, A. Setzer, G. Katz, and I. Mani. 2004. The specification language TimeML. The Language of Time: A Reader. Oxford University Press, Oxford.

H. Reichenbach. 1947. The Tenses of Verbs. Elements of Symbolic Logic, pages 287–98.