Temporal annotation within frame semantics

L. Barque, M. Candito, P. Muller

Annotating temporal information in corpora is one way to acquire knowledge for NLP applications that require time stamping and ordering of events, such as question answering and summarization. Markup language frameworks have been developped for this purpose, among them TimeML (Pustejovsky *et al.* 2005). We present here another framework for coding temporal information that relies on a precise description of lexical signals (e.g temporal prepositions and conjunctions) to compensate for a smaller number of relations. Indeed, the more relations a framework uses, the more the annotation task is error-prone (Lapata and Lascarides, 2006).

The Asfalda Project is an ongoing project whose aim is to build semantic resources and a corresponding semantic analyser for French (Candito *et al* 2014). The project attempts to cover several semantic domains, including the temporal domain, in an unified way, namely the frame semantics framework, as carried out in the Berkeley FrameNet Project (Baker *et al* 1998). The English FrameNet resource provides a structured set of prototypical situations (*frames*) and their prototypical participants (*frame elements*), a lexicon in which lexical units evoke these frames and a corpus of annotated sentences from the BNC. This framework is essentially lexically based since lemmas trigger semantic frames. In (1), the preposition *during* triggers the Temporal_collocation frame, allowing the identification of the Trajector and the Landmark. The semantics of the temporal relation is encoded in the lexicon (*during, after, while, …*).

(1) ***During*** *[last night's walk]*$_{\text{Landmark}}$, *[I stubbed my toe]* $_{\text{Trajector}}$.

In this presentation, we will first describe the modelization of temporal information in the French FrameNet and the effects of our methodology on this modelization (domain-by-domain strategy rather than frame-by-frame or along occurrences of lemmas in corpus, as in the original FrameNet project). We focus on temporal relations between eventualities or between an eventuality and a temporal expression, leaving aside aspectual aspects of the temporal description. We will then compare this description to the TimeML markup language, already used to annotate French data (Bittar et al. 2011), and show how these two semantic metalanguages are compatible. For now, annotated data are too sparse to be exploited by supervised machine learning techniques and it seems important to pool annotation efforts. We will finally expose issues that arise from the annotation of temporal information in such a lexically-based framework (implicit temporal relations, temporal relations expressed by constructions like gerund, …).

Références

Baker C., Fillmore C., et Lowe B. (1998) The Berkeley FrameNet project. In *COLING-ACL '98: Proceedings of the Conference,* pages 86–90, Montreal, Canada.

Bittar A., Amsili P., Denis P. et Danlos L (2011) French TimeBank: An ISO-TimeML Annotated Reference Corpus. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics,* pages 130–134. Association for Computational Linguistics.

Candito M., Amsili P., Barque L., Benamara F., de Chalendar Gaël, Djemaa M., Haas P., Huyghe R., Mathieu Y., Muller P., Sagot B. Vieu L. (2014) Developing a French FrameNet : Methodology and first result, *Proceedings of the 9th edition of the Language Resources and Evaluation Conference*, 26-31 May, Reykjavik, Iceland

Lapata M. and Lascarides A. (2006) Learning Sentence-Internal Temporal Relations. Journal of Artificial Intelligence Research 27: 85–117.

Pustejovsky J., Ingria B., Sauri R., Castano J., Littman J., Gaizauskas R., Setzer A., Katz G., and Mani I. (2005) The Specification Language TimeML. In Mani I, Pustejovsky J. Gaizauskas R. The Language of Time: A Reader (pp 545–57). Oxford University Press.