NPs for Events: experiments in coreference annotation

Laura Hasler*, Constantin Orasan*, Karin Naumann†

*Research Group in Computational Linguistics, SHLSS, University of Wolverhampton, UK
{L.Hasler, C.Orasan}@wlv.ac.uk

†Seminar für Sprachwissenschaft, Universität Tübingen, Abt. Computerlinguistik, Deutschland
knaumann@sfs.uni-tuebingen.de
Introduction to the NP4E Project

• **Observation:** NP and event coreference important for applications such as question answering, information extraction, multi-document summarisation. But - limited number of existing resources, especially for event coreference.

• **Pilot project:** developed a methodology for within-document NP and event coreference, focusing on detailed annotation schemes and guidelines. Narrow definition of coreference. 50,000 word sample annotated corpus, terrorism/security domain. Extensive discussions between annotators to determine most appropriate guidelines.

• **Potential:** methodology provides a basis for future large-scale annotations to produce much-needed resources, ensuring quality and consistency.
The NP4E Pilot Project

- **2 layers of annotation:** NP annotation re-used for events to ease annotation task. NPs annotated in 1st phase fill argument slots for events annotated in 2nd phase, helping to identify coreferential relations.

- **NPs:** relations: identity, synonymy, generalisation, specialisation, other forms: NP, copular, apposition, brackets, speech pronoun, other

- **Events:** 5 categories: ATTACK, DEFEND, INJURE, DIE, CONTACT
  4 triggers: verb, noun, pronoun, adjective (polarity, tense, modality)
  different argument slots for each event category, filled by NPs

- **PALinkA:** multi-purpose annotation tool significantly extended for event argument slots. Plug-in support added to help annotators, e.g. WordNet query for relations, annotation previewer to view markables and relations.
Annotation of NPs using PALinkA

The plane is marked as coreferential The aircraft

The user can override WordNet's decision

WordNet is consulted about the relation between the two NPs

The aircraft will fly out of Xiamen and Shantou, said...
Annotation of an attack event using PALinkA

The hostage crisis erupted on Dec. 17 when MRTA rebels stormed the Japanese ambassador's residence in Lima during a garden party honouring the Japanese emperor's birthday.

The leader of MRTA, Nestor Cerpa, on
Coreference annotation for NPs and events is very difficult!

We still have several outstanding issues:

- producer/recipient distinction in CONTACT events, especially threats etc.
- classification of tense in direct speech in newswire texts
- classification of participles as a separate trigger type
- annotation of relative pronouns
- annotation of general concepts as coreferential
- multiple possible indefinite antecedents for an NP

Check out the project webpage: http://clg.wlv.ac.uk/projects/NP4E