

NPs for Events: experiments in coreference annotation



Laura Hasler*, Constantin Orasan*, Karin Naumann†

***Research Group in Computational Linguistics, SHLSS,
University of Wolverhampton, UK**

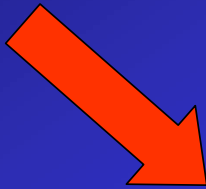
{L.Hasler, C.Orasan}@wlv.ac.uk

**†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.

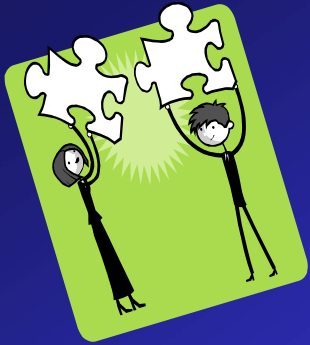


- **Pilot project:** developed a methodology for within-document NP and event coreference, focusing on detailed annotation schemes and guidelines. Narrow definition of coreference. 50,000 word sample annotated corpus, terrorism/security domain. Extensive discussions between annotators to determine most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

Palinka [D:\laura\BA_coreference_project\texts.tar\corpus\China-Taiwan_plane_hijack_/430308newsML-done.xml.xml]

File Edit View Tools Tags Plugins Windows Help

Select the element to which the link SRC is. ESC to cancel

to [Taiwan] rather than detouring through [[Hong Kong] or [another ...

third place "]], [a striking break with [past practice]], [executives] said .

"_ [The aircraft] will fly out_ of [[Xiamen] and [Shantou

midpoint of [the Taiwan strait]] and then fly directly into [Taipei] , "_ said

[Lee Chung-ning] , [general manager of [Far East Air Transport] [whose

[Boeing 757] was hijacked [earlier in [the day]]]].

[Lee] said skipping [a detour through [[Hong Kong] or [Macau]]] was

suggested by [[China's] government] in [the interest of convenience] . "_

[it] was requested by [the air traffic controllers in [Xiamen]] that [the

plane] fly directly to [Taiwan] , "_ [Lee] said .

Monday
the day
Monday
hantou
taipei
Taiwan
Taiwan
Taipei
Taiwan
Taiwan
Taiwan

Choose value

TYPE_REL SPECIALISATION

OK Cancel

to China on

The aircraft
the plane
Far East Air Transport's Boeing 757
Taiwan's decades-old ban on direct flights
across the Taiwan strait
the ban
Xiamen
Xiamen
Xiamen and Shantou
a " third place " , always Hong Kong pr

Wordnet says that plane is more specialised than aircraft

Messages Tags Notes

Current tag:<MARKABLE COMMENT="" ID="10"><COREF CERTAIN="TRUE" COMMENT="" ID="75" SRC="0" TYPE_REF="NP" TYPE_REL="SYNONYM">aircraft

D:\laura\BA_coreference_project\texts.tar\corpus\China-Taiwan_plane_hijack_/43... |BRACKETED_TEXT|SPEECH_PRON|OTHER" COMMENT=""... CERTAIN="TRUE">

The user can override WordNet's decision

WordNet is consulted about the relation between the two NPs

the plane is marked as coreferential The aircraft

Annotation of NPs using PALinka

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 the operation] [MRTA leader Nestor Cerpa], on

the operation

TYPE: attack

TIME: Dec. 17

REF: stormed

TARGET: the Japanese
ambassador's residence
in Lima (FACILITY)

ATTACKER: MRTA rebels
(PERSON)

PLACE: Lima (LOCATION)

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>