

COREA: **C**oreference **R**esolution for **E**xtracting **A**nswers

Gosse Bouma

Stevin midterm, 27/06/08, Rotterdam

COREA Project

- Goals
 - Coreference **Corpus**
 - Automatic Coreference **Resolution**
 - Evaluate effect in **applications**
- Team
 - **University Groningen**: Gosse Bouma, Anne-Marie Mineur, Geert Kloosterman
 - **University of Antwerp**: Walter Daelemans, Iris Hendrickx, Véronique Hoste
 - **Language & Computing N.V.**: Jean-Luc Verschelde, Frederik Coppens, Joeri Van Der Vloet
- **Website**: www.cnts.ua.ac.be/~iris/corea.html

Motivation

Many NLP applications require coreference resolution

- Information Extraction, Summarization, Question Answering

Gas station collapses

Gas station Hoezaar next to highway A58 has collapsed monday afternoon. The building came down after being hit by a truck with a flat tyre.

Alopecia areata is a hair disease. [...]

In extreme cases, the disease may lead to complete baldness.

Annotation

- Coreference relations for names, pronouns, noun phrases (NPs).
- Annotation in XML
 - Extract markables from treebank annotation
 - MMAX annotation tool
 - Stylesheet for visualisation in web browser
- Material
 - newspaper, spoken language, medical
 - Dutch and Flemish
 - 200K words

Types of Coreference

- **Identity (IDENT)**
Xavier Malisse qualified for the semi finals. *The Flemish tennis player* will play against an unknown opponent.
- **Quantification (BOUND)**
Everybody did what *they* could.
- **Superset – Subset (BRIDGE)**
200 people died in that plain crash. *Forty-six* are buried here on this cemetery.
- **Predicative relations (PRED)**
Michel Beuter is *a writer*.
- Special mark-up for negation, modality, time dependency

Corpus statistics

Corpus	DCOI	CGN	MedEnc	Knack
#docs	105	264	497	267
#tokens	35,166	33,048	135,828	122,960
#ident	2,888	3,334	4,910	9,179
#bridge	310	649	1,772	na
#pred	180	199	289	na
#bound	34	15	19	43

Agreement (2 annotators, 29 docs, \pm 500 rels)

Identity	
MUC-score (1 = Gold, 460 rels)	0.71
MUC-score (2 = Gold, 367 rels)	0.82
F-score	0.76

Relation	F-score
Bridge (42-44 rels)	0.33
Pred (11 -31 rels)	0.56
Bound (3 rels)	0.00

Coreference resolution as classification task

Supervised Machine Learning approach

- Identify the NPs in the text,
- Link every NP to the previous NPs,
- **Step one**: classify each pair as coreferential or not
- **Step two**: make coreference chain of positive pairs
- **MUC-score** : 0.48-0.52 (newspaper), 0.39 (medical)

Features Used

- Morphological-lexical
- Syntactic, Semantic
- String Matching, Positional

Effect on Question Answering

- Answers to frequently asked questions are extracted off-line
- Extraction patterns modified to take coreference into account

When was Jan Toorop born?

Jan Toorop, was a Dutch painter whose works straddle the space between the Symbolist painters and Art Nouveau. He was born in Poerworedjo (Java) , on December, 20, 1858 and he died in the Hague on March, 3, 1928.

Effect on Question Answering

patterns	# facts	accuracy
no coref	93K	
pronouns	4K	40%
definite NPS	47K	34%

Table: Extracted facts for different patterns

variant	accuracy
without	65.0%
with	70.0%

Table: Accuracy on relevant questions in QA@CLEF 2005.

Effect on Information Extraction

- **Relation Finder**: predict semantic medical relations in Spectrum Medical Encyclopedia annotated with medical concepts and relations
- **Concepts**: *con_disease, con_person, con_treatment*
- **Relations**: *rel_is_symptom_of, is_cause_of, rel_treats*
- Using annotated corpus of the IMIX Rolaquad (UvT) project

Relation Finder

- Core: Maximum Entropy Modeling algorithm
- Trained on 2000 encyclopedia entries
- Tested on two test sets of 50 and 500 different entries

test set	without	with
small(50)	53.03	53.51
Big(500)	59.15	59.60

Table: F-Scores of Relation Finder wo/w coreference resolution.

Results and Plans

- **Results**
 - Coreference **Corpus** (200K words), Guidelines, Annotation software
 - **Resolution Module**
 - Integrated and evaluated in **2 applications** (QA, IE)
- **Plans**
 - **Groningen**: Integrate coreference resolution in full QA system (Tiedemann & Mur, IR4QA, Coling 2008 ws)
 - **Antwerpen**: coreference resolution for multi-document summarization (Stevin/DEASO)