

Anaphora Resolution

Vašek Němčík
(xnemcik@fi.muni.cz)

NLP Laboratory,
FI MU Brno

RASLAN
December 14, 2007

Outline

- ▶ Introduction & task definition
- ▶ State-of-the-art methods
- ▶ Enhancing rule-based systems
- ▶ Machine learning systems

What is “anaphora”?

- ▶ **anaphor** – relates to another expression in the discourse
 - ▶ pronouns (“he”, “him”, “their”, “itself”)
 - ▶ definite NPs (“the apple”, “the faculty”)
- ▶ **antecedent** – the related expression
- ▶ *Example:*
[Peter]_i ate [a cake]_j.
[He]_i was hungry and [the cake]_j looked delicious.
- ▶ **anaphora** – the phenomenon
- ▶ **anaphora resolution (AR)**
determining the antecedent for each anaphor

The AR task

- ▶ to identify anaphoric expressions in the given text
 - ▶ deictics, non-referential pronouns (expletives, clefts)
- ▶ to determine the antecedent for each such anaphor
- ▶ to determine the relation type
 - ▶ coreference
(two expressions refer to the same discourse entity)
identity of reference \times identity of sense
 - ▶ bridging (associative/indirect anaphora)
(any semantic relation between two discourse entities)
 - ▶ Hypernymy/Hyponymy (animal – dog)
 - ▶ Part/Whole relation (house – door)
 - ▶ Entity/Attribute relation (car – colour)
 - ▶ Cause/Effect relation (sun – light)

AR systems (for English)

- ▶ first AR systems: based on heuristic rules
(*pattern matching, tree traversal,...*)
- ▶ efforts to incorporate semantics and inference
(*centering, focusing*)
~> success only for (very) restricted domains
- ▶ knowledge-poor (90s)
 - ▶ motivated by practical needs (not theory)
 - ▶ RAP, Kennedy&Boguraev's system, CogNIAC, MARS
 - ▶ low-level (easily computable) features used
(especially based on morphology and shallow syntax)
 - ▶ sacrificing completeness
- ▶ statistics and machine learning (*mentioned later*)

AR systems (for Czech)

- ▶ theoretical work (FGD: Sgall, Hajičová)
- ▶ PDT 2.0
large treebank, annotated for coreference
- ▶ Nguy Giang Linh, MFF ÚFAL, Prague
rules & machine learning

- ▶ modular AR system
- ▶ separation of processing levels:
 - ▶ markable level (algorithm)
 - ▶ technical level (text representation)
- ▶ (un)biased evaluation & comparison
- ▶ selected salience methods implemented

Enhancing Rule-Based AR

- ▶ knowledge-poor trend: semantics sacrificed
- ▶ unrealistic to aim at a (computationally plausible) semantic theory accounting for everything
- ▶ realistic to use “some semantics” only
 ~> (a subset of) lexical semantics
- ▶ relevant resources available (for Czech)
 - ▶ WordNet (poodle–dog–animal–being)
 on its own useless for pronominal AR
 - ▶ *Wikipedia* (not hand-crafted, real-world instances, grows faster)
 - ▶ valency lexicons (Vallex and Verbalex)

Enhancing Rule-Based AR (cont.)

The typical problem:

- (1) *Robot_i* odnesl prázdnou misku od *ovoce_j*
Robot (MASC) took the empty bowl of fruits (NEUT)
“The robot took away the empty fruit bowl”
- (2) a Alvar si teprve díky tomu uvědomil,
and Alvar only then thanks to this realized,
“and only after noticing this Alvar realized”
- (3) že *ho_{*i,j}* vůbec snědl.
that him/it (MASC/NEUT) actually ate.
“that he actually ate it.”

Enhancing Rule-Based AR (cont.)

- ▶ Verbalex valency frame for “jíst” (*to eat*):
AG<person:1>kdo1 **Alvar**
VERB
SUBS<food:1>co4 **ovoce** × **robot**
INS<cutlery:2>>čím7
- ▶ check (using WN) whether the antecedent candidate is a hyponym of the synset marked in the valency slot
- ▶ rule it out if not
- ▶ *even better*: reward if it is
- ▶ *alternative*: use corpus-based co-occurrence statistics (high co-occurrence rate \Rightarrow more semantically related \Rightarrow more likely in the anaphor–antecedent relation)

Machine Learning and AR

- ▶ recent corpora \rightsquigarrow training data
- ▶ machine learning works for classification tasks
- ▶ AR is not a classification task
(the AR problem has a different structure)
 \Rightarrow need to transform AR to fit
- ▶ Transformation 1 (Connolly et al., 1994):
Antecedent₁ Antecedent₂ Anaphor **1/2**
- ▶ Transformation 2 (McCarthy and Lehnert, 1995):
Antecedent Anaphor **Y/N**
- ▶ attributes usually based on knowledge-poor features
- ▶ 2 steps: classification & clustering

Enhancing ML-Based AR (cont.)

- ▶ a very detailed description of the anaphor–antecedent relationship
- ▶ advantageous for nominal coreference resolution (the relationship is more important than context)
- ▶ rather unsuitable for grasping pronominal anaphora
 - ▶ (the dynamic progress of) salience
 - ▶ interplay of the antecedent candidates
- ▶ \rightsquigarrow use pre-computed salience values as attributes
- ▶ \rightsquigarrow re-compute these values in the clusterer (to account for anaphoric links already found)
- ▶ different reformulation of AR as an classification task?

Thank you for your attention!