Associative anaphors in the Copenhagen Dependency Treebanks (CDT)

Iørn Korzen and Matthias Buch-Kromann

Copenhagen Business School

ik.ikk@cbs.dk
Associative anaphors in the Copenhagen Dependency Treebanks (CDT)

1. A brief presentation of the Copenhagen Dependency Treebanks, CDT.

2. A few terminological remarks and references.

3. A description of the CDT classification of associative anaphors.

4. Epilogue
   a. Annotation graphs
   b. Inter-annotator agreement count
The Copenhagen Dependency Treebanks

A set of parallel treebanks for Danish, English, German, Italian, and Spanish, annotated for part-of-speech, syntax, morphology, discourse, and nominal anaphora.

The Danish corpus consists of a number of excerpts from mixed-genre texts, amounting to 100,000 words in all, which have been translated into the other languages.

A main objective of the CDT: to arrive at a unified cross-linguistic description and annotation system for syntax, morphology, discourse and anaphora (Buch-Kromann, Korzen & Müller 2009). The CDT manual can be downloaded from Buch-Kromann et al. 2010.
A few terminological remarks and some references

In the theoretical linguistic literature: main distinction between coreferential and associative anaphors

In the computational literature, a frequent term is bridging anaphors
(e.g. Clark 1975; Poesio et al. 1997; Vieira and Poesio 2000; Caselli 2009: 73):
“definite descriptions that either
(i) have an antecedent denoting the same discourse entity, but using a different head noun (as in house . . . building), or
(ii) are related by a relation other than identity to an entity already introduced in the discourse”. (Vieira and Poesio 2000: 558).

(i): subtype of coreferential anaphors; (ii): associative anaphors.
Recent schemes for anaphoric annotation

Some confine themselves to coreference relations:
  - the VENEX corpus (Poesio et al. 2004),
  - the Potsdam Coreference Scheme (PoCoS) (Krasavina and Chiarcos 2007)
  - the Potsdam Commentary Corpus (Stede 2008)
  - the Portuguese and French corpus analysed by Vieira et al. (2002).

Others include certain associative relations such as set membership, subset, ownership, and part-of relations:
  - the GNOME Corpus (Poesio 2004),
  - the ARRAU Corpus (Poesio and Artstein 2008),
  - the Dutch COREA corpus (Hendrickx et al. 2008),
  - the Italian Live Memories Corpus (Rodríguez et al. 2010).

The Prague Dependency Treebank, PDT (Nedoluzhko et al. 2009) includes contrast, location–resident, relatives, and event–argument relations.


It is CDT’s ambition to include all kinds of anaphoric relations, coreferential as well as associative.
The Generative Lexicon and association

• FORMAL QUALE: That which distinguishes the object within a larger domain (shape, dimensionality, color, position).

• CONSTITUTIVE QUALE: The relation between an object and its constituents, or proper parts (material, weight, parts and component elements).

• AGENTIVE QUALE: Factors involved in the origin or “bringing about” of the object.

• TELIC QUALE: Purpose and function of the object.

Figure 1. Pustejovksy’s (1995, 76ff / 85ff) “Qualia Structure”.

On Qualia Structure and associative anaphors:
Bos et al. (1995); Lundquist (2000); Henry & Bassac (2008); Caselli (2009); Korzen (2000; 2003; 2009).

CDT’s approach: a combination of the qualia roles with other semantic roles.
The associative anaphors in the CDT

Two parameters:
1. lexical semantics and generativity, qualia structure;
2. semantic roles in relation to a predicate; the predicate may be either directly expressed by the antecedent or generatable from it, possibly by means of the qualia structure:

<table>
<thead>
<tr>
<th>1. Qualia structure</th>
<th>2. Semantic roles</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASSOC-FORMAL</td>
<td>ASSOC-AGENT</td>
</tr>
<tr>
<td>ASSOC-CONST (itutive)</td>
<td>ASSOC-PATIENT</td>
</tr>
<tr>
<td>ASSOC-AGENTIVE</td>
<td>ASSOC-EXPER (iencer)</td>
</tr>
<tr>
<td>ASSOC-TELIC</td>
<td>ASSOC-REC (ipient)</td>
</tr>
<tr>
<td>ASSOC-EVENT</td>
<td>ASSOC-INST (rument)</td>
</tr>
</tbody>
</table>

Table 1. Associative subtypes, parameters and CDT-labels.
1. The anaphor is associated with the antecedent with regard to its qualia structure
(a. FORMAL and CONSTITUTIVE express static information about the object)

ASSOC-FORMAL: shape, dimension, colour, taste, etc.

(1) The ham to be used in the dish must not be too salty. You cannot use the thin slices, they are too salty and too wet and the flavour [ASSOC-FORMAL] is not good enough. (148)

ASSOC-CONST (parts, elements, material, content, etc.): The predicates of which antecedent and anaphor are arguments are e.g. has as a part, consists of, is part of, etc. In (2) the anaphor is part of the antecedent, in (3) vice versa.

(2) The accident took place at dinner time around 6:45 p.m. last night […]. I saw the plane with its nose pointing downward, the left wing [ASSOC-CONST] up and the right wing [ASSOC-CONST] down over behind the flat building. (1536)

(3) On September 8, DE BEERS CENTENARY opened an office in Moscow. Present were also De Beers’ top people, Russian politicians, diplomats and representatives of the country’s [ASSOC-CONST] diamond industry and trade. (431)
1. The anaphor is associated with the antecedent with regard to its qualia structure (b. AGENTIVE and TELIC)

ASSOC-AGENTIVE and ASSOC-TELIC: Dynamic information about the antecedent. Anaphors designate the quale predicate itself, (4)-(5), or an inferable argument of such a predicate (6)-(8):

(4) We were waiting for an approval from Sony as we submitted a new version of Blood Bowl PSP. This new version has been finally approved and the production [ASSOC-AGENTIVE] started.

(5) Not all debriefings are held after the simulation, but in certain instances, for example, where the aim [ASSOC-TELIC] is to teach a technical skill […] debriefing may occur during the simulation.

(6) In April 2003, marking the tenth anniversary of the Waco Massacre, a new film was released. According to the producer [ASSOC-AGENTIVE.AGENT/(produce)], “Waco: A New Revelation” is a film so disturbing that […] it triggered new investigations in both houses of Congress […].

(7) The accident took place at dinner time around 6:45 p.m. last night, shortly after the El-Al flight […] lifted off from Amsterdam's Schiphol airport. The pilot [ASSOC-TELIC.AGENT/(fly)] suddenly reported to the control tower that he had engine problems […].

(8) Two journeyman tests were passed in August. Both apprentices [ASSOC-TELIC.PATIENT/(examine)] are trained at the Royal Copenhagen. (431)
2. The antecedent is predicative and the anaphor is a semantic role

(9) *The operation* itself requires general anesthesia ... the patient is asleep for the entire course of the operation. *The surgeon* [ASSOC-AGENT] opens the chest by dividing the breast bone or sternum. (http://www.heartsurgeons.com/pr3.html, accessed August 5th, 2010).
(The tree dots appeared as shown in the cited text)

(10) *The operation* itself requires general anesthesia ... *the patient* [ASSOC-PATIENT] is asleep for the entire course of the operation.

(11) *The accident* took place at dinner time around 6:45 p.m. last night […]. “[…] The pilot attempted to right the plane - then I could not see more, but suddenly there were sparks in the air,” says *eyewitness Peter de Neef* [ASSOC-EXPER]. (1536)

(12) “[…] This is *the most violent attack* to this point. *The bombs* [ASSOC-INST] fell half a mile from the hotel,” reported John Hollimann […] (61).
3. “Extensions” (time, location, event)

An ASSOC-TIME anaphor indicates a point in time linked to the antecedent, a predicate or predicative noun, another time indication, (13), or a more general narrative frame, (14):

(13) The season will begin on March 16 with the showdown between AGF and Brøndby, followed the day after [ASSOC-TIME] by games between: Ikast-Lyngby and B 1903-Silkeborg. (43)

(14) Aspiring chef dies hours after making ultra-hot sauce for chilli-eating contest [headline]

Andrew Lee made an ultra-hot sauce with homegrown chillis. The morning after [ASSOC-TIME] he was found unconscious and paramedics were unable to revive him.

An ASSOC-LOC anaphor is located in the antecedent (or vice versa) without being necessarily a constitutive part:

(15) The officers saw the kitchen with many dirty dishes, spoiled food on the floor and in the refrigerator [ASSOC-LOC], and bags of trash on top of the stove [ASSOC-LOC].

A predicative anaphor may express an EVENT which is associable with the antecedent, but not necessarily with regard to its qualia structure:

(16) Hamid Jafar was very eager to show his appreciation of the agreement to his Iraqi partners. Shortly before the invasion [ASSOC-EVENT], he ordered an engraved, Swiss, gold pistol assessed at 7,000 pounds […]. (939)
## The complete picture

<table>
<thead>
<tr>
<th>ASSOC-QUALIA</th>
<th>ASSOC-SEMROLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(± semantic role subtype)</td>
<td>(semantic roles)</td>
</tr>
<tr>
<td>ASSOC-FORMAL</td>
<td>ASSOC-AGENT</td>
</tr>
<tr>
<td>ASSOC-CONST</td>
<td>ASSOC-EXPER</td>
</tr>
<tr>
<td>ASSOC-AGENTIVE</td>
<td>ASSOC-INST(rument)</td>
</tr>
<tr>
<td>ASSOC-AGENTIVE.AGENT</td>
<td>ASSOC-PATIENT</td>
</tr>
<tr>
<td>ASSOC-TELIC</td>
<td>ASSOC-REC(ipient)</td>
</tr>
<tr>
<td>ASSOC-TELIC.AGENT</td>
<td></td>
</tr>
<tr>
<td>ASSOC-TELIC.EXPER</td>
<td></td>
</tr>
<tr>
<td>ASSOC-TELIC.INST(rument)</td>
<td></td>
</tr>
<tr>
<td>ASSOC-TELIC.PATIENT</td>
<td></td>
</tr>
<tr>
<td>ASSOC-TELIC.REC(ipient)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Extensions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ASSOC-EVENT</td>
<td>ASSOC-LOC(ation)</td>
</tr>
<tr>
<td></td>
<td>ASSOC-TIME</td>
</tr>
</tbody>
</table>

Table 2. Associative anaphora in the CDT.
Epilogue a: Annotation graphs

Regarding the DTAG annotation tool, see Kromann (2003)

Figure 2. CDT anaphor annotation (below the nodes) and syntax annotation (above the nodes) of the sentence I saw the plane with its nose pointing downward, the left wing up and the right wing down over behind the flat building.

The NP the plane (nodes 7-8, the noun being “nominal object”, nobj, of the determiner, as indicated in the syntax annotation) is the antecedent of a coreferential pronoun (node 10) and two ASSOC-CONST anaphors (nodes 15-17 and 20-22).

The CDT-manual can be downloaded from the URL of Buch-Kromann et al. (2010).
Epilogue b: Inter-annotator agreement count

<table>
<thead>
<tr>
<th>Relation name</th>
<th>Agreement % ( A - A_U - A_L )</th>
<th>Relation count</th>
<th>Relation name</th>
<th>Agreement % ( A - A_U - A_L )</th>
<th>Relation count</th>
</tr>
</thead>
<tbody>
<tr>
<td>COREF</td>
<td>84 – 85 – 92</td>
<td>141</td>
<td>ASSOC-TELIC</td>
<td>71 – 88 – 83</td>
<td>24</td>
</tr>
<tr>
<td>COREF-VAR</td>
<td>71 – 79 – 79</td>
<td>97</td>
<td>ASSOC</td>
<td>39 – 83 – 39</td>
<td>9</td>
</tr>
<tr>
<td>REF</td>
<td>100 – 100 – 100</td>
<td>63</td>
<td>ASSOC-LOC</td>
<td>100 – 100 – 100</td>
<td>5</td>
</tr>
<tr>
<td>COREF-IDEN</td>
<td>77 – 83 – 81</td>
<td>53</td>
<td>ASSOC-AGENTIVE</td>
<td>25 – 50 – 50</td>
<td>4</td>
</tr>
<tr>
<td>ASSOC-CONST</td>
<td>59 – 77 – 67</td>
<td>39</td>
<td>ASSOC-EVENT</td>
<td>100 – 100 – 100</td>
<td>3</td>
</tr>
<tr>
<td>COREF-RES</td>
<td>65 – 73 – 72</td>
<td>25</td>
<td>ASSOC-FORMAL</td>
<td>100 – 100 – 100</td>
<td>1</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>TOTAL</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Agreement % ( A - A_U - A_L )</th>
<th>Relation count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>77 – 84 – 84</td>
<td>464</td>
</tr>
</tbody>
</table>

Table 3. Inter-annotator agreement based on 25 CDT texts with 464 anaphoric relations.

The associative anaphors are printed in *italics*, the rest are coreferential.

*Full labelled agreement, A:* the probability that another annotator assigns the same label and out-node to the relation;

*Unlabelled agreement, AU:* the probability that another annotator assigns the same out-node (but not necessarily the same label) to the relation;

*Label agreement, AL:* the probability that another annotator assigns the same label (but not necessarily the same out-node) to the relation.
References


References


References


References


http://cswww.essex.ac.uk/staff/poesio/publications/VENEX04.pdf


