

Co-reference resolution

Yohan Chalier

Symbolic Natural Language Processing (SD213)

June 11, 2018

What is co-reference?

Co-reference occurs when two or more expressions refers to the same referent. Usually, one expression is in a full form (the *antecedent*) and the other one in a abbreviated form (a *proform*).

For example:

*The **music** was so loud that **it** couldn't be enjoyed.*

Co-reference resolution is needed to derive a correct interpretation of a text.

The problem of co-reference resolution

Naive algorithm

Look for the nearest preceding individual that is compatible with the referring expression.

It solves sentences like this:

The girl₁ likes her₁ brother₂ and protects him₂.

But it fails to differentiate those sentences:

He_? said that John_? was coming.

His₁ sister said that John₁ was coming.

Domination and c-command

Domination

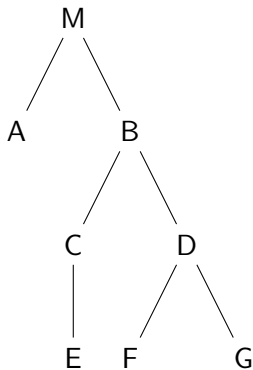
Node N_1 dominates node N_2 if N_1 is above N_2 in the tree and one can trace a path from N_1 to N_2 moving only downwards in the tree (never upwards).

c-command

Node N_1 c-commands node N_2 if

- N_1 does not dominate N_2
- N_2 does not dominate N_1
- The first (i.e. the lowest) branching node that dominates N_1 also dominates N_2

Domination and c-command



Co-reference and c-command

It was hypothesized that one restriction between proform and antecedent is that **the proform cannot appear in a position where it c-commands its antecedent.**

This is not trivial: Bouchard, Denis. (2010). *Une explication cognitive des effets attribués à la c-commande dans les contraintes sur la coréférence*. Corela. 10.4000/corela.965.