

Transformations in Data-Oriented Parsing

*Transforming declarative sentences
into polar questions and vice versa*

Who? Andreas van Cranenburgh (0440949)
Joost Winter (9928545)

From? Statistical Structure in Language Processing, University of Amsterdam

When? May 25, 2010

Abstract

- Extend DOP with **transformations**: given a declarative sentence, get a polar question and vice versa
- Generalize **synchronous grammar** for CFGs to STSGs (DOP).
- Contrast with transformational grammar:
no Deep Structure, pure **performance model**

Example

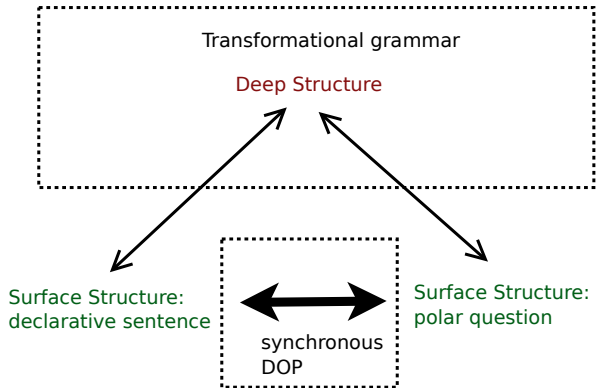


Figure: Transformational Grammar vs. Synchronous DOP

Synchronous grammars

- The idea: to have rules, consisting of a single left hand side, and two right hand sides
- Traditionally used for translation; may also be applied to transformations
- In the case of DOP: we do not consider pairs of rules with a shared left hand side, but pairs of subtrees with a shared root node

Representation

Constituency structure & movement

- Problem: **Structure mismatch** between declarative sentence & polar question
- Solution:
 - **Distituents**: constituents with gaps
 - **Dependency** structures w/phrase structure
 - **Ignore** subtrees for which there is no parallel subtree (does this cause overfitting / under-generalization?)
- Problem: **auxiliary verb** in question but not in declarative sentence
Example: mary walked \Leftrightarrow did mary walk?

PTB Example

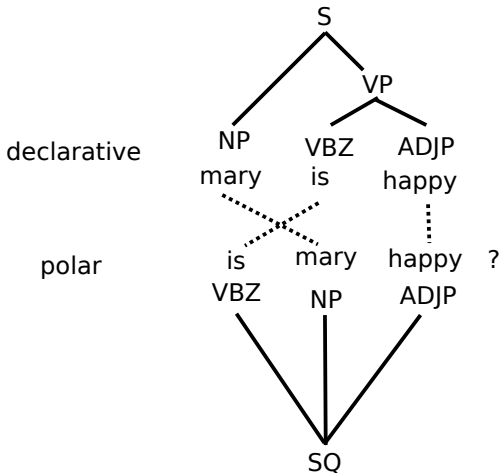


Figure: According to Penn Treebank bracketing

Distituent example

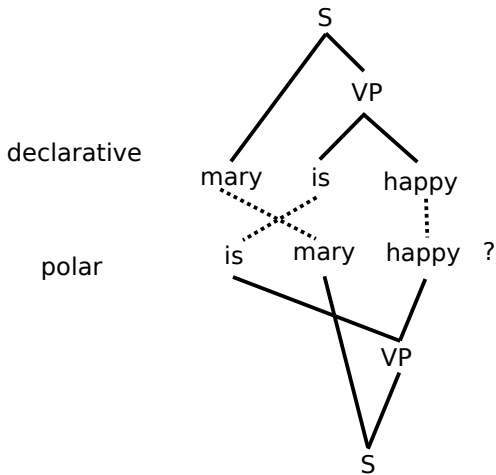


Figure: Using a distituent

Subtrees & Operations

Subtrees:

Instead of with parallel rules, we now work with parallel subtrees.

Operations:

Synchronous substitution

Parallel Treebank

Acquiring a gold standard treebank:

- Pilot: **Hand annotated** gold standard
- or: available already?
- induce treebank through **bootstrapping**?
semi-supervised or unsupervised?
(ie., take Penn Treebank and turn every declarative sentence into a tuple with a polar question as well)

Evaluation: standard PARSEVAL measures.

Planning

- Define formalism
- Implementation
- Acquire treebank
- Evaluation