

Three research topics

Helge Dyvik

1. Treebank-based grammar development

1. Treebank-based grammar development

How useful and realistic is it to assume that a given distinction can be recovered with sufficient precision by parsing and subsequent manual or statistical disambiguation?

1. Treebank-based grammar development

Problematic for disambiguators:

(a) Adjunct attachment (23 % of disagreement cases)

Example: PP-attached focus adverb or S-attached sentence
adverb in

Han så det også uten briller

‘He saw it also without glasses’

Solution: remove focus (PP-attached) possibility when not initial

1. Treebank-based grammar development

Problematic for disambiguators:

(b) *den/det: that* (demonstrative), referring *it* (pronoun), expletive *it* (expl. pronoun), *the* (article before adjectives)?

(13 % of disagreement cases)

Solution: 1. Merge article and demonstrative
2. Always pronoun, never demonstrative when alone

1. Treebank-based grammar development

Problematic for disambiguators:

(c) Epistemic/deontic vs. root modals (5 % of disagreement cases; very frequently wrong in statistical disambiguation)

ville, skulle, kunne, måtte

Han kan reise

'He may travel' (epistemic/deontic) or

'He is able to travel' (root)

Disambiguating syntax:

Han kan ha reist 'He may have travelled' (epistemic)

Han har kunnet reise 'He has been able to travel' (root)

Solution: remove the distinction since syntactically unambiguous cases are a minority.

2. The representation of argument structure

2. The representation of argument structure

Experience from corpus-based grammar development:

It is hard to maintain the idea of a definite, listable set of frames for each verb entry.

3. Cross-linguistic individuation of semantic roles?

3. Cross-linguistic individuation of semantic roles?

There are many competing role inventories and role hierarchies

3. Cross-linguistic individuation of semantic roles?

There are many competing role inventories and role hierarchies

Can role inventories be derived less subjectively from data?

3. Cross-linguistic individuation of semantic roles?

There are many competing role inventories and role hierarchies

Can role inventories be derived less subjectively from data?

Translational approach:

Can the semantic roles of a given verb be classified or indexed by the sets of syntactic functions realizing them across languages?

3. Cross-linguistic individuation of semantic roles?

There are many competing role inventories and role hierarchies

Can role inventories be derived less subjectively from data?

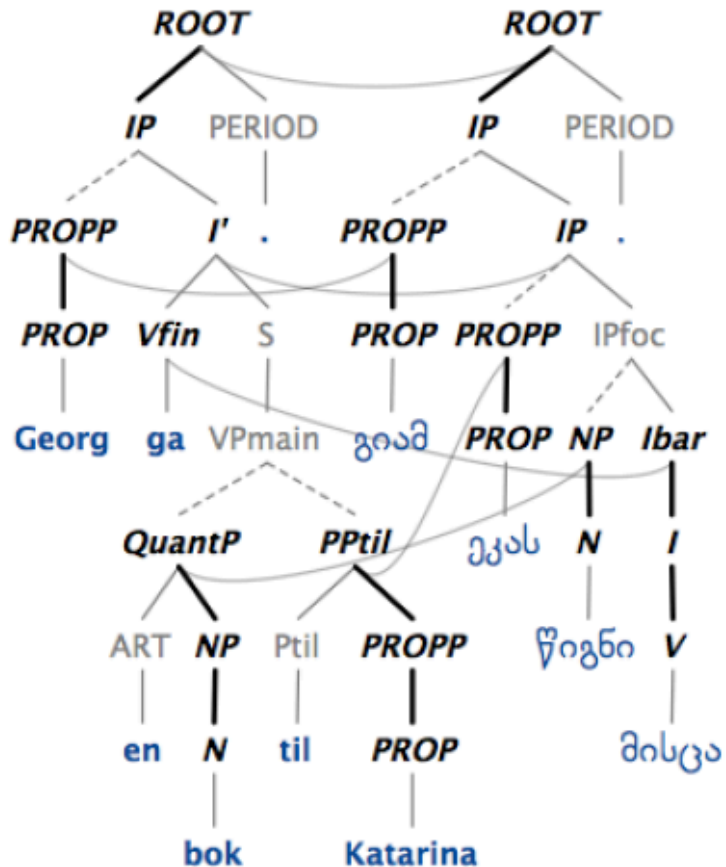
Translational approach:

Can the semantic roles of a given verb be classified or indexed by the sets of syntactic functions realizing them across languages?

Drawback: we need parallel treebanks.

Example from XPar (automatic phrase alignment):

C-structures



F-structures

