Towards a Grand Unified Theory of Scope Underspecification

Alexander Koller
Saarland University
Saarbrücken, Germany

joint work with
Joachim Niehren, Stefan Thater, Ruth Fuchss

University of Pennsylvania
11 October 2004
Overview

- Ambiguity and underspecification
- Dominance graphs
- Are different underspecification formalisms equivalent?
- Encoding of MRS nets into dominance nets
- Relevant descriptions are nets
- Harvesting
Ambiguity

- Natural language sentences often have more than one possible meaning, syntactic structure, etc.

- Example: (syntactic) attachment ambiguity: 
  John watched the man with the telescope.

- This phenomenon is called ambiguity.

- Ambiguity is one of the great challenges for natural language processing today.
Scope ambiguities

- Scope ambiguity is a kind of semantic ambiguity.

- Example:
  "Every student reads a book."

  1. ..., namely, the one their professor wrote.
     \[\exists y \text{ book}(y) \land (\forall x \text{ student}(x) \rightarrow \text{read}(x,y))\]

  2. ..., but not necessarily the same one.
     \[\forall x \text{ student}(x) \rightarrow (\exists y \text{ book}(y) \land \text{read}(x,y))\]
A sentence with more than one scope ambiguity can have an enormous number of readings:

Most politicians can fool most voters on most issues most of the time, but no politician can fool every voter on every single issue all of the time.
(ca. 600 readings, Hobbs)

Modern large-scale grammars predict a lot of scope readings even for harmless-looking sentences:

But that would give us all day Tuesday to be there.
(ca. 65,000 readings, according to ERG grammar)

In general, scope ambiguities contribute a number of readings exponential in the number of quantifiers (and other scope bearers).
Semantics construction

Sentence → Syntax

→ semantic repres. 1
→ semantic repres. 2
→ semantic repres. 3
→ semantic repres. 4
Enumerating readings is expensive

- We'd like to avoid enumerating these many readings.
- Most of the readings were not meant by the speaker.
- Do people enumerate readings of a scope ambiguity?
Scope Underspecification

- So let's avoid enumerating the readings for as long as we can.
- Take a single syntactic analysis and derive a single **underspecified semantic representation** from it.
- Possibly perform inferences on underspecified descriptions to remove unwanted readings.
- Then enumerate readings from description by need.
Scope Underspecification: The big picture

Sentence $\rightarrow$ Syntax $\rightarrow$ USR $\rightarrow$ semantic repres. 1
semantic repres. 2
semantic repres. 3
semantic repres. 4
Dominance graphs

- Semantic representations are e.g. formulas of first-order logic.
- Read these formulas as trees.
- Describe these trees using graphs that can be embedded into them.
- Use special graph edges to represent variable binding (not shown here).
- Equivalent to normal dominance constraints, i.e. both graph view and logic view available.
"Every student reads a book."

\[ \exists y \text{ book}(y) \land \forall x. \text{student}(x) \rightarrow \text{read}(x,y) \]

\[ \forall x. \text{student}(x) \rightarrow \exists y \text{ book}(y) \land \text{read}(x,y) \]
"Every student reads a book."

\[ \forall y \text{ book}(y) \land \forall x \text{ student}(x) \rightarrow \text{read}(x,y) \]

\[ \forall x \text{ student}(x) \rightarrow \exists y \text{ book}(y) \land \text{read}(x,y) \]
∀x student → x
∃y book y ∧ read x y
∀x student → x
∃y book y ∧ read x y
Solutions of dominance graphs

∀x student → x

∃y book y ∧ ∃x read x y

∀x student → x

∃y book y ∧ ∃x read x y
Not a solution
Solutions can be larger than graph
Solutions vs. configurations

- A **solution** is a tree into which the graph can be embedded.

- A **configuration** is a tree into which the graph's fragments can be configured, i.e. arranged without adding new nodes.
Underspecification Formalisms

QLF
(Alshawi & Crouch 92)

UDRT
(Reyle 93)

Hole Semantics
(Bos 96)

Dominance constraints/graphs
(Koller et al. 00)

LDG
(Muskens 95)

MRS
(Copestake et al. 99)

CLLS
(Egg et al. 98)

...
Grand Unified Theory?

- Underspecified descriptions in many of these formalisms look very similar.
- What are the formal relations between them?
- Are the formalisms equivalent?
- At least on some useful fragment of language?
Resource Sharing

- Large-scale grammars
- Usp. description (MRS)
- Usp. description (dominance)
- Efficient solvers
Underspecification Formalisms

QLF
(Alshawi & Crouch 92)

UDRT
(Reyle 93)

LDG
(Muskens 95)

Hole Semantics
(Bos 96)

Dominance constraints/graphs
(Koller et al. 00)

MRS
(Copestake et al. 99)

CLLS
(Egg et al. 98)

...
Underspecification Formalisms

QLF
(Alshawi & Crouch 92)

MRS
(Copestake et al. 99)

Hole Semantics
(Bos 96)

UDRT
(Reyle 93)

Dominance constraints/graphs
(Koller et al. 00)

LDG
(Muskens 95)

CLLS
(Egg et al. 98)

MRS
(Copestake et al. 99)

...
Underspecification Formalisms

QLF
(Alshawi & Crouch 92)

UDRT
(Reyle 93)

LDG
(Muskens 95)

Hole Semantics
(Bos 96)

Dominance constraints/graphs
(Koller et al. 00)

MRS
(Copestake et al. 99)

CLLS
(Egg et al. 98)

. . .
MRS (simplified)

- Underspecified descriptions can be seen as graphs.
- **Configure** nodes into a tree such that dominance wishes are respected.
- All holes must be "plugged".
- A node can have more than one label.
"Every student reads a book."

\[ \forall x \text{ student} \rightarrow x \text{ read } y \land \exists y \text{ book } y \]
Description has exactly two solutions
Translating MRS to dominance graphs

Usp. description (MRS)

Usp. description (dominance)
Solutions of the MRS description correspond to configurations of the dominance graph.

This is pretty nice.

But: Configurations are not the standard notion of solutions for which e.g. dominance solvers are defined.
... but not quite

Usp. description (MRS) ≠ Usp. description (dominance)

solution ≠ solution
Why not?

- So, strictly speaking, the translation is not correct!
- Crucial case: Graph contains subgraph of the following kind:

\[ \text{\begin{tikzpicture}
  \node (1) at (0,0) {};
  \node (2) at (0,-1) {};
  \draw (1) -- (2);
\end{tikzpicture}} \]

- Does this situation ever occur in practice? (If no, the translation works well enough!)
A dominance graph is hypernormally connected (or a net) iff every pair of nodes is connected by a simple hypernormal path.

Hypernormal paths: undirected paths that do not use adjacent dominance edges out of the same leaf:
The MRS-to-dominance translation is correct on all nets.

The Hole Semantics-to-dominance translation and its reverse are correct on all nets.
Are real descriptions nets?

- Equivalence results very satisfying, but: Are the equivalent fragments practically relevant?
- For small fragments of English, this can be proved (K., Niehren, Thater 2003).
- For large-scale grammars, proof is infeasible.
**Net hypothesis**: All useful underspecified descriptions are hypernormally connected.
Evaluation of the Net Hypothesis

Redwoods Treebank (6200 sentences)

English Resource Grammar

83% are nets

17% seem to be systematically incomplete

MRS

Fuchss et al., ACL 2004
Non-nets seem to be incomplete

"A cafeteria and a sauna are available."
Completions are nets

"A cafeteria and a sauna are available."
Taking stock

- Equivalence results for the hypernormally connected fragments of dominance graphs, MRS, and Hole Semantics.
- Most (or all!) relevant underspecified descriptions are indeed hypernormally connected.
- What does this give us?
Efficient solving

Fuchss et al., ACL 2004
Harvesting 2: Grammar verification

- When developing a large grammar, it is not trivial to keep track of interdependencies and long-range effects of changes.

- Semantic output is hard to verify.

- If all correct underspecified descriptions are hnc, non-nets may be a warning sign for errors in the grammar's semantics component.
Different usp. formalisms use different atoms to talk about formulas.

MRS can express things that dominance graphs can't:
- "equality up to quantifiers" (restricted form of dominance): empirically shown to be unnecessary
- can label the same node with multiple labels: never necessary in nets
Conclusion

- Underspecification: deal with (scope) ambiguity by delaying enumeration of readings.

- Many different underspecification formalisms - but are they really different?

- Yes, they are!

- But relevant fragments (nets) of dominance graphs, MRS, and Hole Semantics are indeed equivalent.

- Sharing of resources.
Future Work

- Extend equivalence results to other underspecification formalisms.
- Develop and implement efficient test for hypernormal connectedness, and re-run evaluation on new version of grammar.
- Closer inspection of non-nets from the treebank.
- Let's get on with underspecification work, instead of inventing new formalisms!