Expressive Power and Complexity of Underspecified Representations

Christian Ebert

Department of Linguistics and Literature
University of Bielefeld

christian.ebert@uni-bielefeld.de

I. Underspecified Representations

II. Expressive Power

III. Complexity Issues

IV. Conclusion

Scope Ambiguities

- Scope ambiguities are pervasive in natural language
- Example:

  Each child told two teachers a story

- Possible meaning:

  There are a story and two teachers such that each child told these teachers this story.

- Formally:

$$\exists y(\mathsf{story}'(y) \wedge \mathsf{two}(z, \mathsf{teacher}'(z), \forall x(\mathsf{child}'(x) \rightarrow \mathsf{tell}'(x, y, z)))$$

- Scope Pattern: $(\exists \succ 2 \succ \forall)$

Scope Ambiguities

- Three scopal elements (here: quantified DPs), which can interact freely

  $\Rightarrow$ six different possible permutations of these elements $\Rightarrow$ six (different) readings

$$\forall x(\mathsf{child}'(x) \longrightarrow \mathsf{two}(z, \mathsf{teacher}'(z), \exists y(\mathsf{story}'(y) \wedge \mathsf{tell}'(x, y, z)))) \qquad (\forall 2 \exists)$$

$$\forall x(\mathsf{child}'(x) \longrightarrow \exists y(\mathsf{story}'(y) \wedge \mathsf{two}(z, \mathsf{teacher}'(z), \mathsf{tell}'(x, y, z)))) \qquad (\forall \exists 2)$$

$$\exists y(\mathsf{story}'(y) \wedge \forall x(\mathsf{child}'(x) \longrightarrow \mathsf{two}(z, \mathsf{teacher}'(z), \mathsf{tell}'(x, y, z)))) \qquad (\exists \forall 2)$$

$$\exists y(\mathsf{story}'(y) \wedge \mathsf{two}(z, \mathsf{teacher}'(z), \forall x(\mathsf{child}'(x) \longrightarrow \mathsf{tell}'(x, y, z)))) \qquad (\exists 2 \forall)$$

$$\mathsf{two}(z, \mathsf{teacher}'(z), \exists y(\mathsf{story}'(y) \wedge \forall x(\mathsf{child}'(x) \longrightarrow \mathsf{tell}'(x, y, z)))) \qquad (2 \exists \forall)$$

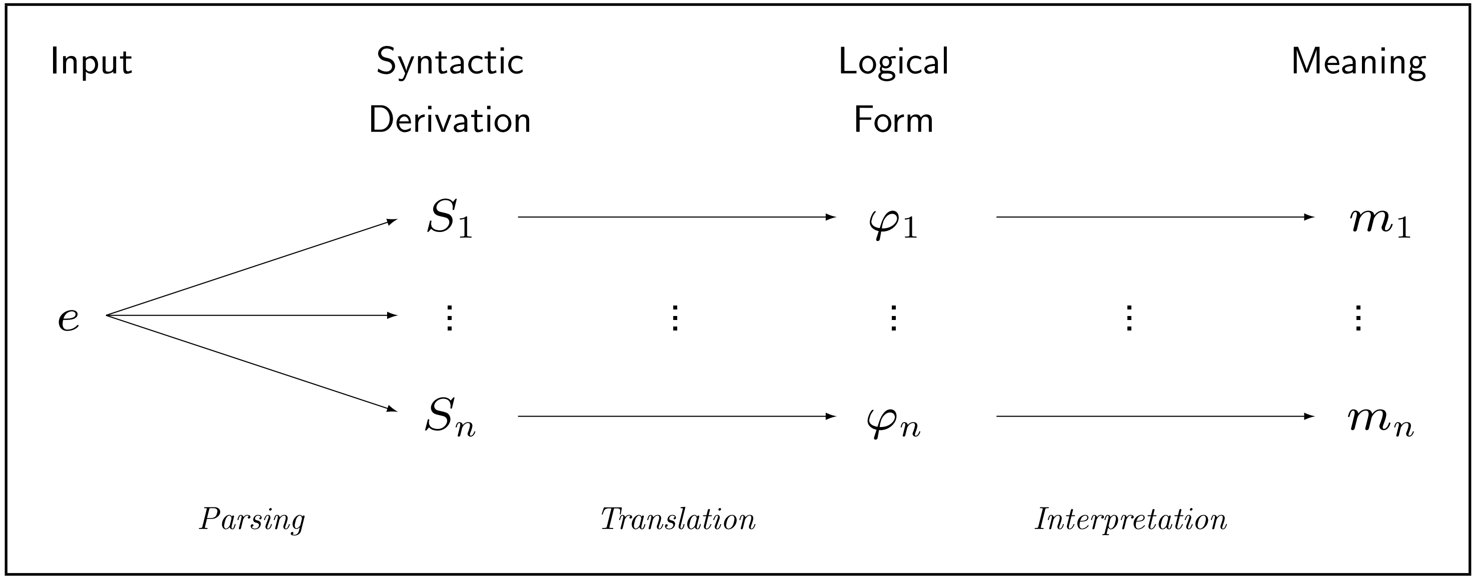
$$\mathsf{two}(z, \mathsf{teacher}'(z), \forall x(\mathsf{child}'(x) \longrightarrow \exists y(\mathsf{story}'(y) \wedge \mathsf{tell}'(x, y, z)))) \qquad (2 \forall \exists)$$

Derivations à la Montague

- Derivation of those readings in the Montagovian Framework:

  Order of *Quantifying-In* determines reading

- Distinct readings $\iff$ distinct derivations



Derivations à la Montague

Problem:

- Number of readings grows massively in the number of scopal elements in the worst case
- Well-known example:

  A politician can fool most voters on some issues most of the time,
  but no politician can fool all voters on every issue all of the time.

- consists of two clauses with five scopal elements each

  $\Rightarrow 5! * 5! = 14400$ ways of combining them

- In general:
  Clause contains n scopal elements $\Rightarrow$ $n!$ readings in the worst case

Derivations à la Montague

Combinatorial Explosion:

Computation/derivation of all $n!$ readings is. . .

1. unwarranted from a psycholinguistic point of view:

   humans do not seem to carry out those computations/derivations

   humans are often not even aware of ambiguities

2. inefficient concerning software implementation:

   feasible algorithms $\approx$ polynomial-time algorithms

   growth of $n!$ is far beyond polynomial growth

▶ Avoid combinatorial explosion by means of **underspecification**

Underspecification

- Only one syntactic analysis/derivation

- Only one underspecified representation that stands for all readings/logical forms simultaneously

- In an ideal world: only one underspecified meaning

Input	Syntactic Derivation	Underspecified Representation	Meaning
e	S	u	m_u

$$e \xrightarrow[\mathit{Parsing}]{} S \xrightarrow[\mathit{Construction}]{} u \xrightarrow[\mathit{Interpretation}]{} m_u$$

- Defer enumeration of readings and hope that further information (discourse/world knowledge) disambiguates

Underspecification

- Use a meta-language to talk about the logical forms of the different readings
- Use meta-variables (handles/holes/labels) to describe parts common to the logical forms
- Use constraints to control the composition of those parts

Example:

Each child told two teachers a story.

- Common parts of the six logical forms:

$$X_1 : \exists y(\mathsf{story}'(y) \wedge X_2) \qquad X_3 : \mathsf{two}(z, \mathsf{teacher}'(z), X_4) \qquad X_5 : \forall x(\mathsf{child}'(x) \rightarrow X_6)$$

$$X_7 : \mathsf{tell}'(x, y, z)))$$

Underspecification

- Constraints (where $X \lhd^* Y$ means roughly *part* Y *occurs at place* X / X *dominates* Y)

$$X_2 \lhd^* X_7 \qquad X_4 \lhd^* X_7 \qquad X_6 \lhd^* X_7$$

- graphically displayed as

$X_1 : \exists y(\mathsf{story}'(y) \wedge X_2)$ $\qquad$ $X_3 : \mathsf{two}(z, \mathsf{teacher}'(z), X_4)$ $\qquad$ $X_5 : \forall x(\mathsf{child}'(x) \rightarrow X_6)$

$X_7 : \mathsf{tell}'(x, y, z)$

- Plugging the parts together while respecting the constraints yields the set of licensed logical forms.

Partial Disambiguation

- Discourse and world knowledge can lead to partial disambiguation, i.e. rule out some of the potentially available readings while leaving others

Example:

Each child told two teachers a story.

Six readings: $\forall \succ 2 \succ \exists$ $\quad \exists \succ \forall \succ 2$ $\quad 2 \succ \forall \succ \exists$

$\forall \succ \exists \succ 2$ $\quad \exists \succ 2 \succ \forall$ $\quad 2 \succ \exists \succ \forall$

... It was *Alice in Wonderland*.

remaining two readings: $\forall \succ \exists \succ 2$ $\quad \exists \succ \forall \succ 2$ $\quad 2 \succ \forall \succ \exists$

$\forall \succ 2 \succ \exists$ $\quad \exists \succ 2 \succ \forall$ $\quad 2 \succ \exists \succ \forall$

Partial Disambiguation

- Partially disambiguated representations are created by monotonic addition of constraints to the constraint set:

$$X_2 \triangleleft^* X_7 \qquad X_4 \triangleleft^* X_7 \qquad X_6 \triangleleft^* X_7 \qquad\qquad X_2 \triangleleft^* X_3 \qquad X_2 \triangleleft^* X_5$$

- graphically represented as:

$$X_1 : \exists y(\mathsf{story}'(y) \wedge X_2)$$

$$X_3 : \mathsf{two}(z, \mathsf{teacher}'(z), X_4) \qquad\qquad X_5 : \forall x(\mathsf{child}'(x) \rightarrow X_6)$$

$$X_7 : \mathsf{tell}'(x, y, z)$$

Approaches

The following approaches all share these basic characteristics but differ in the type of constraints they use:

- Hole Semantics (Bos, 1995; Bos, 2002)

  Constraints: $X \lhd^* Y$ and sharing of label/hole variables

- Normal Dominance Constraints (NDC; Koller et. al., 2000; Koller, 2004)

  Constraints: $X \lhd^* Y, X \neq Y$ and sharing of label/hole variables

- Minimal Recursion Semantics (MRS; Copestake et. al., 1999)

  Constraints: $X =_{\text{qeq}} Y$ (dominance sensitive to the type of intervening element) and sharing of variables

Expressive Completeness

- A representational system is called expressively complete iff it is able to represent what (pre-theoretically) needs to be represented
- For approaches to underspecification this means that they have to provide representations for all possibly occurring scopal ambiguities
- This comprises the ambiguities of isolated sentences as well as partial disambiguations created by further discourse or world knowledge
- To illustrate the point, let us look at a (simple-minded) use of Cooper Storage for underspecification

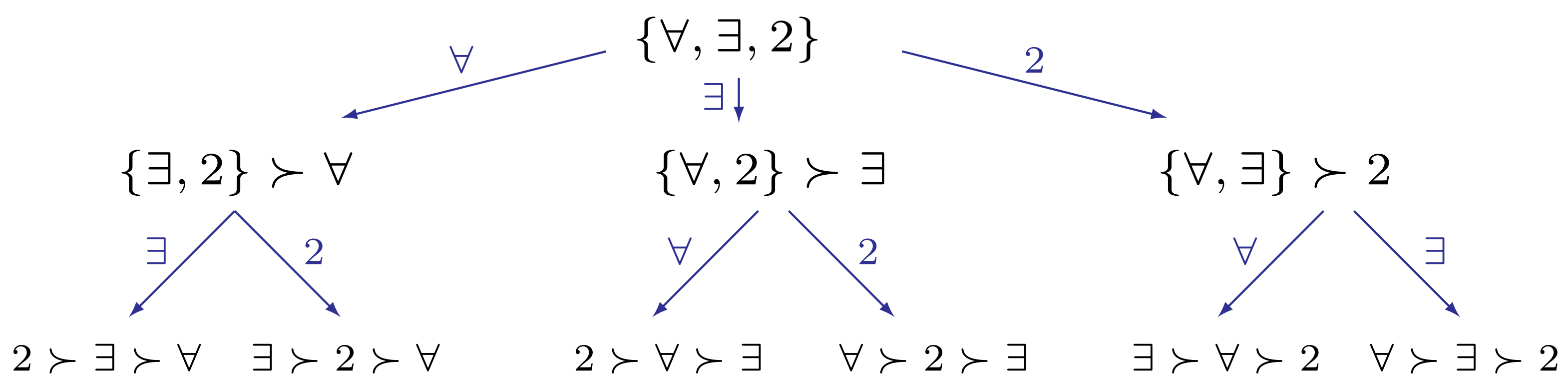
Cooper Storage as an Approach to Underspecification

Idea: Collect all scopal elements in a set...

Each child told two teachers a story.

$$\rightsquigarrow \quad \{\forall, \exists, 2\}$$

...and retrieve them in some order:



Cooper Storage as an Approach to Underspecification

- Underspecification would hence be achieved by retrieval of only some elements while leaving the others in the store.
- Observation: the retrieved elements take scope below those in the store.
- This means that the represented logical forms all end in a common sequence of scopal elements:

$$\{Q_1, \ldots, Q_i\} \succ Q_{i+1} \succ \ldots \succ Q_n$$

Problem: How can you represent the partial disambiguation (i.e. two remaining readings) of

Each child told two teachers a story. It was *Alice in Wonderland*.

$$\rightsquigarrow \quad \{\, \exists \succ \forall \succ 2 \,,\; \exists \succ 2 \succ \forall \,\}$$

Cooper Storage as an Approach to Underspecification

- Answer: You can't.
- According to the observation, there are only representations for sets of logical forms that end in a common sequence of scopal elements.
- Hence there is no such structure for

$$\{\ \exists \succ \forall \succ 2\ ,\ \exists \succ 2 \succ \forall\ \}$$

- In general, one cannot express a scope constraint of the form

Q has widest scope

- Cooper Storage is $\textit{expressively incomplete}$
- What about the other approaches?

Example 1 – an isolated sentence

- Example of a sentence with an embedded DP; from (Park, 1995; Willis & Manandhar, 1999):

  Two representatives of three companies saw most samples.

- According to (Park, 1995), this sentence has only *four* readings:

$$\mathsf{two}(x, \mathsf{three}(y, \mathsf{comp}'(y), \mathsf{rep}'(x) \wedge \mathsf{of}'(x, y)), \mathsf{most}(z, \mathsf{samp}'(z), \mathsf{saw}'(x, z)))$$

$$\mathsf{three}(y, \mathsf{comp}'(y), \mathsf{two}(x, \mathsf{rep}'(x) \wedge \mathsf{of}'(x, y), \mathsf{most}(z, \mathsf{samp}'(z), \mathsf{saw}'(x, z))))$$

$$\mathsf{most}(z, \mathsf{samp}'(z), \mathsf{two}(x, \mathsf{three}(y, \mathsf{comp}'(y), \mathsf{rep}'(x) \wedge \mathsf{of}'(x, y)), \mathsf{saw}'(x, z)))$$

$$\mathsf{most}(z, \mathsf{samp}'(z), \mathsf{three}(y, \mathsf{comp}'(y), \mathsf{two}(x, \mathsf{rep}'(x) \wedge \mathsf{of}'(x, y)), \mathsf{saw}'(x, z)))$$

- In particular, the sentence does not have this reading:

$$\mathsf{three}(y, \mathsf{comp}'(y), \mathsf{most}(z, \mathsf{samp}'(z), \mathsf{two}(x, \mathsf{rep}'(x) \wedge \mathsf{of}'(x, y), \mathsf{saw}'(x, z))))$$

- most samples does not 'intercalate' two representatives and three companies, as (Park, 1995) puts it.

Example 1 – an isolated sentence

- More schematically, the set of readings (call it P) that needs to be represented is

$$\text{two\_rep}(\text{three\_comp}(\text{of}), \text{most\_samp}(\text{saw})) \qquad \text{three\_comp}(\text{two\_rep}(\text{of}, \text{most\_samp}(\text{saw})))$$

$$\text{most\_samp}(\text{two\_rep}(\text{three\_comp}(\text{of}), \text{saw})) \qquad \text{most\_samp}(\text{three\_comp}(\text{two\_rep}(\text{of}, \text{saw})))$$

or even ($\mathtt{f} \mathrel{\hat{=}} \text{two\_rep}; \mathtt{g} \mathrel{\hat{=}} \text{three\_comp}; \mathtt{h} \mathrel{\hat{=}} \text{most\_samp}; \mathtt{x} \mathrel{\hat{=}} \text{of}; \mathtt{y} \mathrel{\hat{=}} \text{saw}$)

$$\mathtt{f}(\mathtt{g}(\mathtt{x}), \mathtt{h}(\mathtt{y})) \qquad \mathtt{g}(\mathtt{f}(\mathtt{x}, \mathtt{h}(\mathtt{y}))) \qquad \mathtt{h}(\mathtt{f}(\mathtt{g}(\mathtt{x}), \mathtt{y})) \qquad \mathtt{h}(\mathtt{g}(\mathtt{f}(\mathtt{x}, \mathtt{y})))$$

- Assumption: there is an underspecified representation u of Hole Semantics for the set P

- u must contain the part information

$$X_1 : \mathtt{f}(X_2, X_3) \qquad X_4 : \mathtt{g}(X_5) \qquad X_6 : \mathtt{h}(X_7) \qquad X_8 : \mathtt{x} \qquad X_9 : \mathtt{y}$$

- Take a closer look at the form of the representation u

Example 1 – an isolated sentence

- First, in Hole Semantics it might be the case, that some label/argument variables of those parts are shared (i.e. that they are identical).
- For instance, if $X_5 = X_6$ in those parts, this would specify that $\mathtt{h}$ has to be the daughter of $\mathtt{g}$ in *every* solution of the representation u.

  This is not the case in P (see $\mathtt{g}(\mathtt{f}(\mathtt{x}, \mathtt{h}(\mathtt{y}))))$ $\Rightarrow$ X_5 and X_6 are distinct variables in u
- Inspection of P shows:
  for no two functors f, g is it the case that f is the immediate daughter of g in *every* $t \in P$.
- Hence no sharing of variables occurs an the parts actually are as given
- Considering dominance constraints it holds that
  if u contained the constraint $X_2 \lhd^* X_4$ (for instance), then $\mathtt{f}$ would have to dominate $\mathtt{g}$ in *every* solution of u.

  This is not the case in P (see $\mathtt{g}(\mathtt{f}(\mathtt{x}, \mathtt{h}(\mathtt{y}))))$ $\Rightarrow$ u does not contain $X_2 \lhd^* X_4$

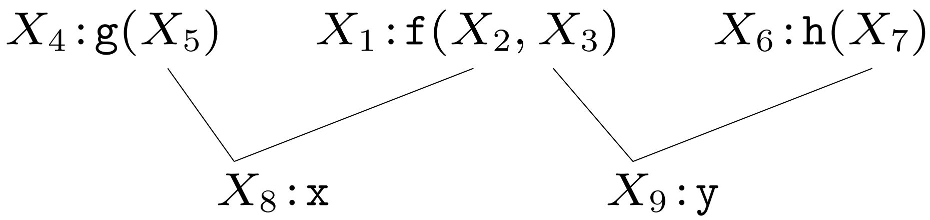
Example 1 – an isolated sentence

- In general, the dominance information D_P common to the terms in P sets an upper bound to the possible constraints D_u of u, i.e. $D_u \subseteq D_P$.

- For P from above we have

$$D_P = \{ \quad \underset{\texttt{f} \lhd^* \texttt{x}}{X_2 \lhd^* X_8,} \qquad \underset{\texttt{f} \lhd^* \texttt{y}}{X_3 \lhd^* X_9,} \qquad \underset{\texttt{g} \lhd^* \texttt{x}}{X_5 \lhd^* X_8,} \qquad \underset{\texttt{h} \lhd^* \texttt{y}}{X_7 \lhd^* X_9} \quad \}$$

- displayed as a constraint graph:



Example 1 – an isolated sentence

- But note that the unavailable fifth reading $\mathtt{g}(\mathtt{h}(\mathtt{f}(\mathtt{x},\mathtt{y})))$ also satisfies the constraints D_P
- Hence $\mathtt{g}(\mathtt{h}(\mathtt{f}(\mathtt{x},\mathtt{y})))$ would also be a solution of u, contrary to our assumption.

$\Rightarrow$ there is no Hole Semantics representation for P

- We get the same result for MRS which also fails to represent this pattern of ambiguity
- The problem is that the common dominance information alone is not sufficient to rule out the unavailable reading

Example 1 – an isolated sentence

- Note however, that there is a representation in the NDC approach making use of the inequality

$$X_5 \neq X_6$$

stating that $\mathtt{h}$ (here: three companies) may not be the daughter of $\mathtt{g}$ (here: most samples)

$$\begin{aligned} \psi := \quad & X_1 : \mathtt{f}(X_2, X_3) \wedge X_4 : \mathtt{g}(X_5) \wedge X_6 : \mathtt{h}(X_7) \wedge X_8 : \mathtt{x} \wedge X_9 : \mathtt{y} \\ & \wedge \;\; X_2 \lhd^* X_8 \wedge X_3 \lhd^* X_9 \wedge X_5 \lhd^* X_8 \wedge X_7 \lhd^* X_9 \\ & \wedge \;\; X_5 \neq X_6 \\ & \wedge \;\; \text{(additional inequalities needed by a well-formed NDC)} \end{aligned}$$

- These considerations straightforwardly lead to a comparison of the approaches as a side effect: NDCs with inequalities are strictly more expressive than Hole Semantics (see Ebert, 2005; proves Theorem of (Koller et. al., 2003) wrong),
  (without inequalities they are equivalent to Hole Semantics; cf. Koller, 2004)

Example 2 – a partially disambiguated sentence

- Take our previous example

  Each child told two teachers a story.

- Suppose this examples is disambiguated in a way that rules out the $(\exists 2 \forall)$ reading.

- For instance, it may be followed by an unambiguous negated paraphrase of this reading:

  . . . But it is not the case that there was one story und two teachers, such that each child told this story to these teachers.

- Hence the initially six-fold ambiguous sentence is partially disambiguated, leaving only a five-fold ambiguity

$$(\forall 2 \exists)\ (\forall \exists 2)\ (2 \forall \exists)\ (2 \exists \forall)\ (\exists \forall 2)$$

- How would a representation for this set of readings look like?

Example 2 – a partially disambiguated sentence

- Let us look at the more abstract set of terms (call it Q)

$$\mathtt{f(g(h(x)))} \qquad \mathtt{f(h(g(x)))} \qquad \mathtt{g(f(h(x)))} \qquad \mathtt{g(h(f(x)))} \qquad \mathtt{h(f(g(x)))}$$

- The parts do again have no shared variables

$$X_1 : \mathtt{f}(X_2) \qquad X_3 : \mathtt{g}(X_4) \qquad X_5 : \mathtt{h}(X_6) \qquad X_7 : \mathtt{x}$$

- Again the common dominance information in Q is vacuous w.r.t. all six possible readings
- but here also the common *inequality* information for NDCs is vacuous
- If an alleged NDC representation φ contained the constraint $X_2 \neq X_3$ (for instance), then in *no* solution `f` could immediately dominate g.
- This is not the case in Q (see $\mathtt{f(g(h(x)))}$) $\Rightarrow$ φ does not contain $X_2 \neq X_3$

Example 2 – a partially disambiguated sentence

$X_i \triangleleft^* X_j$ in φ $\Rightarrow$ functor of X_i dominates functor of X_j in each solution

$X_i \neq X_j$ in φ $\Rightarrow$ functor of X_i immediately dominates functor of X_j in no solution

immediate dominance relation					*dominance* relation				
fghx	fhgx	hgfx	ghfx	gfhx	fghx	fhgx	hgfx	ghfx	gfhx
fg					fg	fg			
	fh			fh	fh	fh			fh
		fx	fx		fx	fx	fx	fx	fx
		gf		gf			gf	gf	gf
gh			gh		gh			gh	gh
	gx				gx	gx	gx	gx	gx
			hf				hf	hf	
	hg	hg				hg	hg		
hx				hx	hx	hx	hx	hx	hx

Example 2 – a partially disambiguated sentence

- Hence the dominance constraints that could possibly occur in φ are

$$D_Q = \left\{ \begin{array}{ccc} X_2 \triangleleft^* X_7, & X_4 \triangleleft^* X_7, & X_6 \triangleleft^* X_7 \\ \mathtt{f} \triangleleft^* \mathtt{x} & \mathtt{g} \triangleleft^* \mathtt{x} & \mathtt{h} \triangleleft^* \mathtt{x} \end{array} \right\}$$

- Furthermore φ does not contain any non-vacuous inequality constraints
- Again, the unwanted reading $\mathtt{h}(\mathtt{g}(\mathtt{f}(\mathtt{x})))$ also fulfills those constraints and cannot be excluded

$\Rightarrow$ there is no NDC that represents Q

- The same is true for Hole Semantics and MRS

Expressive Completeness

- Conclusion:
  All of the investigated approaches fail to represent some (fairly simple) patterns of ambiguity

  For instance, all fail to represent the ambiguity that arises if three scopal elements interact freely and context/world knowledge excludes one of the resulting readings (cf. Example 2)

- So what are the patterns of ambiguity that actually need to be represented?

- König & Reyle (1999):
  if you deal with n distinct scopal elements, there are $n!$ distinct permutations...

$$\forall,\ \exists,\ 2 \qquad \rightsquigarrow \qquad (\forall\exists 2) \quad (\forall 2 \exists) \quad (\exists\forall 2) \quad (\exists 2 \forall) \quad (2\forall\exists) \quad (2\exists\forall)$$

- ...and there are $2^{n!}$ distinct subsets of readings...

$$\{(\forall\exists 2)\} \qquad \{(\forall\exists 2),\ (2\exists\forall)\} \qquad \{(\exists\forall 2),\ (\exists 2\forall),\ (2\exists\forall),\ (2\forall\exists)\} \qquad \ldots$$

- ... and so a formalism is expressively complete if it provides an underspecified representation for each of these subsets

Expressive Completeness

- So is natural language really that unrestricted w.r.t. scopal ambiguities?
- In isolated sentences scopal ambiguities are clearly limited (and to some extent still a mystery...)
- For instance, it may well be that there is no isolated sentence with five scopal elements $Q_1, \ldots, Q_5$ that is two-fold ambiguous between the readings (call this set R)

$$Q_1 Q_2 Q_3 Q_4 Q_5 \qquad\qquad Q_5 Q_4 Q_3 Q_2 Q_1$$

- But it *is* reasonable to assume that this 'weird' set of readings R can be produced by partial disambiguations due to context/world knowledge
- In other words: the claim, that R is no linguistically relevant set of readings comes down to the claim that

  1. there is no isolated sentence that is ambiguous between the readings in R, and
  2. there is no way context/world knowledge could evolve such that partial disambiguation leads to the ambiguity in R

- In particular the latter point seems to be an unreasonable claim

Example 3 – Negative Concord

- Even this unrestricted view of König & Reyle may not be general enough to capture all occurring ambiguities
- French data from (Corblin, 1996):

  Personne n'aime personne
  nobody NEG love nobody

- This sentence is two-fold ambiguous:
  1. Negative Concord Reading:

     $$\neg\exists x \exists y(\mathsf{love}'(x, y)) \quad \equiv \quad \forall x \forall y(\neg \mathsf{love}'(x, y))$$

     'Nobody loves nobody' (there is no love in the world)
  2. Double Negation Reading:

     $$\neg\exists x \neg\exists y(\mathsf{love}'(x, y)) \quad \equiv \quad \forall x \exists y(\mathsf{love}'(x, y))$$

     'Nobody is such that he loves nobody' (everybody loves somebody)
- The common set of parts must provide two negations for the double negation reading but in the negative concord reading these must be identified

Expressive Completeness

- Let us ignore cases like Example 3 for the moment and adopt König & Reyle's (1999) reasonable idea of expressive completeness
- Observation:

  Hole Semantics, NDCs, and MRS are *expressively incomplete*

  as each approach fails to provide a representation for the pattern of ambiguity

  $$\mathtt{f(g(h(x)))} \qquad \mathtt{f(h(g(x)))} \qquad \mathtt{g(f(h(x)))} \qquad \mathtt{g(h(f(x)))} \qquad \mathtt{h(f(g(x)))}$$

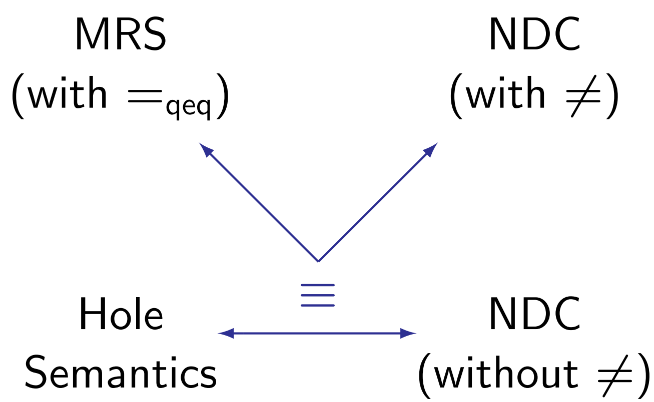
  an instance of which were the readings of the partially disambiguated Example 2.
- Another example which Hole Semantics and MRS cannot represent is

  $$\mathtt{f(h(g(x)))} \qquad \mathtt{g(h(f(x)))} \qquad \text{(}\mathtt{h}\text{ takes intermediate scope)}$$

- Obviously, those patterns are independent of the concrete scopal elements

A Comparison

- As a side effect, we get the following results concerning a comparison of the approaches with respect to their expressive power (see also Koller, 2004; Niehren & Thater, 2003; Player, 2004 for related research).



- Question:
  What would an expressively complete formalism look like?

I. Underspecified Representations

II. Expressive Power

III. Complexity Issues

IV. Conclusion

Complexity Considerations in the Literature

- So far, the work on complexity of underspecified representation has focussed on satisfiability:

  Given an underspecified representation, does it actually represent anything?

- For instance, the satisfiability of (non-normal) dominance constraints is NP-complete (Koller et. al., 1998)
- Hence Koller et. al. (2000) devise the restricted fragment of *normal* dominance constraints, which can be checked for satisfiability in polynomial time.
- With the same motivation, Willis (2000) has designed a polynomially satisfiable formalism
- Although these questions are interesting from a practical point of view in particular, they do not contribute to a solution of the combinatorial explosion problem.

A naïve proposal

- Extremely Naïve Proposal:
  Represent each ambiguity by the set of logical forms

- E.g. Example 1: Two representatives of three companies saw most samples. $\rightsquigarrow$

$$\{\ \mathsf{two}(x, \mathsf{three}(y, \mathsf{comp}'(y), \mathsf{rep}'(x) \wedge \mathsf{of}'(x,y)), \mathsf{most}(z, \mathsf{samp}'(z), \mathsf{saw}'(x,z))),$$

$$\mathsf{three}(y, \mathsf{comp}'(y), \mathsf{two}(x, \mathsf{rep}'(x) \wedge \mathsf{of}'(x,y), \mathsf{most}(z, \mathsf{samp}'(z), \mathsf{saw}'(x,z)))),$$

$$\mathsf{most}(z, \mathsf{samp}'(z), \mathsf{two}(x, \mathsf{three}(y, \mathsf{comp}'(y), \mathsf{rep}'(x) \wedge \mathsf{of}'(x,y)), \mathsf{saw}'(x,z))),$$

$$\mathsf{most}(z, \mathsf{samp}'(z), \mathsf{three}(y, \mathsf{comp}'(y), \mathsf{two}(x, \mathsf{rep}'(x) \wedge \mathsf{of}'(x,y)), \mathsf{saw}'(x,z)))\ \}$$

- This 'approach' is expressively complete!

- Satisfiability can be checked in constant time!

- The enumeration of the encoded readings can be performed in linear time!

A less naïve proposal?

- Another (less naïve?) proposal:
  Use *disjunction* in addition to conjunction in the interpretation of the constraints

- This approach also is expressively complete, as one can disjoin fully specified conjunctive representations of single terms

- For instance, the following constraint represents $\{\mathtt{fghx}, \mathtt{fhgx}, \mathtt{hgfx}, \mathtt{ghfx}, \mathtt{gfhx}\}$ (Example 2):

$$\begin{aligned}\varphi = \Big(\; & (X:\mathtt{f}(Y) \wedge Y:\mathtt{g}(Z) \wedge Z:\mathtt{h}(V) \wedge V:\mathtt{x}) \\ \vee\; & (X:\mathtt{f}(Z) \wedge Y:\mathtt{g}(V) \wedge Z:\mathtt{h}(Y) \wedge V:\mathtt{x}) \\ \vee\; & (X:\mathtt{f}(V) \wedge Y:\mathtt{g}(X) \wedge Z:\mathtt{h}(Y) \wedge V:\mathtt{x}) \\ \vee\; & (X:\mathtt{f}(V) \wedge Y:\mathtt{g}(Z) \wedge Z:\mathtt{h}(X) \wedge V:\mathtt{x}) \\ \vee\; & (X:\mathtt{f}(Z) \wedge Y:\mathtt{g}(X) \wedge Z:\mathtt{h}(V) \wedge V:\mathtt{x})\Big)\end{aligned}$$

A less naïve proposal?

- Obviously, this representation is not any better than the naïve enumeration of all readings
- With disjunction around, one starts to 'enumerate' information instead of 'underspecifying' it
- The main point of these ridiculous examples:

  The construction of the 'underspecified representation' has to do all the work

  Whereas the Combinatorial Explosion Problem is a 'feature' of the Montagovian **framework** itself, it lurks in the **construction** process in the Underspecification framework.

  Hence the Underspecification framework does not solve the combinatorial explosion problem $per\ se$, but only if the involved processes – in particular the **construction** process – are efficient (cf. Dörre, 1997).
- The compactness (i.e. 'small size w.r.t. the number of scopal elements') of the representations is a necessary requirement for efficient construction

Compact Representations

- An underspecified representation u represents a set of readings built out of some scopal elements $Q_1, \ldots, Q_n$
- We can hence define an approach to produce compact representations, if the size of its representations is polynomial ($\approx$ 'feasible') in the number n of underlying scopal elements.
- Formally:
  Suppose the underspecified representations are drawn from some formal language $U \subseteq (\mathcal{A} \cup \mathcal{V})^*$ over some finite alphabet of symbols $\mathcal{A}$ and some infinite set of variables $\mathcal{V}$
- Then let

$$\lambda(n) := \max\big\{|u| \mid u \in U \text{ represents readings over } n \text{ scopal elements}\big\}$$

be the maximal size of a representation that is needed by formalism U to represent a set of readings over n scopal elements

Compact Representations

- Definition:

$$U \text{ is compact } :\Longleftrightarrow \lambda(n) \in O(n^c) \text{ for some } c \in \mathbb{N}$$

- With this definition:

approach not compact $\Rightarrow$ construction not feasible $\Rightarrow$ combinatorial explosion

- For instance, this definition classifies the naïve proposal as not compact as the representations are of factorial size in the worst case

Expressive Completeness $\leftrightarrow$ Compactness

- Take a closer look at the worst case growth rates:

No. scopal elements n	No. readings $n!$	No. patterns of ambiguity $2^{n!}$
1	1	2
2	2	4
3	6	64
4	24	16777216
5	120	$1,329 \cdot 10^{36}$
6	720	$5,516 \cdot 10^{216}$
7	5040	$1,553 \cdot 10^{1517}$
8	40320	$3,384 \cdot 10^{12137}$

- e.g. $n = 5$ scopal elements may generate $n! = 120$ readings...
- ...and those may form $1,329 \cdot 10^{36}$ sets/patterns of ambiguity/partial disambiguations

Expressive Completeness $\leftrightarrow$ Compactness

- Hence an expressively complete formalism must provide $1,329 \cdot 10^{36}$ representations...
- ...but each of those must be shorter than listing all $n! = 120$ readings in order to be compact (i.e. do better than the naïve proposal and avoid combinatorial explosion)
- Problem:
  there are not enough 'polynomially short' representations, so you necessarily have to use some longer ones.
- For a given length l, an upper bound on the number of distinct representations of size l is

$$(|\mathcal{A}| + l)^l$$

  (for each of the l symbols either a symbol from $\mathcal{A}$, or one of maximally l variables)
- Hence, up to a given length m, the formalism U provides no more than ... representations

$$\sum_{i=1}^{m} (|\mathcal{A}| + i)^i \;\leq\; m \cdot (|\mathcal{A}| + m)^m$$

Expressive Completeness $\leftrightarrow$ Compactness

- Now recall that U must provide $2^{n!}$ representations in order to expressively complete (for any given n)
- Suppose we can use *every* representation up to length $\lambda(n)$ and that each such representation represents a different set from the $2^{n!}$ we need

$$\lambda(n) \cdot (|\mathcal{A}| + \lambda(n))^{\lambda(n)} \;\geq\; 2^{n!}$$

- Then we get ($\lg = \log_2$)

$$\lg \lambda(n) + \lambda(n) \cdot \lg\left(|\mathcal{A}| + \lambda(n)\right) \;\geq\; n!$$

- If U is compact this yields the following contradiction

$$\underbrace{c \cdot \lg n + n^c \cdot \lg\left(|\mathcal{A}| + n^c\right)}_{\in O(n^c \cdot \lg n)} \;\geq\; \underbrace{n!}_{\in O(n!)}$$

Expressive Completeness $\leftrightarrow$ Compactness

- Hence, under the assumption that U is expressively complete, $\lambda(n)$ must grow faster than polynomial $\implies$ U is not compact

- So we have:

  An underspecified representations formalism cannot both
  be expressively complete and compact.

  (see Ebert, 2005 for further details)

- In other words:
  Either the formalisms fails to represent some potential sets of readings or it is not feasible (i.e. does not avoid the combinatorial explosion problem).

Conclusion

- Avoidance of combinatorial explosion is *the* most important motivation for using underspecification
- Expressive completeness is a requirement for every linguistically adequate representation formalism
- Unfortunately, it seems that you cannot have both at the same time
- Virtually all of the seminal approaches to underspecification have focussed on the first point and striven for efficient procedures working on the representations (e.g. satisfaction/enumeration of readings)
- Approaches focussing on the second have emerged only very recently

Conclusion

- Two examples:
  - Underspecification in Property Theory with Curry Typing (PTCT; Fox & Lappin, 2005)

    uses essentially first-order logic to express complex filter conditions on possible readings
  - Lexical Resource Semantics (LRS; Richter & Sailer, 2004)

    integrates an underspecification module as essential part of HPSGs using the Relational Speciate Re-entrant Language (RSRL)

    enables formulation of complex constraints and particularly allows for treatment of negative concord phenomena (cf. Example 3) due to the possible unification of parts.
- Both approaches are expressively complete, but forget about the complexity...
- Morale: you have to know, which side you are on

Thanks!

Bibliography

- Bos, Johan (1995). Predicate Logic Unplugged. In: *Proceedings of the 10th Amsterdam Colloquium*, Amsterdam.
- Bos, Johan (2002). *Underspecification and Resolution in Discourse Semantics*. PhD thesis, Universität des Saarlandes.
- Copestake, Ann, Dan Flickinger, and Ivan A. Sag (1999). Minimal Recursion Semantics: An Introduction. Draft.
- Corblin, Francis (1996). Multiple Negation Processing in Natural Language, *Theoria* 17, pp. 214–259.
- Dörre, Jochen (1997). Efficient Construction of Underspecified Semantics under Massive Ambiguity. In: *Proceedings of the 35th Conference of the Association for Computational Linguistics*, pp. 386–393, Morristown, NJ, USA.
- Ebert, Christian (2003). On the Expressive Completeness of Underspecified Representations. In: *Proceedings of the 14th Amsterdam Colloquium.*, Amsterdam.
- Ebert, Christian (2005). *Formal Investigations of Underspecified Representations*. PhD thesis, King's College London.
- Fox, Chris & Shalom Lappin (2005). *Foundations of Intensional Semantics*. Blackwell.
- König, Esther & Uwe Reyle (1999). A General Reasoning Scheme for Underspecified Representations. In: *Logic, Language, and Reasoning. Essays in Honour of Dov Gabbay.*, Kluwer. pp. 251–277.
- Koller, Alexander (2004). *Constraint-Based and Graph-Based Resolution of Ambiguities in Natural Language*. PhD thesis, Universität des Saarlandes.
- Koller, Alexander, Joachim Niehren, and Stefan Thater (2003). Bridging the Gap between Underspecification Formalisms: Hole Semantics as Dominance Constraints. In *Proceedings of the 11th EACL*, Budapest.
- Koller, Alexander, Joachim Niehren, and Ralf Treinen (1998). Dominance Constraints: Algorithms and Complexity. In *Proceedings of the Third Conference on Logical Aspects of Computational Linguistics (LACL '98)*, Grenoble, France.
- Koller, Alexander, Kurt Mehlhorn, and Joachim Niehren (2000). A Polynomial-Time Fragment of Dominance Constraints. In: *Proceedings of the 38th Conference of the ACL*, Hong Kong.
- Niehren, Joachim and Stefan Thater (2003). Bridging the Gap between Underspecification Formalisms: Minimal Recursion Semantics as Dominance Constraints. In: *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo.
- Park, Jong C. (1995). Quantifier Scope and Constituency. In: *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics.*, pp. 205-212, Cambridge, MA.

- Player, Nickie J. (2004). *Logics of Ambiguity*. PhD thesis, University of Manchester.
- Richter, Frank & Manfred Sailer (2004). Basic Concepts of Lexical Resource Semantics. In: Arnold Beckmann, Norbert Preining (eds): *ESSLLI 2003 – Course Material I*. Collegium Logicum, Volume 5, Kurt Gödel Society. Wien. pp. 87–143.
- Willis, Alistair (2000). *An Efficient Treatment of Quantification in Underspecified Semantic Representations*. PhD thesis, University of York.
- Willis, Alistair and Suresh Manandhar (1999). Two Accounts of Scope Availability and Semantic Underspecification. In: *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*. pp. 293–300.